Package ‘powerSurvEpi’

March 1, 2021

Version 0.1.3

Date 2021-02-28

Title Power and Sample Size Calculation for Survival Analysis of Epidemiological Studies

Author Weiliang Qiu, Jorge Chavarro, Ross Lazarus, Bernard Rosner, Jing Ma.

Maintainer Weiliang Qiu <weiliang.qiu@gmail.com>

Depends R (>= 3.6.0)

Imports stats, survival, pracma

Description Functions to calculate power and sample size for testing main effect or interaction effect in the survival analysis of epidemiological studies (non-randomized studies), taking into account the correlation between the covariate of the interest and other covariates. Some calculations also take into account the competing risks and stratified analysis. This package also includes a set of functions to calculate power and sample size for testing main effect in the survival analysis of randomized clinical trials and conditional logistic regression for nested case-control study.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2021-03-01 06:30:02 UTC

R topics documented:

`numDEpi` *Calculate Number of Deaths Required for Cox Proportional Hazards Regression with Two Covariates for Epidemiological Studies*

Description

Calculate number of deaths required for Cox proportional hazards regression with two covariates for epidemiological Studies. The covariate of interest should be a binary variable. The other covariate can be either binary or non-binary. The formula takes into account competing risks and the correlation between the two covariates. Some parameters will be estimated based on a pilot data set.

Usage

```
numDEpi(X1,
X2,
power,
theta,
alpha = 0.05)
```

Arguments

`X1`	numeric. a `nPilot` by 1 vector, where `nPilot` is the number of subjects in the pilot data set. This vector records the values of the covariate of interest for the `nPilot` subjects in the pilot study. `X1` should be binary and take only two possible values: zero and one.
`X2`	numeric. a `nPilot` by 1 vector, where `nPilot` is the number of subjects in the pilot study. This vector records the values of the second covariate for the `nPilot` subjects in the pilot study. `X2` can be binary or non-binary.
`power`	numeric. the postulated power.
`theta`	numeric. postulated hazard ratio
`alpha`	numeric. type I error rate.

Details

This is an implementation of the calculation of the number of required deaths derived by Latouche et al. (2004) for the following Cox proportional hazards regression in the epidemiological studies:

$$h(t|x_1, x_2) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2),$$

where the covariate X_1 is of our interest. The covariate X_1 should be a binary variable taking two possible values: zero and one, while the covariate X_2 can be binary or continuous.

Suppose we want to check if the hazard of $X_1 = 1$ is equal to the hazard of $X_1 = 0$ or not. Equivalently, we want to check if the hazard ratio of $X_1 = 1$ to $X_1 = 0$ is equal to 1 or is equal to $\exp(\beta_1) = \theta$. Given the type I error rate α for a two-sided test, the total number of deaths required to achieve a power of $1 - \beta$ is

$$D = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{[\log(\theta)]^2 p(1-p)(1-\rho^2),}$$

where z_a is the $100a$-th percentile of the standard normal distribution,

$$\rho = corr(X_1, X_2) = (p_1 - p_0) \times \sqrt{\frac{q(1-q)}{p(1-p)}},$$

and $p = Pr(X_1 = 1)$, $q = Pr(X_2 = 1)$, $p_0 = Pr(X_1 = 1|X_2 = 0)$, and $p_1 = Pr(X_1 = 1|X_2 = 1)$.

p and rho will be estimated from a pilot data set.

Value

`D`	the number of deaths required to achieve the desired power with given type I error rate.
`p`	proportion of subjects taking $X_1 = 1$.
`rho2`	square of the correlation between X_1 and X_2.

Note

(1) The formula can be used to calculate power for a randomized trial study by setting `rho2=0`.

(2) When `rho2=0`, the formula derived by Latouche et al. (2004) looks the same as that derived by Schoenfeld (1983). Latouche et al. (2004) pointed out that in this situation, the interpretations are different hence the two formulae are actually different. In Latouched et al. (2004), the hazard ratio θ measures the difference of effect of a covariate at two different levels on the subdistribution hazard for a particular failure, while in Schoenfeld (1983), the hazard ratio θ measures the difference of effect on the cause-specific hazard.

References

Schoenfeld DA. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics*. 39:499-503.

Latouche A., Porcher R. and Chevret S. (2004). Sample size formula for proportional hazards modelling of competing risks. *Statistics in Medicine*. 23:3263-3274.

See Also

`numDEpi.default`

Examples

```
 # generate a toy pilot data set
 X1 <- c(rep(1, 39), rep(0, 61))
 set.seed(123456)
 X2 <- sample(c(0, 1), 100, replace = TRUE)
 res <- numDEpi(X1 = X1,
X2 = X2,
power = 0.8,
theta = 2,
alpha = 0.05)
 print(res)

 # proportion of subjects died of the disease of interest.
 psi <- 0.505

 # total number of subjects required to achieve the desired power
 ceiling(res$D / psi)
```

`numDEpi.default` *Calculate Number of Deaths Required for Cox Proportional Hazards Regression with Two Covariates for Epidemiological Studies*

Description

Calculate number of deaths required for Cox proportional hazards regression with two covariates for epidemiological Studies. The covariate of interest should be a binary variable. The other covariate can be either binary or non-binary. The formula takes into account competing risks and the correlation between the two covariates.

Usage

```
numDEpi.default(power,
theta,
p,
rho2,
alpha = 0.05)
```

Arguments

`power`	numeric. the postulated power.
`theta`	numeric. postulated hazard ratio
`p`	numeric. proportion of subjects taking the value one for the covariate of interest.
`rho2`	numeric. square of the correlation between the covariate of interest and the other covariate.
`alpha`	numeric. type I error rate.

Details

This is an implementation of the calculation of the number of required deaths derived by Latouche et al. (2004) for the following Cox proportional hazards regression in the epidemiological studies:

$$h(t|x_1, x_2) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2),$$

where the covariate X_1 is of our interest. The covariate X_1 should be a binary variable taking two possible values: zero and one, while the covariate X_2 can be binary or continuous.

Suppose we want to check if the hazard of $X_1 = 1$ is equal to the hazard of $X_1 = 0$ or not. Equivalently, we want to check if the hazard ratio of $X_1 = 1$ to $X_1 = 0$ is equal to 1 or is equal to $\exp(\beta_1) = \theta$. Given the type I error rate α for a two-sided test, the total number of deaths required to achieve a power of $1 - \beta$ is

$$D = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{[\log(\theta)]^2 p(1-p)(1-\rho^2),}$$

where z_a is the $100a$-th percentile of the standard normal distribution,

$$\rho = corr(X_1, X_2) = (p_1 - p_0) \times \sqrt{\frac{q(1-q)}{p(1-p)}},$$

and $p = Pr(X_1 = 1)$, $q = Pr(X_2 = 1)$, $p_0 = Pr(X_1 = 1|X_2 = 0)$, and $p_1 = Pr(X_1 = 1|X_2 = 1)$.

Value

The number of deaths required to achieve the desired power with given type I error rate.

Note

(1) The formula can be used to calculate power for a randomized trial study by setting `rho2=0`.

(2) When `rho2=0`, the formula derived by Latouche et al. (2004) looks the same as that derived by Schoenfeld (1983). Latouche et al. (2004) pointed out that in this situation, the interpretations are different hence the two formulae are actually different. In Latouched et al. (2004), the hazard ratio θ measures the difference of effect of a covariate at two different levels on the subdistribution hazard for a particular failure, while in Schoenfeld (1983), the hazard ratio θ measures the difference of effect on the cause-specific hazard.

References

Schoenfeld DA. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics*. 39:499-503.

Latouche A., Porcher R. and Chevret S. (2004). Sample size formula for proportional hazards modelling of competing risks. *Statistics in Medicine*. 23:3263-3274.

See Also

numDEpi

Examples

```
# Example at the end of Section 5.2 of Latouche et al. (2004)
# for a cohort study.
D <- numDEpi.default(power = 0.8,
     theta = 2,
     p = 0.39,
                     rho2 = 0.132^2,
     alpha = 0.05)

# proportion of subjects died of the disease of interest.
psi <- 0.505

# total number of subjects required to achieve the desired power
ceiling(D / psi)
```

Oph *Ophthalmology Data*

Description

The Ophthalmology data set is described in Example 14.41 on page 807 in Rosner (2006).

Usage

```
data(Oph)
```

Format

A data frame with 354 observations on the following 3 variables.

`times` a numeric vector recording the survival/censoring time for each event/censoring.

`status` a numeric vector recording if a observed time is event time (`status=1`) or censoring time (`status=0`).

`group` a factor with levels `C` (indicating control group) and `E` (indicating experimental group).

Details

This data set was from a clinical trial (Berson et al., 1993) conducted to test the efficacy of different vitamin supplements in preventing visual loss in patients with retinitis pigmentosa. Rosner (2006) used the data from this clinical trial to illustrate the analysis of survival data (Sections 14.9-14.12 of Rosner (2006)).

The data set consists of two groups of participants: (1) the experimental group (i.e., group E in which participants receiving 15,000 IU of vitamin A per day) and (2) the control group (i.e., group C in which participants receiving 75 IU of vitamin A per day).

The participants were enrolled over a 2-year period (1984-1987) and followed for a maximum of 6 years. The follow-up was terminated in September 1991. Some participants dropped out of the study before September 1991 and had not failed. Dropouts were due to death, other diseases, or side effects possibly due to the study medications, or unwillingness to comply (take study medications). There are 6 time points (at 1st year, 2nd year, 3rd year, 4th year, 5-th year, and 6-th year) in this data set.

Rosner (2006, page 786) defined the participants who do not reach a disease endpoint during their period of follow-up as censored observations. A participant has been censored at time t if the participant has been followed up to time t and has not failed. Noninformative censoring is assumed. That is, participants who are censored have the same underlying survival curve after their censoring time as patients who are not censored.

Source

Created based on Table 14.12 on page 787 of Rosner (2006).

References

Berson, E.L., Rosner, B., Sandberg, M.A., Hayes, K.C., Nicholson, B.W., Weigel-DiFranco, C., and Willett, W.C. (1993). A randomized trial of vitamin A and vitamin E supplementation for retinitis pigmentosa. *Archives of Ophthalmology*. 111:761-772.

Rosner B. (2006). *Fundamentals of Biostatistics*. (6-th edition). Thomson Brooks/Cole.

Examples

```
data(Oph)
```

power.stratify *Power Calculation for Survival Analysis with Binary Predictor and Exponential Survival Function*

Description

Power calculation for survival analysis with binary predictor and exponential survival function.

Usage

```
power.stratify(
    n,
    timeUnit,
    gVec,
    PVec,
    HR,
    lambda0Vec,
    power.ini = 0.8,
    power.low = 0.001,
    power.upp = 0.999,
    alpha = 0.05,
    verbose = TRUE)
```

Arguments

`n`	integer. Sample size.
`timeUnit`	numeric. Total study length.
`gVec`	numerc. m by 1 vector. The s-th element is the proportion of the total sample size for the s-th stratum, where m is the number of strata.
`PVec`	numeric. m by 1 vector. The s-th element is the proportion of subjects in treatment group 1 for the s-th stratum, where m is the number of strata.
`HR`	numeric. Hazard ratio (Ratio of the hazard for treatment group 1 to the hazard for treatment group 0, i.e. reference group).
`lambda0Vec`	numeric. m by 1 vector. The s-th element is the hazard for treatment group 0 (i.e., reference group) in the s-th stratum.
`power.ini`	numeric. Initial power estimate.
`power.low`	numeric. Lower bound for power.
`power.upp`	numeric. Upper bound for power.
`alpha`	numeric. Type I error rate.
`verbose`	Logical. Indicating if intermediate results will be output or not.

Details

We assume (1) there is only one predictor and no covariates in the survival model (exponential survival function); (2) there are `m` strata; (3) the predictor `x` is a binary variable indicating treatment group 1 ($x = 1$) or treatment group 0 ($x = 0$); (3) the treatment effect is constant over time (proportional hazards); (4) the hazard ratio is the same in all strata, and (5) the data will be analyzed by the stratified log rank test.

The sample size formula is Formula (1) on page 801 of Palta M and Amini SB (1985):

$$n = (Z_\alpha + Z_\beta)^2/\mu^2$$

where α is the Type I error rate, β is the Type II error rate (power= $1-\beta$), Z_α is the $100(1-\alpha)$-th percentile of standard normal distribution, and

$$\mu = \log(\delta)\sqrt{\sum_{s=1}^{m} g_s P_s (1-P_s) V_s}$$

and

$$V_s = P_s\left[1-\frac{1}{\lambda_{1s}}\left\{\exp\left[-\lambda_{1s}(T-1)\right]-\exp(-\lambda_{1s}T)\right\}\right]+(1-P_s)\left[1-\frac{1}{\lambda_{0s}}\left\{\exp\left[-\lambda_{0s}(T-1)\right]-\exp(-\lambda_{0s}T\right\}\right]$$

In the above formulas, m is the number of strata, T is the total study length, δ is the hazard ratio, g_s is the proportion of the total sample size in stratum s, P_s is the proportion of stratum s, which is in treatment group 1, and λ_{is} is the hazard for the i-th treatment group in stratum s.

Value

A list of 2 elments.

`power` Estimated power

`res.optim` Object returned by funciton `optim`. We used numerical optimization method to calculate power based on sample size calculation formula.

References

Palta M and Amini SB. (1985). Consideration of covariates and stratification in sample size determination for survival time studies. *Journal of Chronic Diseases*. 38(9):801-809.

See Also

`ssize.stratify`

Examples

```
# example on page 803 of Palta M and Amini SB. (1985).
res.power <- power.stratify(
  n = 146,
  timeUnit = 1.25,
  gVec = c(0.5, 0.5),
```

```
  PVec = c(0.5, 0.5),
  HR = 1 / 1.91,
  lambda0Vec = c(2.303, 1.139),
  power.ini = 0.8,
  power.low = 0.001,
  power.upp = 0.999,
  alpha = 0.05,
  verbose = TRUE
  )
```

`powerConLogistic.bin` *Sample Size Calculation for Conditional Logistic Regression with Binary Covariate*

Description

Sample Size Calculation for Conditional Logistic Regression with Binary Covariate, such as matched logistic regression or nested case-control study.

Usage

```
powerConLogistic.bin(
  N = NULL,
  power = 0.8,
  OR,
  pE,
  nD,
  nH,
  R2 = 0,
  alpha = 0.05,
  nTests = 1,
  OR.low = 1.01,
  OR.upp = 100
)
```

Arguments

`N`	integer. Number of sets. Each set contains `nD` cases and `nH` controls.
`power`	numeric. Power of the test for if the exposure variable is associated with the risk of diseases
`OR`	numeric. Odds ratio $= exp(\theta)$, where θ is the regression coefficient of the exposure variable.
`pE`	numeric. Population prevalence of exposure.
`nD`	integer. Number of cases per set.
`nH`	integer. Number of controls per set.

`R2`	numeric. Coefficient of determination of the exposure variable and other covariates
`alpha`	numeric. family-wise type I error rate.
`nTests`	integer. Number of tests.
`OR.low`	numeric. Lower bound of odds ratio. Only used when `OR=NULL` and `power` and `N` are not equal to NULL.
`OR.upp`	numeric. Upper bound of odds ratio. Only used when `OR=NULL` and `power` and `N` are not equal to NULL.

Details

The power and sample size calculation formulas are provided by Lachin (2008, Section 3.3, Formula (38))

$$power = \Phi\left(\sqrt{Nc} - z_{\alpha/(2nTests)}\right)$$

and

$$N = (z_{power} + z_{\alpha/(2nTests)})^2/c$$

where Φ is the cumulative distribution function of the standard normal distribution $N(0,1)$, z_a is the upper $100a$-th percentile of $N(0,1)$,

$$c = \theta^2 pE(1-pE)(1-R^2)nD * nH/(nD+nH)$$

and R^2 is the coefficient of determination for linear regression linking the exposure with other covariates.

Value

If the inputs `is.null(N) = TRUE` and `is.null(power) = FALSE`, then the function returns the number `N` of sets.

If the inputs `is.null(N) = FALSE` and `is.null(power) = TRUE`, then the function returns the power.

Otherwise, an error message is output.

References

Lachin, JM Sample Size Evaluation for a Multiply Matched Case-Control Study Using the Score Test From a Conditional Logistic (Discrete Cox PH) Regression Model. Stat Med. 2008 27(14): 2509-2523

Examples

```
# estimate power
power = powerConLogistic.bin(
  N = 59,
  power = NULL,
  OR = 3.5,
  pE = 0.15,
```

```
  nD = 1,
  nH = 2,
  R2 = 0,
  alpha = 0.05,
  nTests = 1)

print(power) # 0.80

# estimate N (number of sets)
N = powerConLogistic.bin(
  N = NULL,
  power = 0.80,
  OR = 3.5,
  pE = 0.15,
  nD = 1,
  nH = 2,
  R2 = 0,
  alpha = 0.05,
  nTests = 1)

print(ceiling(N)) # 59

# estimate OR
OR = powerConLogistic.bin(
  N = 59,
  power = 0.80,
  OR = NULL,
  pE = 0.15,
  nD = 1,
  nH = 2,
  R2 = 0,
  alpha = 0.05,
  nTests = 1,
  OR.low = 1.01,
  OR.upp = 100)

print(OR) # 3.49
```

`powerConLogistic.con` *Sample Size Calculation for Conditional Logistic Regression with Continuous Covariate*

Description

Sample Size Calculation for Conditional Logistic Regression with Continuous Covariate, such as matched logistic regression or nested case-control study.

Usage

```
powerConLogistic.con(
  N = NULL,
  power = 0.8,
  OR,
  sigma,
  nD,
  nH,
  R2 = 0,
  alpha = 0.05,
  nTests = 1,
  OR.low = 1.01,
  OR.upp = 100
)
```

Arguments

`N`	integer. Number of sets. Each set contains `nD` cases and `nH` controls.
`power`	numeric. Power of the test for if the exposure variable is associated with the risk of diseases
`OR`	numeric. Odds ratio $= exp(\theta)$, where θ is the regression coefficient of the exposure variable.
`sigma`	numeric. Standard deviation of the continuous exposure variable.
`nD`	integer. Number of cases per set.
`nH`	integer. Number of controls per set.
`R2`	numeric. Coefficient of determination of the exposure variable and other covariates
`alpha`	numeric. family-wise type I error rate.
`nTests`	integer. Number of tests.
`OR.low`	numeric. Lower bound of odds ratio. Only used when `OR=NULL` and `power` and `N` are not equal to NULL.
`OR.upp`	numeric. Upper bound of odds ratio. Only used when `OR=NULL` and `power` and `N` are not equal to NULL.

Details

The power and sample size calculation formulas are provided by Lachin (2008, Section 3.1, Formulas (24) and (25))

$$power = \Phi\left(\sqrt{Nc} - z_{\alpha/(2nTests)}\right)$$

and

$$N = \left(z_{power} + z_{\alpha/(2nTests)}\right)^2/c$$

where Φ is the cumulative distribution function of the standard normal distribution $N(0,1)$, z_a is the upper $100a$-th percentile of $N(0,1)$,

$$c = \theta^2 \sigma^2 nD(1 - 1/b)(1 - R^2)$$

and b is the Binomial coefficient (n chooses nD), $n = nD + nH$, and R^2 is the coefficient of determination for linear regression linking the exposure with other covariates.

Value

If the inputs `is.null(N) = TRUE` and `is.null(power) = FALSE`, then the function returns the number `N` of sets.

If the inputs `is.null(N) = FALSE` and `is.null(power) = TRUE`, then the function returns the power.

Otherwise, an error message is output.

References

Lachin, JM Sample Size Evaluation for a Multiply Matched Case-Control Study Using the Score Test From a Conditional Logistic (Discrete Cox PH) Regression Model. Stat Med. 2008 27(14): 2509-2523

Examples

```
library(pracma)

# Section 4.1 in Lachin (2008)

# estimate number of sets
N = powerConLogistic.con(N = NULL,
                                power = 0.85,
                                OR = 1.39,
                                sigma = 1,
                                nD = 1,
                                nH = 2,
                                R2 = 0,
                                alpha = 0.05,
                                nTests = 1)
print(ceiling(N)) # 125

# estimate power
power = powerConLogistic.con(N = 125,
                                power = NULL,
                                OR = 1.39,
                                sigma = 1,
                                nD = 1,
                                nH = 2,
                                R2 = 0,
                                alpha = 0.05,
                                nTests = 1)
print(power) # 0.85
```

```
# estimate OR
OR = powerConLogistic.con(N = 125,
                                power = 0.85,
                                OR = NULL,
                                sigma = 1,
                                nD = 1,
                                nH = 2,
                                R2 = 0,
                                alpha = 0.05,
                                nTests = 1)
print(OR) # 1.39
```

---

powerCT *Power Calculation in the Analysis of Survival Data for Clinical Trials*

Description

Power calculation for the Comparison of Survival Curves Between Two Groups under the Cox Proportional-Hazards Model for clinical trials. Some parameters will be estimated based on a pilot data set.

Usage

```
powerCT(formula,
dat,
nE,
nC,
RR,
alpha = 0.05)
```

Arguments

formula	A formula object, e.g. `Surv(time,status) ~ x`, where `time` is a vector of survival/censoring time, `status` is a vector of censoring indicator, `x` is the group indicator, which is a factor object in R and takes only two possible values (`C` for control group and `E` for experimental group). See also the documentation of the function `survfit` in the library `survival`.
dat	a data frame representing the pilot data set and containing at least 3 columns: (1) survival/censoring time; (2) censoring indicator; (3) group indicator which is a factor object in R and takes only two possible values (`C` for control group and `E` for experimental group).
nE	integer. number of participants in the experimental group.
nC	integer. number of participants in the control group.
RR	numeric. postulated hazard ratio.
alpha	numeric. type I error rate.

Details

This is an implementation of the power calculation method described in Section 14.12 (page 807) of Rosner (2006). The method was proposed by Freedman (1982).

The movitation of this function is that some times we do not have information about m or p_E and p_C available, but we have a pilot data set that can be used to estimate p_E and p_C hence m, where $m = n_E p_E + n_C p_C$ is the expected total number of events over both groups, n_E and n_C are numbers of participants in group E (experimental group) and group C (control group), respectively. p_E is the probability of failure in group E (experimental group) over the maximum time period of the study (t years). p_C is the probability of failure in group C (control group) over the maximum time period of the study (t years).

Suppose we want to compare the survival curves between an experimental group (E) and a control group (C) in a clinical trial with a maximum follow-up of t years. The Cox proportional hazards regression model is assumed to have the form:

$$h(t|X_1) = h_0(t)\exp(\beta_1 X_1).$$

Let n_E be the number of participants in the E group and n_C be the number of participants in the C group. We wish to test the hypothesis $H0 : RR = 1$ versus $H1 : RR$ not equal to 1, where $RR = \exp(\beta_1)$ =underlying hazard ratio for the E group versus the C group. Let RR be the postulated hazard ratio, α be the significance level. Assume that the test is a two-sided test. If the ratio of participants in group E compared to group C $= n_E/n_C = k$, then the power of the test is

$$power = \Phi(\sqrt{k * m} * |RR - 1|/(k * RR + 1) - z_{1-\alpha/2}),$$

where

$$m = n_E p_E + n_C p_C,$$

and $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$-th percentile of the standard normal distribution $N(0,1)$, Φ is the cumulative distribution function (CDF) of $N(0,1)$.

p_C and p_E can be calculated from the following formulaes:

$$p_C = \sum_{i=1}^{t} D_i, p_E = \sum_{i=1}^{t} E_i,$$

where $D_i = \lambda_i A_i C_i$, $E_i = RR\lambda_i B_i C_i$, $A_i = \prod_{j=0}^{i-1}(1-\lambda_j)$, $B_i = \prod_{j=0}^{i-1}(1-RR\lambda_j)$, $C_i = \prod_{j=0}^{i-1}(1-\delta_j)$. And λ_i is the probability of failure at time `i` among participants in the control group, given that a participant has survived to time $i-1$ and is not censored at time $i-1$, i.e., the approximate hazard time i in the control group, $i = 1, ..., t$; $RRlambda_i$ is the probability of failure at time `i` among participants in the experimental group, given that a participant has survived to time $i-1$ and is not censored at time $i-1$, i.e., the approximate hazard time i in the experimental group, $i = 1, ..., t$; $delta$ is the prbability that a participant is censored at time i given that he was followed up to time i and has not failed, $i = 0, 1, ..., t$, which is assumed the same in each group.

Value

`mat.lambda` a matrix with 9 columns and `nTimes+1` rows, where `nTimes` is the number of observed time points for the control group in the data set. The 9 columns are (1) `time` - observed time point for the control group; (2) `lambda`; (3) `RRlambda`; (4)

delta; (5) A; (6) B; (7) C; (8) D; (9) E. Please refer to the Details section for the definitions of elements of these quantities. See also Table 14.24 on page 809 of Rosner (2006).

mat.event
: a matrix with 5 columns and nTimes+1 rows, where nTimes is the number of observed time points for control group in the data set. The 5 columns are (1) time - observed time point for the control group; (2) nEvent.C - number of events in the control group at each time point; (3) nCensored.C - number of censorings in the control group at each time point; (4) nSurvive.C - number of alived in the control group at each time point; (5) nRisk.C - number of participants at risk in the control group at each time point. Please refer to Table 14.12 on page 787 of Rosner (2006).

pC
: estimated probability of failure in group C (control group) over the maximum time period of the study (t years).

pE
: estimated probability of failure in group E (experimental group) over the maximum time period of the study (t years).

power
: the power of the test.

Note

(1) The estimates of $RRlambda_i = RR * \lambda_i$. That is, RRlambda is not directly estimated based on data from the experimental group; (2) The power formula assumes that the central-limit theorem is valid and hence is appropriate for large samples.

References

Freedman, L.S. (1982). Tables of the number of patients required in clinical trials using the log-rank test. *Statistics in Medicine*. 1: 121-129

Rosner B. (2006). *Fundamentals of Biostatistics*. (6-th edition). Thomson Brooks/Cole.

See Also

powerCT.default0, powerCT.default

Examples

```
 # Example 14.42 in Rosner B. Fundamentals of Biostatistics.
 # (6-th edition). (2006) page 809

 library(survival)

 data(Oph)
 res <- powerCT(formula = Surv(times, status) ~ group,
dat = Oph,
                nE = 200,
nC = 200,
RR = 0.7,
alpha = 0.05)

 # Table 14.24 on page 809 of Rosner (2006)
```

```
print(round(res$mat.lambda, 4))

# Table 14.12 on page 787 of Rosner (2006)
print(round(res$mat.event, 4))

# the power
print(round(res$power, 2))
```

---

powerCT.default *Power Calculation in the Analysis of Survival Data for Clinical Trials*

Description

Power calculation for the Comparison of Survival Curves Between Two Groups under the Cox Proportional-Hazards Model for clinical trials.

Usage

```
powerCT.default(nE,
nC,
pE,
pC,
RR,
alpha = 0.05)
```

Arguments

nE	integer. number of participants in the experimental group.
nC	integer. number of participants in the control group.
pE	numeric. probability of failure in group E (experimental group) over the maximum time period of the study (t years).
pC	numeric. probability of failure in group C (control group) over the maximum time period of the study (t years).
RR	numeric. postulated hazard ratio.
alpha	numeric. type I error rate.

Details

This is an implementation of the power calculation method described in Section 14.12 (page 807) of Rosner (2006). The method was proposed by Freedman (1982).

Suppose we want to compare the survival curves between an experimental group (E) and a control group (C) in a clinical trial with a maximum follow-up of t years. The Cox proportional hazards regression model is assumed to have the form:

$$h(t|X_1) = h_0(t)\exp(\beta_1 X_1).$$

Let n_E be the number of participants in the E group and n_C be the number of participants in the C group. We wish to test the hypothesis $H0 : RR = 1$ versus $H1 : RR$ not equal to 1, where $RR = \exp(\beta_1)$ =underlying hazard ratio for the E group versus the C group. Let RR be the postulated hazard ratio, α be the significance level. Assume that the test is a two-sided test. If the ratio of participants in group E compared to group C $= n_E/n_C = k$, then the power of the test is

$$power = \Phi(\sqrt{k * m} * |RR - 1|/(k * RR + 1) - z_{1-\alpha/2}),$$

where

$$m = n_E p_E + n_C p_C,$$

and $z_{1-\alpha/2}$ is the $100(1 - \alpha/2)$-th percentile of the standard normal distribution $N(0, 1)$, Φ is the cumulative distribution function (CDF) of $N(0, 1)$.

Value

The power of the test.

Note

The power formula assumes that the central-limit theorem is valid and hence is appropriate for large samples.

References

Freedman, L.S. (1982). Tables of the number of patients required in clinical trials using the log-rank test. *Statistics in Medicine*. 1: 121-129

Rosner B. (2006). *Fundamentals of Biostatistics*. (6-th edition). Thomson Brooks/Cole.

See Also

powerCT.default0, powerCT

Examples

```
# Example 14.42 in Rosner B. Fundamentals of Biostatistics.
# (6-th edition). (2006) page 809
powerCT.default(nE = 200,
nC = 200,
pE = 0.3707,
pC = 0.4890,
                RR = 0.7,
alpha = 0.05)
```

`powerCT.default0` *Power Calculation in the Analysis of Survival Data for Clinical Trials*

Description

Power calculation for the Comparison of Survival Curves Between Two Groups under the Cox Proportional-Hazards Model for clinical trials.

Usage

```
powerCT.default0(k,
 m,
 RR,
 alpha = 0.05)
```

Arguments

`k`	numeric. ratio of participants in group E (experimental group) compared to group C (control group).
`m`	integer. expected total number of events over both groups.
`RR`	numeric. postulated hazard ratio.
`alpha`	numeric. type I error rate.

Details

This is an implementation of the power calculation method described in Section 14.12 (page 807) of Rosner (2006). The method was proposed by Freedman (1982).

Suppose we want to compare the survival curves between an experimental group (E) and a control group (C) in a clinical trial with a maximum follow-up of t years. The Cox proportional hazards regression model is assumed to have the form:

$$h(t|X_1) = h_0(t)\exp(\beta_1 X_1).$$

Let n_E be the number of participants in the E group and n_C be the number of participants in the C group. We wish to test the hypothesis $H0 : RR = 1$ versus $H1 : RR$ not equal to 1, where $RR = \exp(\beta_1)$ =underlying hazard ratio for the E group versus the C group. Let RR be the postulated hazard ratio, α be the significance level. Assume that the test is a two-sided test. If the ratio of participants in group E compared to group C = $n_E/n_C = k$, then the power of the test is

$$power = \Phi(\sqrt{k*m} * |RR - 1|/(k*RR + 1) - z_{1-\alpha/2}),$$

where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$-th percentile of the standard normal distribution $N(0,1)$, Φ is the cumulative distribution function (CDF) of $N(0,1)$.

Value

The power of the test.

Note

The power formula assumes that the central-limit theorem is valid and hence is appropriate for large samples.

References

Freedman, L.S. (1982). Tables of the number of patients required in clinical trials using the log-rank test. *Statistics in Medicine*. 1: 121-129

Rosner B. (2006). *Fundamentals of Biostatistics*. (6-th edition). Thomson Brooks/Cole.

See Also

powerCT.default, powerCT

Examples

```
# Example 14.42 in Rosner B. Fundamentals of Biostatistics.
# (6-th edition). (2006) page 809
powerCT.default0(k = 1,
 m = 171.9,
 RR = 0.7,
 alpha = 0.05)
```

---

powerEpi *Power Calculation for Cox Proportional Hazards Regression with Two Covariates for Epidemiological Studies*

Description

Power calculation for Cox proportional hazards regression with two covariates for epidemiological Studies. The covariate of interest should be a binary variable. The other covariate can be either binary or non-binary. The formula takes into account competing risks and the correlation between the two covariates. Some parameters will be estimated based on a pilot data set.

Usage

```
powerEpi(X1, X2, failureFlag, n, theta, alpha = 0.05)
```

Arguments

X1	numeric. a nPilot by 1 vector, where nPilot is the number of subjects in the pilot data set. This vector records the values of the covariate of interest for the nPilot subjects in the pilot study. X1 should be binary and take only two possible values: zero and one.
X2	numeric. a nPilot by 1 vector, where nPilot is the number of subjects in the pilot study. This vector records the values of the second covariate for the nPilot subjects in the pilot study. X2 can be binary or non-binary.

failureFlag	numeric. a nPilot by 1 vector of indicators indicating if a subject is failure (failureFlag=1) or alive (failureFlag=0).
n	integer. total number of subjects
theta	numeric. postulated hazard ratio
alpha	numeric. type I error rate.

Details

This is an implementation of the power calculation formula derived by Latouche et al. (2004) for the following Cox proportional hazards regression in the epidemiological studies:

$$h(t|x_1, x_2) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2),$$

where the covariate X_1 is of our interest. The covariate X_1 should be a binary variable taking two possible values: zero and one, while the covariate X_2 can be binary or continuous.

Suppose we want to check if the hazard of $X_1 = 1$ is equal to the hazard of $X_1 = 0$ or not. Equivalently, we want to check if the hazard ratio of $X_1 = 1$ to $X_1 = 0$ is equal to 1 or is equal to $\exp(\beta_1) = \theta$. Given the type I error rate α for a two-sided test, the power required to detect a hazard ratio as small as $\exp(\beta_1) = \theta$ is

$$power = \Phi\left(-z_{1-\alpha/2} + \sqrt{n[\log(\theta)]^2 p(1-p)\psi(1-\rho^2)}\right),$$

where z_a is the $100a$-th percentile of the standard normal distribution, ψ is the proportion of subjects died of the disease of interest, and

$$\rho = corr(X_1, X_2) = (p_1 - p_0) \times \sqrt{\frac{q(1-q)}{p(1-p)}},$$

and $p = Pr(X_1 = 1)$, $q = Pr(X_2 = 1)$, $p_0 = Pr(X_1 = 1|X_2 = 0)$, and $p_1 = Pr(X_1 = 1|X_2 = 1)$.

p, ρ^2, and ψ will be estimated from a pilot data set.

Value

power	the power of the test.
p	proportion of subjects taking $X_1 = 1$.
rho2	square of the correlation between X_1 and X_2.
psi	proportion of subjects died of the disease of interest.

Note

(1) The formula can be used to calculate power for a randomized trial study by setting rho2=0.

(2) When $\rho^2 = 0$, the formula derived by Latouche et al. (2004) looks the same as that derived by Schoenfeld (1983). Latouche et al. (2004) pointed out that in this situation, the interpretations are different hence the two formulae are actually different. In Latouched et al. (2004), the hazard ratio θ measures the difference of effect of a covariate at two different levels on the subdistribution hazard for a particular failure, while in Schoenfeld (1983), the hazard ratio θ measures the difference of effect on the cause-specific hazard.

References

Schoenfeld DA. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics*. 39:499-503.

Latouche A., Porcher R. and Chevret S. (2004). Sample size formula for proportional hazards modelling of competing risks. *Statistics in Medicine*. 23:3263-3274.

See Also

powerEpi.default

Examples

```
# generate a toy pilot data set
X1 <- c(rep(1, 39), rep(0, 61))
set.seed(123456)
X2 <- sample(c(0, 1), 100, replace = TRUE)
failureFlag <- sample(c(0, 1), 100, prob = c(0.5, 0.5), replace = TRUE)

powerEpi(X1 = X1, X2 = X2, failureFlag = failureFlag,
  n = 139, theta = 2, alpha = 0.05)
```

---

powerEpi.default — *Power Calculation for Cox Proportional Hazards Regression with Two Covariates for Epidemiological Studies*

Description

Power calculation for Cox proportional hazards regression with two covariates for epidemiological Studies. The covariate of interest should be a binary variable. The other covariate can be either binary or non-binary. The formula takes into account competing risks and the correlation between the two covariates.

Usage

```
powerEpi.default(n,
 theta,
 p,
 psi,
 rho2,
 alpha = 0.05)
```

Arguments

n	integer. total number of subjects
theta	numeric. postulated hazard ratio
p	numeric. proportion of subjects taking the value one for the covariate of interest.

`psi`	numeric. proportion of subjects died of the disease of interest.
`rho2`	numeric. square of the correlation between the covariate of interest and the other covariate.
`alpha`	numeric. type I error rate.

Details

This is an implementation of the power calculation formula derived by Latouche et al. (2004) for the following Cox proportional hazards regression in the epidemiological studies:

$$h(t|x_1, x_2) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2),$$

where the covariate X_1 is of our interest. The covariate X_1 should be a binary variable taking two possible values: zero and one, while the covariate X_2 can be binary or continuous.

Suppose we want to check if the hazard of $X_1 = 1$ is equal to the hazard of $X_1 = 0$ or not. Equivalently, we want to check if the hazard ratio of $X_1 = 1$ to $X_1 = 0$ is equal to 1 or is equal to $\exp(\beta_1) = \theta$. Given the type I error rate α for a two-sided test, the power required to detect a hazard ratio as small as $\exp(\beta_1) = \theta$ is

$$power = \Phi\left(-z_{1-\alpha/2} + \sqrt{n[\log(\theta)]^2 p(1-p)\psi(1-\rho^2)}\right),$$

where z_a is the $100a$-th percentile of the standard normal distribution, ψ is the proportion of subjects died of the disease of interest, and

$$\rho = corr(X_1, X_2) = (p_1 - p_0) \times \sqrt{\frac{q(1-q)}{p(1-p)}},$$

and $p = Pr(X_1 = 1)$, $q = Pr(X_2 = 1)$, $p_0 = Pr(X_1 = 1|X_2 = 0)$, and $p_1 = Pr(X_1 = 1|X_2 = 1)$.

Value

The power of the test.

Note

(1) The formula can be used to calculate power for a randomized trial study by setting `rho2=0`.

(2) When `rho2=0`, the formula derived by Latouche et al. (2004) looks the same as that derived by Schoenfeld (1983). Latouche et al. (2004) pointed out that in this situation, the interpretations are different hence the two formulae are actually different. In Latouched et al. (2004), the hazard ratio θ measures the difference of effect of a covariate at two different levels on the subdistribution hazard for a particular failure, while in Schoenfeld (1983), the hazard ratio θ measures the difference of effect on the cause-specific hazard.

References

Schoenfeld DA. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics*. 39:499-503.

Latouche A., Porcher R. and Chevret S. (2004). Sample size formula for proportional hazards modelling of competing risks. *Statistics in Medicine*. 23:3263-3274.

See Also

powerEpi

Examples

```
# Example at the end of Section 5.2 of Latouche et al. (2004)
# for a cohort study.
powerEpi.default(n = 139,
 theta = 2,
 p = 0.39,
 psi = 0.505,
                rho2 = 0.132^2,
 alpha = 0.05)
```

---

powerEpiCont — *Power Calculation for Cox Proportional Hazards Regression with Nonbinary Covariates for Epidemiological Studies*

Description

Power calculation for Cox proportional hazards regression with nonbinary covariates for Epidemiological Studies. Some parameters will be estimated based on a pilot data set.

Usage

```
powerEpiCont(formula,
     dat,
     var.X1,
     var.failureFlag,
     n,
     theta,
     alpha = 0.05)
```

Arguments

`formula`	a formula object relating the covariate of interest to other covariates to calculate the multiple correlation coefficient. The variables in formula must be in the data frame `dat`.
`dat`	a `nPilot` by `p` data frame representing the pilot data set, where `nPilot` is the number of subjects in the pilot study and the `p` (> 1) columns contains the covariate of interest and other covariates.
`var.X1`	character. name of the column in the data frame `dat`, indicating the covariate of interest.
`var.failureFlag`	character. name of the column in the data frame `dat`, indicating if a subject is failure (taking value 1) or alive (taking value 0).

n	integer. total number of subjects.
theta	numeric. postulated hazard ratio.
alpha	numeric. type I error rate.

Details

This is an implementation of the power calculation formula derived by Hsieh and Lavori (2000) for the following Cox proportional hazards regression in the epidemiological studies:

$$h(t|x_1, \boldsymbol{x}_2) = h_0(t)\exp(\beta_1 x_1 + \boldsymbol{\beta}_2 \boldsymbol{x}_2),$$

where the covariate X_1 is a nonbinary variable and $\boldsymbol{X}_2$ is a vector of other covariates.

Suppose we want to check if the hazard ratio of the main effect $X_1 = 1$ to $X_1 = 0$ is equal to 1 or is equal to $\exp(\beta_1) = \theta$. Given the type I error rate α for a two-sided test, the power required to detect a hazard ratio as small as $\exp(\beta_1) = \theta$ is

$$power = \Phi\left(-z_{1-\alpha/2} + \sqrt{n[\log(\theta)]^2\sigma^2\psi(1-\rho^2)}\right),$$

where z_a is the $100a$-th percentile of the standard normal distribution, $\sigma^2 = Var(X_1)$, ψ is the proportion of subjects died of the disease of interest, and ρ is the multiple correlation coefficient of the following linear regression:

$$x_1 = b_0 + \boldsymbol{b}^T \boldsymbol{x}_2.$$

That is, $\rho^2 = R^2$, where R^2 is the proportion of variance explained by the regression of X_1 on the vector of covriates $\boldsymbol{X}_2$.

rho will be estimated from a pilot study.

Value

power	The power of the test.
rho2	square of the correlation between X_1 and X_2.
sigma2	variance of the covariate of interest.
psi	proportion of subjects died of the disease of interest.

Note

(1) Hsieh and Lavori (2000) assumed one-sided test, while this implementation assumed two-sided test. (2) The formula can be used to calculate power for a randomized trial study by setting `rho2=0`.

References

Hsieh F.Y. and Lavori P.W. (2000). Sample-size calculation for the Cox proportional hazards regression model with nonbinary covariates. *Controlled Clinical Trials*. 21:552-560.

See Also

powerEpiCont.default

Examples

```
# generate a toy pilot data set
set.seed(123456)
X1 <- rnorm(100, mean = 0, sd = 0.3126)
X2 <- sample(c(0, 1), 100, replace = TRUE)
failureFlag <- sample(c(0, 1), 100, prob = c(0.25, 0.75), replace = TRUE)
dat <- data.frame(X1 = X1, X2 = X2, failureFlag = failureFlag)

powerEpiCont(formula = X1 ~ X2,
     dat = dat,
     var.X1 = "X1",
     var.failureFlag = "failureFlag",
             n = 107,
     theta = exp(1),
     alpha = 0.05)
```

---

powerEpiCont.default *Power Calculation for Cox Proportional Hazards Regression with Nonbinary Covariates for Epidemiological Studies*

Description

Power calculation for Cox proportional hazards regression with nonbinary covariates for Epidemiological Studies.

Usage

```
powerEpiCont.default(n,
     theta,
     sigma2,
     psi,
     rho2,
     alpha = 0.05)
```

Arguments

`n`	integer. total number of subjects.
`theta`	numeric. postulated hazard ratio.
`sigma2`	numeric. variance of the covariate of interest.
`psi`	numeric. proportion of subjects died of the disease of interest.
`rho2`	numeric. square of the multiple correlation coefficient between the covariate of interest and other covariates.
`alpha`	numeric. type I error rate.

Details

This is an implementation of the power calculation formula derived by Hsieh and Lavori (2000) for the following Cox proportional hazards regression in the epidemiological studies:

$$h(t|x_1, \boldsymbol{x}_2) = h_0(t) \exp(\beta_1 x_1 + \boldsymbol{\beta}_2 \boldsymbol{x}_2),$$

where the covariate X_1 is a nonbinary variable and $\boldsymbol{X}_2$ is a vector of other covariates.

Suppose we want to check if the hazard ratio of the main effect $X_1 = 1$ to $X_1 = 0$ is equal to 1 or is equal to $\exp(\beta_1) = \theta$. Given the type I error rate α for a two-sided test, the power required to detect a hazard ratio as small as $\exp(\beta_1) = \theta$ is

$$power = \Phi\left(-z_{1-\alpha/2} + \sqrt{n[\log(\theta)]^2\sigma^2\psi(1-\rho^2)}\right),$$

where z_a is the $100a$-th percentile of the standard normal distribution, $\sigma^2 = Var(X_1)$, ψ is the proportion of subjects died of the disease of interest, and ρ is the multiple correlation coefficient of the following linear regression:

$$x_1 = b_0 + \boldsymbol{b}^T \boldsymbol{x}_2.$$

That is, $\rho^2 = R^2$, where R^2 is the proportion of variance explained by the regression of X_1 on the vector of covriates $\boldsymbol{X}_2$.

Value

The power of the test.

Note

(1) Hsieh and Lavori (2000) assumed one-sided test, while this implementation assumed two-sided test. (2) The formula can be used to calculate power for a randomized trial study by setting `rho2=0`.

References

Hsieh F.Y. and Lavori P.W. (2000). Sample-size calculation for the Cox proportional hazards regression model with nonbinary covariates. *Controlled Clinical Trials*. 21:552-560.

See Also

powerEpiCont

Examples

```
# example in the EXAMPLE section (page 557) of Hsieh and Lavori (2000).
# Hsieh and Lavori (2000) assumed one-sided test,
# while this implementation assumed two-sided test.
# Hence alpha=0.1 here (two-sided test) will correspond
# to alpha=0.05 of one-sided test in Hsieh and Lavori's (2000) example.
powerEpiCont.default(n = 107,
     theta = exp(1),
     sigma2 = 0.3126^2,
                     psi = 0.738,
```

```
rho2 = 0.1837,
alpha = 0.1)
```

---

powerEpiInt *Power Calculation Testing Interaction Effect for Cox Proportional Hazards Regression with two covariates for Epidemiological Studies (Both covariates should be binary)*

Description

Power calculation testing interaction effect for Cox proportional hazards regression with two covariates for Epidemiological Studies. Both covariates should be binary variables. The formula takes into account the correlation between the two covariates. Some parameters will be estimated based on a pilot study.

Usage

```
powerEpiInt(X1,
    X2,
    failureFlag,
    n,
    theta,
    alpha = 0.05)
```

Arguments

X1	numeric. a nPilot by 1 vector, where nPilot is the number of subjects in the pilot data set. This vector records the values of the covariate of interest for the nPilot subjects in the pilot study. X1 should be binary and take only two possible values: zero and one.
X2	numeric. a nPilot by 1 vector, where nPilot is the number of subjects in the pilot study. This vector records the values of the second covariate for the nPilot subjects in the pilot study. X2 should be binary and take only two possible values: zero and one.
failureFlag	numeric.a nPilot by 1 vector of indicators indicating if a subject is failure (failureFlag=1) or alive (failureFlag=0).
n	integer. total number of subjects.
theta	numeric. postulated hazard ratio.
alpha	numeric. type I error rate.

Details

This is an implementation of the power calculation formula derived by Schmoor et al. (2000) for the following Cox proportional hazards regression in the epidemoilogical studies:

$$h(t|x_1, x_2) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \gamma(x_1 x_2)),$$

where both covariates X_1 and X_2 are binary variables.

Suppose we want to check if the hazard ratio of the interaction effect $X_1 X_2 = 1$ to $X_1 X_2 = 0$ is equal to 1 or is equal to $\exp(\gamma) = \theta$. Given the type I error rate α for a two-sided test, the power required to detect a hazard ratio as small as $\exp(\gamma) = \theta$ is:

$$power = \Phi\left(-z_{1-\alpha/2} + \sqrt{\frac{n}{\delta}[\log(\theta)]^2\psi}\right),$$

where z_a is the $100a$-th percentile of the standard normal distribution,

$$\delta = \frac{1}{p_{00}} + \frac{1}{p_{01}} + \frac{1}{p_{10}} + \frac{1}{p_{11}},$$

ψ is the proportion of subjects died of the disease of interest, and $p_{00} = Pr(X_1 = 0, \text{and}, X_2 = 0)$, $p_{01} = Pr(X_1 = 0, \text{and}, X_2 = 1)$, $p_{10} = Pr(X_1 = 1, \text{and}, X_2 = 0)$, $p_{11} = Pr(X_1 = 1, \text{and}, X_2 = 1)$.

$p_{00}, p_{01}, p_{10}, p_{11}$, and ψ will be estimated from the pilot data.

Value

`power`	the power of the test.
`p`	estimated $Pr(X_1 = 1)$
`q`	estimated $Pr(X_2 = 1)$
`p0`	estimated $Pr(X_1 = 1\|X_2 = 0)$
`p1`	estimated $Pr(X_1 = 1\|X_2 = 1)$
`rho2`	square of the estimated $corr(X_1, X_2)$
`G`	a factor adjusting the sample size. The sample size needed to detect an effect of a prognostic factor with given error probabilities has to be multiplied by the factor `G` when an interaction of the same magnitude is to be detected.
`mya`	estimated number of subjects taking values $X_1 = 0$ and $X_2 = 0$.
`myb`	estimated number of subjects taking values $X_1 = 0$ and $X_2 = 1$.
`myc`	estimated number of subjects taking values $X_1 = 1$ and $X_2 = 0$.
`myd`	estimated number of subjects taking values $X_1 = 1$ and $X_2 = 1$.
`psi`	proportion of subjects died of the disease of interest.

References

Schmoor C., Sauerbrei W., and Schumacher M. (2000). Sample size considerations for the evaluation of prognostic factors in survival analysis. *Statistics in Medicine*. 19:441-452.

See Also

powerEpiInt.default0, powerEpiInt2

Examples

```
# generate a toy pilot data set
X1 <- c(rep(1, 39), rep(0, 61))
set.seed(123456)
X2 <- sample(c(0, 1), 100, replace = TRUE)
failureFlag <- sample(c(0, 1), 100, prob = c(0.25, 0.75), replace = TRUE)

powerEpiInt(X1 = X1,
    X2 = X2,
    failureFlag = failureFlag,
    n = 184,
    theta = 3,
    alpha = 0.05)
```

`powerEpiInt.default0` *Power Calculation Testing Interaction Effect for Cox Proportional Hazards Regression*

Description

Power calculation testing interaction effect for Cox proportional hazards regression with two covariates for Epidemiological Studies. Both covariates should be binary variables. The formula takes into account the correlation between the two covariates.

Usage

```
powerEpiInt.default0(n,
     theta,
     p,
     psi,
     G,
     rho2,
     alpha = 0.05)
```

Arguments

n	integer. total number of subjects.
theta	numeric. postulated hazard ratio.
p	numeric. proportion of subjects taking the value one for the covariate of interest.
psi	numeric. proportion of subjects died of the disease of interest.

G	numeric. a factor adjusting the sample size. The sample size needed to detect an effect of a prognostic factor with given error probabilities has to be multiplied by the factor G when an interaction of the same magnitude is to be detected.
rho2	numeric. square of the correlation between the covariate of interest and the other covariate.
alpha	numeric. type I error rate.

Details

This is an implementation of the power calculation formula derived by Schmoor et al. (2000) for the following Cox proportional hazards regression in the epidemiological studies:

$$h(t|x_1, x_2) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \gamma(x_1 x_2)),$$

where both covariates X_1 and X_2 are binary variables.

Suppose we want to check if the hazard ratio of the interaction effect $X_1 X_2 = 1$ to $X_1 X_2 = 0$ is equal to 1 or is equal to $\exp(\gamma) = \theta$. Given the type I error rate α for a two-sided test, the power required to detect a hazard ratio as small as $\exp(\gamma) = \theta$ is

$$power = \Phi\left(-z_{1-\alpha/2} + \sqrt{\frac{n}{G}[\log(\theta)]^2 p(1-p)\psi(1-\rho^2)}\right),$$

where z_a is the $100a$-th percentile of the standard normal distribution, ψ is the proportion of subjects died of the disease of interest, and

$$\rho = corr(X_1, X_2) = (p_1 - p_0) \times \sqrt{\frac{q(1-q)}{p(1-p)}},$$

and $p = Pr(X_1 = 1)$, $q = Pr(X_2 = 1)$, $p_0 = Pr(X_1 = 1|X_2 = 0)$, and $p_1 = Pr(X_1 = 1|X_2 = 1)$, and

$$G = \frac{[(1-q)(1-p_0)p_0 + q(1-p_1)p_1]^2}{(1-q)q(1-p_0)p_0(1-p_1)p_1}.$$

If X_1 and X_2 are uncorrelated, we have $p_0 = p_1 = p$ leading to $1/[(1-q)q]$. For $q = 0.5$, we have $G = 4$.

Value

The power of the test.

References

Schmoor C., Sauerbrei W., and Schumacher M. (2000). Sample size considerations for the evaluation of prognostic factors in survival analysis. *Statistics in Medicine*. 19:441-452.

See Also

powerEpiInt.default1, powerEpiInt2

Examples

```
# Example at the end of Section 4 of Schmoor et al. (2000).
powerEpiInt.default0(n = 184,
     theta = 3,
     p = 0.61,
     psi = 139 / 184,
                     G = 4.79177,
     rho2 = 0.015^2,
     alpha = 0.05)
```

`powerEpiInt.default1` *Power Calculation Testing Interaction Effect for Cox Proportional Hazards Regression*

Description

Power calculation testing interaction effect for Cox proportional hazards regression with two covariates for Epidemiological Studies. Both covariates should be binary variables. The formula takes into account the correlation between the two covariates.

Usage

```
powerEpiInt.default1(n,
     theta,
     psi,
     p00,
     p01,
     p10,
     p11,
     alpha = 0.05)
```

Arguments

`n`	integer. total number of subjects.
`theta`	numeric. postulated hazard ratio.
`psi`	numeric. proportion of subjects died of the disease of interest.
`p00`	numeric. proportion of subjects taking values $X_1 = 0$ and $X_2 = 0$, i.e., $p_{00} = Pr(X_1 = 0, \text{and}, X_2 = 0)$.
`p01`	numeric. proportion of subjects taking values $X_1 = 0$ and $X_2 = 1$, i.e., $p_{01} = Pr(X_1 = 0, \text{and}, X_2 = 1)$.
`p10`	numeric. proportion of subjects taking values $X_1 = 1$ and $X_2 = 0$, i.e., $p_{10} = Pr(X_1 = 1, \text{and}, X_2 = 0)$.
`p11`	numeric. proportion of subjects taking values $X_1 = 1$ and $X_2 = 1$, i.e., $p_{11} = Pr(X_1 = 1, \text{and}, X_2 = 1)$.
`alpha`	numeric. type I error rate.

Details

This is an implementation of the power calculation formula derived by Schmoor et al. (2000) for the following Cox proportional hazards regression in the epidemoilogical studies:

$$h(t|x_1, x_2) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \gamma(x_1 x_2)),$$

where both covariates X_1 and X_2 are binary variables.

Suppose we want to check if the hazard ratio of the interaction effect $X_1 X_2 = 1$ to $X_1 X_2 = 0$ is equal to 1 or is equal to $\exp(\gamma) = \theta$. Given the type I error rate α for a two-sided test, the power required to detect a hazard ratio as small as $\exp(\gamma) = \theta$ is:

$$power = \Phi\left(-z_{1-\alpha/2} + \sqrt{\frac{n}{\delta}[\log(\theta)]^2 \psi}\right),$$

where z_a is the $100a$-th percentile of the standard normal distribution,

$$\delta = \frac{1}{p_{00}} + \frac{1}{p_{01}} + \frac{1}{p_{10}} + \frac{1}{p_{11}},$$

ψ is the proportion of subjects died of the disease of interest, and $p_{00} = Pr(X_1 = 0, \text{and}, X_2 = 0)$, $p_{01} = Pr(X_1 = 0, \text{and}, X_2 = 1)$, $p_{10} = Pr(X_1 = 1, \text{and}, X_2 = 0)$, $p_{11} = Pr(X_1 = 1, \text{and}, X_2 = 1)$.

Value

The power of the test.

References

Schmoor C., Sauerbrei W., and Schumacher M. (2000). Sample size considerations for the evaluation of prognostic factors in survival analysis. *Statistics in Medicine*. 19:441-452.

See Also

powerEpiInt.default0, powerEpiInt2

Examples

```
# Example at the end of Section 4 of Schmoor et al. (2000).
# p00, p01, p10, and p11 are calculated based on Table III on page 448
# of Schmoor et al. (2000).
powerEpiInt.default1(n = 184,
     theta = 3,
     psi = 139 / 184,
                     p00 = 50 / 184,
     p01 = 21 / 184,
     p10 = 78 / 184,
     p11 = 35 / 184,
                     alpha = 0.05)
```

powerEpiInt2 *Power Calculation Testing Interaction Effect for Cox Proportional Hazards Regression*

Description

Power calculation testing interaction effect for Cox proportional hazards regression with two covariates for Epidemiological Studies. Both covariates should be binary variables. The formula takes into account the correlation between the two covariates.

Usage

```
powerEpiInt2(n,
     theta,
     psi,
     mya,
     myb,
     myc,
     myd,
     alpha = 0.05)
```

Arguments

n	integer. total number of subjects.
theta	numeric. postulated hazard ratio.
psi	numeric. proportion of subjects died of the disease of interest.
mya	integer. number of subjects taking values $X_1 = 0$ and $X_2 = 0$ obtained from a pilot study.
myb	integer. number of subjects taking values $X_1 = 0$ and $X_2 = 1$ obtained from a pilot study.
myc	integer. number of subjects taking values $X_1 = 1$ and $X_2 = 0$ obtained from a pilot study.
myd	integer. number of subjects taking values $X_1 = 1$ and $X_2 = 1$ obtained from a pilot study.
alpha	numeric. type I error rate.

Details

This is an implementation of the power calculation formula derived by Schmoor et al. (2000) for the following Cox proportional hazards regression in the epidemiological studies:

$$h(t|x_1, x_2) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \gamma(x_1 x_2)),$$

where both covariates X_1 and X_2 are binary variables.

Suppose we want to check if the hazard ratio of the interaction effect $X_1X_2 = 1$ to $X_1X_2 = 0$ is equal to 1 or is equal to $\exp(\gamma) = \theta$. Given the type I error rate α for a two-sided test, the power required to detect a hazard ratio as small as $\exp(\gamma) = \theta$ is

$$power = \Phi\left(-z_{1-\alpha/2} + \sqrt{\frac{n}{G}[\log(\theta)]^2 p(1-p)\psi(1-\rho^2)}\right),$$

where z_a is the $100a$-th percentile of the standard normal distribution, ψ is the proportion of subjects died of the disease of interest, and

$$\rho = corr(X_1, X_2) = (p_1 - p_0) \times \sqrt{\frac{q(1-q)}{p(1-p)}},$$

and $p = Pr(X_1 = 1)$, $q = Pr(X_2 = 1)$, $p_0 = Pr(X_1 = 1|X_2 = 0)$, and $p_1 = Pr(X_1 = 1|X_2 = 1)$, and

$$G = \frac{[(1-q)(1-p_0)p_0 + q(1-p_1)p_1]^2}{(1-q)q(1-p_0)p_0(1-p_1)p_1},$$

and $p0 = Pr(X_1 = 1|X_2 = 0) = myc/(mya + myc)$, $p1 = Pr(X_1 = 1|X_2 = 1) = myd/(myb + myd)$, $p = Pr(X_1 = 1) = (myc + myd)/n_{obs}$, $q = Pr(X_2 = 1) = (myb + myd)/n_{obs}$, $n_{obs} = mya + myb + myc + myd$.

$p_{00} = Pr(X_1 = 0, \text{and}, X_2 = 0)$, $p_{01} = Pr(X_1 = 0, \text{and}, X_2 = 1)$, $p_{10} = Pr(X_1 = 1, \text{and}, X_2 = 0)$, $p_{11} = Pr(X_1 = 1, \text{and}, X_2 = 1)$.

Value

The power of the test.

References

Schmoor C., Sauerbrei W., and Schumacher M. (2000). Sample size considerations for the evaluation of prognostic factors in survival analysis. *Statistics in Medicine*. 19:441-452.

See Also

powerEpiInt.default0, powerEpiInt.default1

Examples

```
# Example at the end of Section 4 of Schmoor et al. (2000).
# mya, myb, myc, and myd are obtained from Table III on page 448
# of Schmoor et al. (2000).
powerEpiInt2(n = 184,
     theta = 3,
     psi = 139 / 184,
             mya = 50,
     myb = 21,
     myc = 78,
     myd = 35,
     alpha = 0.05)
```

ssize.stratify *Sample size calculation for Survival Analysis with Binary Predictor and Exponential Survival Function*

Description

Sample size calculation for survival analysis with binary predictor and exponential survival function.

Usage

```
ssize.stratify(
    power,
    timeUnit,
    gVec,
    PVec,
    HR,
    lambda0Vec,
    alpha = 0.05,
    verbose = TRUE)
```

Arguments

power	numeric. Power of the test.
timeUnit	numeric. Total study length.
gVec	numeric. m by 1 vector. The s-th element is the proportion of the total sample size for the s-th stratum, where m is the number of strata.
PVec	numeric. m by 1 vector. The s-th element is the proportion of subjects in treatment group 1 for the s-th stratum, where m is the number of strata.
HR	numeric. Hazard ratio (Ratio of the hazard for treatment group 1 to the hazard for treatment group 0, i.e. reference group).
lambda0Vec	numeric. m by 1 vector. The s-th element is the hazard for treatment group 0 (i.e., reference group) in the s-th stratum.
alpha	numeric. Type I error rate.
verbose	Logical. Indicating if intermediate results will be output or not.

Details

We assume (1) there is only one predictor and no covariates in the survival model (exponential survival function); (2) there are `m` strata; (3) the predictor `x` is a binary variable indicating treatment group 1 ($x = 1$) or treatment group 0 ($x = 0$); (3) the treatment effect is constant over time (proportional hazards); (4) the hazard ratio is the same in all strata, and (5) the data will be analyzed by the stratified log rank test.

The sample size formula is Formula (1) on page 801 of Palta M and Amini SB (1985):

$$n = (Z_\alpha + Z_\beta)^2/\mu^2$$

where α is the Type I error rate, β is the Type II error rate (power= $1-\beta$), Z_α is the $100(1-\alpha)$-th percentile of standard normal distribution, and

$$\mu = \log(\delta)\sqrt{\sum_{s=1}^{m} g_s P_s (1-P_s) V_s}$$

and

$$V_s = P_s\left[1 - \frac{1}{\lambda_{1s}}\left\{\exp\left[-\lambda_{1s}(T-1)\right] - \exp(-\lambda_{1s}T)\right\}\right] + (1-P_s)\left[1 - \frac{1}{\lambda_{0s}}\left\{\exp\left[-\lambda_{0s}(T-1)\right] - \exp(-\lambda_{0s}T\right\}\right]$$

In the above formulas, m is the number of strata, T is the total study length, δ is the hazard ratio, g_s is the proportion of the total sample size in stratum s, P_s is the proportion of stratum s, which is in treatment group 1, and λ_{is} is the hazard for the i-th treatment group in stratum s.

Value

The sample size.

References

Palta M and Amini SB. (1985). Consideration of covariates and stratification in sample size determination for survival time studies. *Journal of Chronic Diseases*. 38(9):801-809.

See Also

power.stratify

Examples

```
# example on page 803 of Palta M and Amini SB. (1985).
n <- ssize.stratify(
  power = 0.9,
  timeUnit = 1.25,
  gVec = c(0.5, 0.5),
  PVec = c(0.5, 0.5),
  HR = 1 / 1.91,
  lambda0Vec = c(2.303, 1.139),
  alpha = 0.05,
  verbose = TRUE
)
```

ssizeCT — *Sample Size Calculation in the Analysis of Survival Data for Clinical Trials*

Description

Sample size calculation for the Comparison of Survival Curves Between Two Groups under the Cox Proportional-Hazards Model for clinical trials. Some parameters will be estimated based on a pilot data set.

Usage

```
ssizeCT(formula,
dat,
power,
k,
RR,
alpha = 0.05)
```

Arguments

`formula`	A formula object, e.g. `Surv(time,status) ~ x`, where `time` is a vector of survival/censoring time, `status` is a vector of censoring indicator, `x` is the group indicator, which is a factor object in R and takes only two possible values (`C` for control group and `E` for experimental group). See also the documentation of the function `survfit` in the library `survival`.
`dat`	a data frame representing the pilot data set and containing at least 3 columns: (1) survival/censoring time; (2) censoring indicator; (3) group indicator which is a factor object in R and takes only two possible values (`C` for control group and `E` for experimental group).
`power`	numeric. power to detect the magnitude of the hazard ratio as small as that specified by `RR`.
`k`	numeric. ratio of participants in group E (experimental group) compared to group C (control group).
`RR`	numeric. postulated hazard ratio.
`alpha`	numeric. type I error rate.

Details

This is an implementation of the sample size calculation method described in Section 14.12 (page 807) of Rosner (2006). The method was proposed by Freedman (1982).

The movitation of this function is that some times we do not have information about m or p_E and p_C available, but we have a pilot data set that can be used to estimate p_E and p_C hence m, where $m = n_E p_E + n_C p_C$ is the expected total number of events over both groups, n_E and n_C are numbers of participants in group E (experimental group) and group C (control group), respectively.

p_E is the probability of failure in group E (experimental group) over the maximum time period of the study (t years). p_C is the probability of failure in group C (control group) over the maximum time period of the study (t years).

Suppose we want to compare the survival curves between an experimental group (E) and a control group (C) in a clinical trial with a maximum follow-up of t years. The Cox proportional hazards regression model is assumed to have the form:

$$h(t|X_1) = h_0(t)\exp(\beta_1 X_1).$$

Let n_E be the number of participants in the E group and n_C be the number of participants in the C group. We wish to test the hypothesis $H0 : RR = 1$ versus $H1 : RR$ not equal to 1, where $RR = \exp(\beta_1)$ =underlying hazard ratio for the E group versus the C group. Let RR be the postulated hazard ratio, α be the significance level. Assume that the test is a two-sided test. If the ratio of participants in group E compared to group C = $n_E/n_C = k$, then the number of participants needed in each group to achieve a power of $1 - \beta$ is

$$n_E = \frac{mk}{kp_E + p_C}, n_C = \frac{m}{kp_E + p_C}$$

where

$$m = \frac{1}{k}\left(\frac{kRR + 1}{RR - 1}\right)^2 \left(z_{1-\alpha/2} + z_{1-\beta}\right)^2,$$

and $z_{1-\alpha/2}$ is the $100(1 - \alpha/2)$-th percentile of the standard normal distribution $N(0, 1)$.

p_C and p_E can be calculated from the following formulaes:

$$p_C = \sum_{i=1}^{t} D_i, p_E = \sum_{i=1}^{t} E_i,$$

where $D_i = \lambda_i A_i C_i$, $E_i = RR\lambda_i B_i C_i$, $A_i = \prod_{j=0}^{i-1}(1 - \lambda_j)$, $B_i = \prod_{j=0}^{i-1}(1 - RR\lambda_j)$, $C_i = \prod_{j=0}^{i-1}(1 - \delta_j)$. And λ_i is the probability of failure at time `i` among participants in the control group, given that a participant has survived to time $i - 1$ and is not censored at time $i - 1$, i.e., the approximate hazard time i in the control group, $i = 1, ..., t$; $RRlambda_i$ is the probability of failure at time `i` among participants in the experimental group, given that a participant has survived to time $i-1$ and is not censored at time $i-1$, i.e., the approximate hazard time i in the experimental group, $i = 1, ..., t$; $delta$ is the prbability that a participant is censored at time i given that he was followed up to time i and has not failed, $i = 0, 1, ..., t$, which is assumed the same in each group.

Value

`mat.lambda` a matrix with 9 columns and `nTimes+1` rows, where `nTimes` is the number of observed time points for the control group in the data set. The 9 columns are (1) `time` - observed time point for the control group; (2) `lambda`; (3) `RRlambda`; (4) `delta`; (5) `A`; (6) `B`; (7) `C`; (8) `D`; (9) `E`. Please refer to the Details section for the definitions of elements of these quantities. See also Table 14.24 on page 809 of Rosner (2006).

`mat.event` a matrix with 5 columns and `nTimes+1` rows, where `nTimes` is the number of observed time points for control group in the data set. The 5 columns are (1) `time` - observed time point for the control group; (2) `nEvent.C` - number of events in the control group at each time point; (3) nCensored.C - number of censorings in the control group at each time point; (4) nSurvive.C - number of alived in the control group at each time point; (5) nRisk.C - number of participants at risk in the control group at each time point. Please refer to Table 14.12 on page 787 of Rosner (2006).

pC
: estimated probability of failure in group C (control group) over the maximum time period of the study (t years).

pE
: estimated probability of failure in group E (experimental group) over the maximum time period of the study (t years).

ssize
: a two-element vector. The first element is n_E and the second element is n_C.

Note

(1) The estimates of $RRlambda_i = RR * \lambda_i$. That is, RRlambda is not directly estimated based on data from the experimental group; (2) The sample size formula assumes that the central-limit theorem is valid and hence is appropriate for large samples. (3) n_E and n_C will be rounded up to integers.

References

Freedman, L.S. (1982). Tables of the number of patients required in clinical trials using the log-rank test. *Statistics in Medicine*. 1: 121-129

Rosner B. (2006). *Fundamentals of Biostatistics*. (6-th edition). Thomson Brooks/Cole.

See Also

ssizeCT.default

Examples

```
 # Example 14.42 in Rosner B. Fundamentals of Biostatistics.
 # (6-th edition). (2006) page 809

 library(survival)

 data(Oph)
 res <- ssizeCT(formula = Surv(times, status) ~ group,
dat = Oph,
                power = 0.8,
k = 1,
RR = 0.7,
alpha = 0.05)

 # Table 14.24 on page 809 of Rosner (2006)
 print(round(res$mat.lambda, 4))

 # Table 14.12 on page 787 of Rosner (2006)
 print(round(res$mat.event, 4))

 # the sample size
```

```
print(res$ssize)
```

`ssizeCT.default` *Sample Size Calculation in the Analysis of Survival Data for Clinical Trials*

Description

Sample size calculation for the Comparison of Survival Curves Between Two Groups under the Cox Proportional-Hazards Model for clinical trials.

Usage

```
ssizeCT.default(power,
k,
pE,
pC,
RR,
alpha = 0.05)
```

Arguments

`power`	numeric. power to detect the magnitude of the hazard ratio as small as that specified by `RR`.
`k`	numeric. ratio of participants in group E (experimental group) compared to group C (control group).
`pE`	numeric. probability of failure in group E (experimental group) over the maximum time period of the study (t years).
`pC`	numeric. probability of failure in group C (control group) over the maximum time period of the study (t years).
`RR`	numeric. postulated hazard ratio.
`alpha`	numeric. type I error rate.

Details

This is an implementation of the sample size calculation method described in Section 14.12 (page 807) of Rosner (2006). The method was proposed by Freedman (1982).

Suppose we want to compare the survival curves between an experimental group (E) and a control group (C) in a clinical trial with a maximum follow-up of t years. The Cox proportional hazards regression model is assumed to have the form:

$$h(t|X_1) = h_0(t)\exp(\beta_1 X_1).$$

Let n_E be the number of participants in the E group and n_C be the number of participants in the C group. We wish to test the hypothesis $H0 : RR = 1$ versus $H1 : RR$ not equal to 1, where

$RR = \exp(\beta_1)$ =underlying hazard ratio for the E group versus the C group. Let RR be the postulated hazard ratio, α be the significance level. Assume that the test is a two-sided test. If the ratio of participants in group E compared to group C $= n_E/n_C = k$, then the number of participants needed in each group to achieve a power of $1-\beta$ is

$$n_E = \frac{mk}{kp_E + p_C}, n_C = \frac{m}{kp_E + p_C}$$

where

$$m = \frac{1}{k}\left(\frac{kRR+1}{RR-1}\right)^2 \left(z_{1-\alpha/2} + z_{1-\beta}\right)^2,$$

and $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$-th percentile of the standard normal distribution $N(0,1)$.

Value

A two-element vector. The first element is n_E and the second element is n_C.

Note

(1) The sample size formula assumes that the central-limit theorem is valid and hence is appropriate for large samples. (2) n_E and n_C will be rounded up to integers.

References

Freedman, L.S. (1982). Tables of the number of patients required in clinical trials using the log-rank test. *Statistics in Medicine*. 1: 121-129

Rosner B. (2006). *Fundamentals of Biostatistics*. (6-th edition). Thomson Brooks/Cole.

See Also

ssizeCT

Examples

```
# Example 14.42 in Rosner B. Fundamentals of Biostatistics.
# (6-th edition). (2006) page 809
ssizeCT.default(power = 0.8,
k = 1,
pE = 0.3707,
pC = 0.4890,
                RR = 0.7,
alpha = 0.05)
```

ssizeEpi *Sample Size Calculation for Cox Proportional Hazards Regression*

Description

Sample size calculation for Cox proportional hazards regression with two covariates for Epidemiological Studies. The covariate of interest should be a binary variable. The other covariate can be either binary or non-binary. The formula takes into account competing risks and the correlation between the two covariates.

Usage

```
ssizeEpi(X1,
 X2,
 failureFlag,
 power,
 theta,
 alpha = 0.05)
```

Arguments

`X1`	numeric. a `nPilot` by 1 vector, where `nPilot` is the number of subjects in the pilot data set. This vector records the values of the covariate of interest for the `nPilot` subjects in the pilot study. `X1` should be binary and take only two possible values: zero and one.
`X2`	numeric. a `nPilot` by 1 vector, where `nPilot` is the number of subjects in the pilot study. This vector records the values of the second covariate for the `nPilot` subjects in the pilot study. `X2` can be binary or non-binary.
`failureFlag`	numeric. a `nPilot` by 1 vector of indicators indicating if a subject is failure (`failureFlag=1`) or alive (`failureFlag=0`).
`power`	numeric. postulated power.
`theta`	numeric. postulated hazard ratio.
`alpha`	numeric. type I error rate.

Details

This is an implementation of the sample size formula derived by Latouche et al. (2004) for the following Cox proportional hazards regression in the epidemiological studies:

$$h(t|x_1, x_2) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2),$$

where the covariate X_1 is of our interest. The covariate X_1 has to be a binary variable taking two possible values: zero and one, while the covariate X_2 can be binary or continuous.

Suppose we want to check if the hazard of $X_1 = 1$ is equal to the hazard of $X_1 = 0$ or not. Equivalently, we want to check if the hazard ratio of $X_1 = 1$ to $X_1 = 0$ is equal to 1 or is equal to $\exp(\beta_1) = \theta$. Given the type I error rate α for a two-sided test, the total number of subjects required to achieve a power of $1 - \beta$ is

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{[\log(\theta)]^2 p(1-p)\psi(1-\rho^2)},$$

where z_a is the $100a$-th percentile of the standard normal distribution, ψ is the proportion of subjects died of the disease of interest, and

$$\rho = corr(X_1, X_2) = (p_1 - p_0) \times \sqrt{\frac{q(1-q)}{p(1-p)}},$$

and $p = Pr(X_1 = 1)$, $q = Pr(X_2 = 1)$, $p_0 = Pr(X_1 = 1|X_2 = 0)$, and $p_1 = Pr(X_1 = 1|X_2 = 1)$.

p, ρ^2, and ψ will be estimated from a pilot study.

Value

`n`	the total number of subjects required.
`p`	the proportion that X_1 takes value one.
`rho2`	square of the correlation between X_1 and X_2.
`psi`	proportion of subjects died of the disease of interest.

Note

(1) The calculated sample size will be round up to an integer.

(2) The formula can be used to calculate sample size required for a randomized trial study by setting `rho2=0`.

(3) When `rho2=0`, the formula derived by Latouche et al. (2004) looks the same as that derived by Schoenfeld (1983). Latouche et al. (2004) pointed out that in this situation, the interpretations are different hence the two formulae are actually different. In Latouched et al. (2004), the hazard ratio $\exp(\beta_1) = \theta$ measures the difference of effect of a covariate at two different levels on the subdistribution hazard for a particular failure, while in Schoenfeld (1983), the hazard ratio θ measures the difference of effect on the cause-specific hazard.

References

Schoenfeld DA. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics*. 39:499-503.

Latouche A., Porcher R. and Chevret S. (2004). Sample size formula for proportional hazards modelling of competing risks. *Statistics in Medicine*. 23:3263-3274.

See Also

`ssizeEpi.default`

Examples

```
# generate a toy pilot data set
X1 <- c(rep(1, 39), rep(0, 61))
set.seed(123456)
X2 <- sample(c(0, 1), 100, replace = TRUE)
failureFlag <- sample(c(0, 1), 100, prob = c(0.5, 0.5), replace = TRUE)

ssizeEpi(X1 = X1,
 X2 = X2,
 failureFlag = failureFlag,
         power = 0.80,
 theta = 2,
 alpha = 0.05)
```

`ssizeEpi.default` *Sample Size Calculation for Cox Proportional Hazards Regression*

Description

Sample size calculation for Cox proportional hazards regression with two covariates for Epidemiological Studies. The covariate of interest should be a binary variable. The other covariate can be either binary or non-binary. The formula takes into account competing risks and the correlation between the two covariates.

Usage

```
ssizeEpi.default(power,
 theta,
 p,
 psi,
 rho2,
 alpha = 0.05)
```

Arguments

`power`	numeric. postulated power.
`theta`	numeric. postulated hazard ratio.
`p`	numeric. proportion of subjects taking value one for the covariate of interest.
`psi`	numeric. proportion of subjects died of the disease of interest.
`rho2`	numeric. square of the correlation between the covariate of interest and the other covariate.
`alpha`	numeric. type I error rate.

Details

This is an implementation of the sample size formula derived by Latouche et al. (2004) for the following Cox proportional hazards regression in the epidemiological studies:

$$h(t|x_1, x_2) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2),$$

where the covariate X_1 is of our interest. The covariate X_1 has to be a binary variable taking two possible values: zero and one, while the covariate X_2 can be binary or continuous.

Suppose we want to check if the hazard of $X_1 = 1$ is equal to the hazard of $X_1 = 0$ or not. Equivalently, we want to check if the hazard ratio of $X_1 = 1$ to $X_1 = 0$ is equal to 1 or is equal to $\exp(\beta_1) = \theta$. Given the type I error rate α for a two-sided test, the total number of subjects required to achieve a power of $1 - \beta$ is

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{[\log(\theta)]^2 p(1-p)\psi(1-\rho^2)},$$

where z_a is the $100a$-th percentile of the standard normal distribution, ψ is the proportion of subjects died of the disease of interest, and

$$\rho = corr(X_1, X_2) = (p_1 - p_0) \times \sqrt{\frac{q(1-q)}{p(1-p)}},$$

and $p = Pr(X_1 = 1)$, $q = Pr(X_2 = 1)$, $p_0 = Pr(X_1 = 1|X_2 = 0)$, and $p_1 = Pr(X_1 = 1|X_2 = 1)$.

Value

The required sample size to achieve the specified power with the given type I error rate.

Note

(1) The calculated sample size will be round up to an integer.

(2) The formula can be used to calculate sample size required for a randomized trial study by setting `rho2=0`.

(3) When `rho2=0`, the formula derived by Latouche et al. (2004) looks the same as that derived by Schoenfeld (1983). Latouche et al. (2004) pointed out that in this situation, the interpretations are different hence the two formulae are actually different. In Latouched et al. (2004), the hazard ratio $\exp(\beta_1) = \theta$ measures the difference of effect of a covariate at two different levels on the subdistribution hazard for a particular failure, while in Schoenfeld (1983), the hazard ratio θ measures the difference of effect on the cause-specific hazard.

References

Schoenfeld DA. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics*. 39:499-503.

Latouche A., Porcher R. and Chevret S. (2004). Sample size formula for proportional hazards modelling of competing risks. *Statistics in Medicine*. 23:3263-3274.

See Also

ssizeEpi

Examples

```
# Examples at the end of Section 5.2 of Latouche et al. (2004)
# for a cohort study.
ssizeEpi.default(power = 0.80,
 theta = 2,
 p = 0.39,
 psi = 0.505,
                rho2 = 0.132^2,
 alpha = 0.05)
```

---

ssizeEpiCont — *Sample Size Calculation for Cox Proportional Hazards Regression with Nonbinary Covariates for Epidemiological Studies*

Description

Sample size calculation for Cox proportional hazards regression with nonbinary covariates for Epidemiological Studies.

Usage

```
ssizeEpiCont(formula,
     dat,
     var.X1,
     var.failureFlag,
     power,
     theta,
     alpha = 0.05)
```

Arguments

`formula`	a formula object relating the covariate of interest to other covariates to calculate the multiple correlation coefficient. The variables in formula must be in the data frame `dat`.
`dat`	a `nPilot` by `p` data frame representing the pilot data set, where `nPilot` is the number of subjects in the pilot study and the `p` (> 1) columns contains the covariate of interest and other covariates.
`var.X1`	character. name of the column in the data frame `dat`, indicating the covariate of interest.
`var.failureFlag`	character. name of the column in the data frame `dat`, indicating if a subject is failure (taking value 1) or alive (taking value 0).
`power`	numeric. postulated power.

`theta`	numeric. postulated hazard ratio.
`alpha`	numeric. type I error rate.

Details

This is an implementation of the sample size calculation formula derived by Hsieh and Lavori (2000) for the following Cox proportional hazards regression in the epidemiological studies:

$$h(t|x_1, \boldsymbol{x}_2) = h_0(t) \exp(\beta_1 x_1 + \boldsymbol{\beta}_2 \boldsymbol{x}_2,$$

where the covariate X_1 is a nonbinary variable and $\boldsymbol{X}_2$ is a vector of other covariates.

Suppose we want to check if the hazard ratio of the main effect $X_1 = 1$ to $X_1 = 0$ is equal to 1 or is equal to $\exp(\beta_1) = \theta$. Given the type I error rate α for a two-sided test, the total number of subjects required to achieve a sample size of $1 - \beta$ is

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{[\log(\theta)]^2 \sigma^2 \psi (1 - \rho^2)},$$

where z_a is the $100a$-th percentile of the standard normal distribution, $\sigma^2 = Var(X_1)$, ψ is the proportion of subjects died of the disease of interest, and ρ is the multiple correlation coefficient of the following linear regression:

$$x_1 = b_0 + \boldsymbol{b}^T \boldsymbol{x}_2.$$

That is, $\rho^2 = R^2$, where R^2 is the proportion of variance explained by the regression of X_1 on the vector of covriates $\boldsymbol{X}_2$.

rho^2, σ^2, and ψ will be estimated from a pilot study.

Value

`n`	the total number of subjects required.
`rho2`	square of the correlation between X_1 and X_2.
`sigma2`	variance of the covariate of interest.
`psi`	proportion of subjects died of the disease of interest.

Note

(1) Hsieh and Lavori (2000) assumed one-sided test, while this implementation assumed two-sided test. (2) The formula can be used to calculate ssize for a randomized trial study by setting `rho2=0`.

References

Hsieh F.Y. and Lavori P.W. (2000). Sample-size calculation for the Cox proportional hazards regression model with nonbinary covariates. *Controlled Clinical Trials*. 21:552-560.

See Also

`ssizeEpiCont.default`

Examples

```
# generate a toy pilot data set
set.seed(123456)
X1 <- rnorm(100, mean = 0, sd = 0.3126)
X2 <- sample(c(0, 1), 100, replace = TRUE)
failureFlag <- sample(c(0, 1), 100, prob = c(0.25, 0.75), replace = TRUE)
dat <- data.frame(X1 = X1, X2 = X2, failureFlag = failureFlag)

ssizeEpiCont(formula = X1 ~ X2,
     dat = dat,
     var.X1 = "X1",
     var.failureFlag = "failureFlag",
             power = 0.806,
     theta = exp(1),
     alpha = 0.05)
```

---

ssizeEpiCont.default *Sample Size Calculation for Cox Proportional Hazards Regression with Nonbinary Covariates for Epidemiological Studies*

Description

Sample size calculation for Cox proportional hazards regression with nonbinary covariates for Epidemiological Studies.

Usage

```
ssizeEpiCont.default(power,
     theta,
     sigma2,
     psi,
     rho2,
     alpha = 0.05)
```

Arguments

power	numeric. postulated power.
theta	numeric. postulated hazard ratio.
sigma2	numeric. variance of the covariate of interest.
psi	numeric. proportion of subjects died of the disease of interest.
rho2	numeric. square of the multiple correlation coefficient between the covariate of interest and other covariates.
alpha	numeric. type I error rate.

Details

This is an implementation of the sample size calculation formula derived by Hsieh and Lavori (2000) for the following Cox proportional hazards regression in the epidemiological studies:

$$h(t|x_1, \boldsymbol{x}_2) = h_0(t)\exp(\beta_1 x_1 + \boldsymbol{\beta}_2 \boldsymbol{x}_2,$$

where the covariate X_1 is a nonbinary variable and $\boldsymbol{X}_2$ is a vector of other covariates.

Suppose we want to check if the hazard ratio of the main effect $X_1 = 1$ to $X_1 = 0$ is equal to 1 or is equal to $\exp(\beta_1) = \theta$. Given the type I error rate α for a two-sided test, the total number of subjects required to achieve a sample size of $1 - \beta$ is

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{[\log(\theta)]^2 \sigma^2 \psi (1 - \rho^2)},$$

where z_a is the $100a$-th percentile of the standard normal distribution, $\sigma^2 = Var(X_1)$, ψ is the proportion of subjects died of the disease of interest, and ρ is the multiple correlation coefficient of the following linear regression:

$$x_1 = b_0 + \boldsymbol{b}^T \boldsymbol{x}_2.$$

That is, $\rho^2 = R^2$, where R^2 is the proportion of variance explained by the regression of X_1 on the vector of covriates $\boldsymbol{X}_2$.

Value

The total number of subjects required.

Note

(1) Hsieh and Lavori (2000) assumed one-sided test, while this implementation assumed two-sided test. (2) The formula can be used to calculate ssize for a randomized trial study by setting `rho2=0`.

References

Hsieh F.Y. and Lavori P.W. (2000). Sample-size calculation for the Cox proportional hazards regression model with nonbinary covariates. *Controlled Clinical Trials*. 21:552-560.

See Also

ssizeEpiCont

Examples

```
# example in the EXAMPLE section (page 557) of Hsieh and Lavori (2000).
# Hsieh and Lavori (2000) assumed one-sided test,
# while this implementation assumed two-sided test.
# Hence alpha=0.1 here (two-sided test) will correspond
# to alpha=0.05 of one-sided test in Hsieh and Lavori's (2000) example.
ssizeEpiCont.default(power = 0.806,
     theta = exp(1),
     sigma2 = 0.3126^2,
                     psi = 0.738,
```

```
        rho2 = 0.1837,
        alpha = 0.1)
```

---

ssizeEpiInt *Sample Size Calculation Testing Interaction Effect for Cox Proportional Hazards Regression*

Description

Sample size calculation testing interaction effect for Cox proportional hazards regression with two covariates for Epidemiological Studies. Both covariates should be binary variables. The formula takes into account the correlation between the two covariates.

Usage

```
ssizeEpiInt(X1,
    X2,
    failureFlag,
    power,
    theta,
    alpha = 0.05)
```

Arguments

`X1`	numeric. a `nPilot` by 1 vector, where `nPilot` is the number of subjects in the pilot data set. This vector records the values of the covariate of interest for the `nPilot` subjects in the pilot study. `X1` should be binary and take only two possible values: zero and one.
`X2`	numeric. a `nPilot` by 1 vector, where `nPilot` is the number of subjects in the pilot study. This vector records the values of the second covariate for the `nPilot` subjects in the pilot study. `X2` should be binary and take only two possible values: zero and one.
`failureFlag`	numeric. a `nPilot` by 1 vector of indicators indicating if a subject is failure (`failureFlag=1`) or alive (`failureFlag=0`).
`power`	numeric. postulated power.
`theta`	numeric. postulated hazard ratio.
`alpha`	numeric. type I error rate.

Details

This is an implementation of the sample size calculation formula derived by Schmoor et al. (2000) for the following Cox proportional hazards regression in the epidemoilogical studies:

$$h(t|x_1, x_2) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \gamma(x_1 x_2)),$$

where both covariates X_1 and X_2 are binary variables.

Suppose we want to check if the hazard ratio of the interaction effect $X_1X_2 = 1$ to $X_1X_2 = 0$ is equal to 1 or is equal to $\exp(\gamma) = \theta$. Given the type I error rate α for a two-sided test, the total number of subjects required to achieve the desired power $1-\beta$ is:

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 G}{[\log(\theta)]^2 \psi (1-p) p (1-\rho^2)},$$

where z_a is the $100a$-th percentile of the standard normal distribution, ψ is the proportion of subjects died of the disease of interest, and

$$\rho = corr(X_1, X_2) = (p_1 - p_0) \times \sqrt{\frac{q(1-q)}{p(1-p)}},$$

and $p = Pr(X_1 = 1)$, $q = Pr(X_2 = 1)$, $p_0 = Pr(X_1 = 1|X_2 = 0)$, and $p_1 = Pr(X_1 = 1|X_2 = 1)$, and

$$G = \frac{[(1-q)(1-p_0)p_0 + q(1-p_1)p_1]^2}{(1-q)q(1-p_0)p_0(1-p_1)p_1},$$

and $p0 = Pr(X_1 = 1|X_2 = 0) = myc/(mya + myc)$, $p1 = Pr(X_1 = 1|X_2 = 1) = myd/(myb + myd)$, $p = Pr(X_1 = 1) = (myc + myd)/n$, $q = Pr(X_2 = 1) = (myb + myd)/n$, $n = mya + myb + myc + myd$.

$p_{00} = Pr(X_1 = 0, \text{and}, X_2 = 0)$, $p_{01} = Pr(X_1 = 0, \text{and}, X_2 = 1)$, $p_{10} = Pr(X_1 = 1, \text{and}, X_2 = 0)$, $p_{11} = Pr(X_1 = 1, \text{and}, X_2 = 1)$.

$p_{00}, p_{01}, p_{10}, p_{11}$, and ψ will be estimated from the pilot data.

Value

`n`	the total number of subjects required.
`p`	estimated $Pr(X_1 = 1)$
`q`	estimated $Pr(X_2 = 1)$
`p0`	estimated $Pr(X_1 = 1\|X_2 = 0)$
`p1`	estimated $Pr(X_1 = 1\|X_2 = 1)$
`rho2`	square of the estimated $corr(X_1, X_2)$
`G`	a factor adjusting the sample size. The sample size needed to detect an effect of a prognostic factor with given error probabilities has to be multiplied by the factor `G` when an interaction of the same magnitude is to be detected.
`mya`	estimated number of subjects taking values $X_1 = 0$ and $X_2 = 0$.
`myb`	estimated number of subjects taking values $X_1 = 0$ and $X_2 = 1$.
`myc`	estimated number of subjects taking values $X_1 = 1$ and $X_2 = 0$.
`myd`	estimated number of subjects taking values $X_1 = 1$ and $X_2 = 1$.
`psi`	proportion of subjects died of the disease of interest.

References

Schmoor C., Sauerbrei W., and Schumacher M. (2000). Sample size considerations for the evaluation of prognostic factors in survival analysis. *Statistics in Medicine*. 19:441-452.

See Also

ssizeEpiInt.default0, ssizeEpiInt2

Examples

```
# generate a toy pilot data set
X1 <- c(rep(1, 39), rep(0, 61))
set.seed(123456)
X2 <- sample(c(0, 1), 100, replace = TRUE)
failureFlag <- sample(c(0, 1), 100, prob = c(0.25, 0.75), replace = TRUE)

ssizeEpiInt(X1 = X1,
    X2 = X2,
    failureFlag = failureFlag,
    power = 0.88,
    theta = 3,
    alpha = 0.05)
```

---

ssizeEpiInt.default0 *Sample Size Calculation Testing Interaction Effect for Cox Proportional Hazards Regression*

Description

Sample size calculation testing interaction effect for Cox proportional hazards regression with two covariates for Epidemiological Studies. Both covariates should be binary variables. The formula takes into account the correlation between the two covariates.

Usage

```
ssizeEpiInt.default0(power,
     theta,
     p,
     psi,
     G,
     rho2,
     alpha = 0.05)
```

Arguments

power	numeric. postulated power.
theta	numeric. postulated hazard ratio.
p	numeric. proportion of subjects taking value one for the covariate of interest.
psi	numeric. proportion of subjects died of the disease of interest.

`G` numeric. a factor adjusting the sample size. The sample size needed to detect an effect of a prognostic factor with given error probabilities has to be multiplied by the factor `G` when an interaction of the same magnitude is to be detected.

`rho2` numeric. square of the correlation between the covariate of interest and the other covariate.

`alpha` numeric. type I error rate.

Details

This is an implementation of the sample size calculation formula derived by Schmoor et al. (2000) for the following Cox proportional hazards regression in the epidemiological studies:

$$h(t|x_1, x_2) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \gamma(x_1 x_2)),$$

where both covariates X_1 and X_2 are binary variables.

Suppose we want to check if the hazard ratio of the interaction effect $X_1 X_2 = 1$ to $X_1 X_2 = 0$ is equal to 1 or is equal to $\exp(\gamma) = \theta$. Given the type I error rate α for a two-sided test, the total number of subjects required to achieve a power of $1 - \beta$ is

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 G}{[\log(\theta)]^2 \psi(1-p)p(1-\rho^2)},$$

where z_a is the $100a$-th percentile of the standard normal distribution, ψ is the proportion of subjects died of the disease of interest, and

$$\rho = corr(X_1, X_2) = (p_1 - p_0) \times \sqrt{\frac{q(1-q)}{p(1-p)}},$$

and $p = Pr(X_1 = 1)$, $q = Pr(X_2 = 1)$, $p_0 = Pr(X_1 = 1|X_2 = 0)$, and $p_1 = Pr(X_1 = 1|X_2 = 1)$, and

$$G = \frac{[(1-q)(1-p_0)p_0 + q(1-p_1)p_1]^2}{(1-q)q(1-p_0)p_0(1-p_1)p_1}$$

.

If X_1 and X_2 are uncorrelated, we have $p_0 = p_1 = p$ leading to $1/[(1-q)q]$. For $q = 0.5$, we have $G = 4$.

Value

The total number of subjects required.

References

Schmoor C., Sauerbrei W., and Schumacher M. (2000). Sample size considerations for the evaluation of prognostic factors in survival analysis. *Statistics in Medicine*. 19:441-452.

See Also

`ssizeEpiInt.default1`, `ssizeEpiInt2`

Examples

```
# Example at the end of Section 4 of Schmoor et al. (2000).
ssizeEpiInt.default0(power = 0.8227,
     theta = 3,
     p = 0.61,
     psi = 139 / 184,
                     G = 4.79177,
     rho2 = 0.015^2,
     alpha = 0.05)
```

`ssizeEpiInt.default1` *Sample Size Calculation Testing Interaction Effect for Cox Proportional Hazards Regression*

Description

Sample size calculation testing interaction effect for Cox proportional hazards regression with two covariates for Epidemiological Studies. Both covariates should be binary variables. The formula takes into account the correlation between the two covariates.

Usage

```
ssizeEpiInt.default1(power,
     theta,
     psi,
     p00,
     p01,
     p10,
     p11,
     alpha = 0.05)
```

Arguments

`power`	numeric. postulated power.
`theta`	numeric. postulated hazard ratio.
`psi`	numeric. proportion of subjects died of the disease of interest.
`p00`	numeric. proportion of subjects taking values $X_1 = 0$ and $X_2 = 0$, i.e., $p_{00} = Pr(X_1 = 0, \text{and}, X_2 = 0)$.
`p01`	numeric. proportion of subjects taking values $X_1 = 0$ and $X_2 = 1$, i.e., $p_{01} = Pr(X_1 = 0, \text{and}, X_2 = 1)$.
`p10`	numeric. proportion of subjects taking values $X_1 = 1$ and $X_2 = 0$, i.e., $p_{10} = Pr(X_1 = 1, \text{and}, X_2 = 0)$.
`p11`	numeric. proportion of subjects taking values $X_1 = 1$ and $X_2 = 1$, i.e., $p_{11} = Pr(X_1 = 1, \text{and}, X_2 = 1)$.
`alpha`	type I error rate.

Details

This is an implementation of the sample size calculation formula derived by Schmoor et al. (2000) for the following Cox proportional hazards regression in the epidemoilogical studies:

$$h(t|x_1, x_2) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \gamma(x_1 x_2)),$$

where both covariates X_1 and X_2 are binary variables.

Suppose we want to check if the hazard ratio of the interaction effect $X_1 X_2 = 1$ to $X_1 X_2 = 0$ is equal to 1 or is equal to $\exp(\gamma) = \theta$. Given the type I error rate α for a two-sided test, the total number of subjects required to achieve a power of $1 - \beta$ is

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \delta}{[\log(\theta)]^2 \psi},$$

where z_a is the $100a$-th percentile of the standard normal distribution, ψ is the proportion of subjects died of the disease of interest,

$$\delta = \frac{1}{p_{00}} + \frac{1}{p_{01}} + \frac{1}{p_{10}} + \frac{1}{p_{11}},$$

and $p_{00} = Pr(X_1 = 0, \text{and}, X_2 = 0)$, $p_{01} = Pr(X_1 = 0, \text{and}, X_2 = 1)$, $p_{10} = Pr(X_1 = 1, \text{and}, X_2 = 0)$, $p_{11} = Pr(X_1 = 1, \text{and}, X_2 = 1)$.

Value

The ssize of the test.

References

Schmoor C., Sauerbrei W., and Schumacher M. (2000). Sample size considerations for the evaluation of prognostic factors in survival analysis. *Statistics in Medicine*. 19:441-452.

See Also

ssizeEpiInt.default0, ssizeEpiInt2

Examples

```
# Example at the end of Section 4 of Schmoor et al. (2000).
# p00, p01, p10, and p11 are calculated based on Table III on page 448
# of Schmoor et al. (2000).
ssizeEpiInt.default1(power = 0.8227,
     theta = 3,
     psi = 139 / 184,
                     p00 = 50/184,
     p01 = 21 / 184,
     p10 = 78 / 184,
     p11 = 35 / 184,
                     alpha = 0.05)
```

ssizeEpiInt2 *Sample Size Calculation Testing Interaction Effect for Cox Proportional Hazards Regression*

Description

Sample size calculation testing interaction effect for Cox proportional hazards regression with two covariates for Epidemiological Studies. Both covariates should be binary variables. The formula takes into account the correlation between the two covariates.

Usage

```
ssizeEpiInt2(power,
     theta,
     psi,
     mya,
     myb,
     myc,
     myd,
     alpha = 0.05)
```

Arguments

power	numeric. postulated power.
theta	numeric. postulated hazard ratio.
psi	numeric. proportion of subjects died of the disease of interest.
mya	integer. number of subjects taking values $X_1 = 0$ and $X_2 = 0$ from the pilot study.
myb	integer. number of subjects taking values $X_1 = 0$ and $X_2 = 1$ from the pilot study.
myc	integer. number of subjects taking values $X_1 = 1$ and $X_2 = 0$ from the pilot study.
myd	integer. number of subjects taking values $X_1 = 1$ and $X_2 = 1$ from the pilot study.
alpha	numeric. type I error rate.

Details

This is an implementation of the sample size calculation formula derived by Schmoor et al. (2000) for the following Cox proportional hazards regression in the epidemiological studies:

$$h(t|x_1, x_2) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \gamma(x_1 x_2)),$$

where both covariates X_1 and X_2 are binary variables.

Suppose we want to check if the hazard ratio of the interaction effect $X_1X_2 = 1$ to $X_1X_2 = 0$ is equal to 1 or is equal to $\exp(\gamma) = \theta$. Given the type I error rate α for a two-sided test, the total number of subjects required to achieve a power of $1 - \beta$ is

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 G}{[\log(\theta)]^2 \psi (1-p) p (1-\rho^2)},$$

where z_a is the $100a$-th percentile of the standard normal distribution, ψ is the proportion of subjects died of the disease of interest, and

$$\rho = corr(X_1, X_2) = (p_1 - p_0) \times \sqrt{\frac{q(1-q)}{p(1-p)}},$$

and $p = Pr(X_1 = 1)$, $q = Pr(X_2 = 1)$, $p_0 = Pr(X_1 = 1|X_2 = 0)$, and $p_1 = Pr(X_1 = 1|X_2 = 1)$, and

$$G = \frac{[(1-q)(1-p_0)p_0 + q(1-p_1)p_1]^2}{(1-q)q(1-p_0)p_0(1-p_1)p_1},$$

and $p0 = Pr(X_1 = 1|X_2 = 0) = myc/(mya + myc)$, $p1 = Pr(X_1 = 1|X_2 = 1) = myd/(myb + myd)$, $p = Pr(X_1 = 1) = (myc + myd)/n$, $q = Pr(X_2 = 1) = (myb + myd)/n$, $n = mya + myb + myc + myd$.

$p_{00} = Pr(X_1 = 0, \text{and}, X_2 = 0)$, $p_{01} = Pr(X_1 = 0, \text{and}, X_2 = 1)$, $p_{10} = Pr(X_1 = 1, \text{and}, X_2 = 0)$, $p_{11} = Pr(X_1 = 1, \text{and}, X_2 = 1)$.

Value

The total number of subjects required.

References

Schmoor C., Sauerbrei W., and Schumacher M. (2000). Sample size considerations for the evaluation of prognostic factors in survival analysis. *Statistics in Medicine*. 19:441-452.

See Also

ssizeEpiInt.default0, ssizeEpiInt.default1

Examples

```
# Example at the end of Section 4 of Schmoor et al. (2000).
# mya, myb, myc, and myd are obtained from Table III on page 448
# of Schmoor et al. (2000).
ssizeEpiInt2(power = 0.8227,
     theta = 3,
     psi = 139 / 184,
             mya = 50,
     myb = 21,
     myc = 78,
     myd = 35,
     alpha = 0.05)
```

Index