

Package ‘PracTools’

June 21, 2014

Type Package

Title Tools for Designing and Weighting Survey Samples

Version 0.0-2

Date 2014-03-30

Author Richard Valliant, Jill A. Dever, Frauke Kreuter

Maintainer Richard Valliant <rvallian@umd.edu>

Description PracTools contains functions for sample size calculation for survey samples using stratified or clustered one-, two-, and three-stage sample designs. Other functions compute variance components for multistage designs and sample sizes in two-phase designs. A number of example datasets are included.

Suggests doBy, foreign, pps, reshape, sampling, samplingbook, survey

License GPL (>= 2)

LazyLoad yes

Depends R (>= 2.10)

R topics documented:

BW2stagePPS	2
BW2stagePPSe	4
BW2stageSRS	6
BW3stagePPS	7
BW3stagePPSe	9
clusOpt2	11
clusOpt2fixedPSU	13
clusOpt3	14
clusOpt3fixedPSU	16
CVcalc2	17
CVcalc3	19
Domainy1y2	20
dub	21
gamEst	22
gammaFit	23
HMT	24

hospital	25
labor	26
MDarea.pop	27
nCont	28
nDep2sam	29
nDomain	31
nhis	32
nhis.large	33
nLogOdds	34
nProp	35
nProp2sam	36
nPropMoe	38
NRFUopt	39
nWilson	41
smho.N874	42
smho98	43
strAlloc	44
wtdvar	46

Index 47

BW2stagePPS	<i>Relvariance components for 2-stage sample</i>
-------------	--

Description

Compute components of relvariance for a sample design where primary sampling units (PSUs) are selected with probability proportional to size (*pps*) and elements are selected via simple random sampling (*srs*). The input is an entire sampling frame.

Usage

```
BW2stagePPS(X, pp, psuID)
```

Arguments

X	data vector; length is the number of elements in the population.
pp	vector of 1-draw probabilities for the PSUs. This vector must be as long as X. Each element in a given PSU should have the same value in pp. PSUs must be in the same order as in X.
psuID	vector of PSU identification numbers. This vector must be as long as X. Each element in a given PSU should have the same value in psuID. PSUs must be in the same order as in X.

Details

BW2stagePPS computes the between and within population relvariance components appropriate for a two-stage sample in which PSUs are selected with varying probabilities and with replacement. Elements within PSUs are selected by simple random sampling. The components are appropriate for approximating the relvariance of the probability-with-replacement (*pwr*)-estimator of a total when the same number of elements are selected within each sample PSU. The function requires that an entire frame of PSUs and elements be input. (Use [BW2stagePPSe](#) if only a sample of PSUs and elements are available.)

Value

List object with values:

B2	between PSU unit relvariance
W2	within PSU unit relvariance
unit relvar	unit relvariance for population
B2+W2	sum of between and within relvariance estimates
k	ratio of $B^2 + W^2$ to unit relvariance
delta	measure of homogeneity with PSUs estimated as $B^2/(B^2 + W^2)$

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

- Cochran, W.G. (1977, pp.308-310). *Sampling Techniques*. New York: John Wiley & Sons.
- Saerndal, C.E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.
- Valliant, R., Dever, J., Kreuter, F. (2013, sect. 9.3.2). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[BW2stagePPSe](#), [BW2stageSRS](#), [BW3stagePPS](#), [BW3stagePPSe](#)

Examples

```
data(MDarea.pop)
MDsub <- MDarea.pop[1:100000,]
# Use PSU and SSU variables to define psu's
pp.PSU <- table(MDsub$PSU) / nrow(MDsub)
pp.SSU <- table(MDsub$SSU) / nrow(MDsub)
# components with psu's defined by the PSU variable
BW2stagePPS(MDsub$y1, pp=pp.PSU, psuID=MDsub$PSU)
# components with psu's defined by the SSU variable
BW2stagePPS(MDsub$y1, pp=pp.SSU, psuID=MDsub$SSU)

# Use census tracts and block groups to define psu's
trtBG <- 10*MDsub$TRACT + MDsub$BLKGROUP
pp.trt <- table(MDsub$TRACT) / nrow(MDsub)
pp.BG <- table(trtBG) / nrow(MDsub)
# components with psu's defined by tracts
BW2stagePPS(MDsub$ins.cov, pp=pp.trt, psuID=MDsub$TRACT)
# components with psu's defined by block groups
BW2stagePPS(MDsub$ins.cov, pp=pp.BG, psuID=trtBG)
```

BW2stagePPSe

*Estimated relvariance components for 2-stage sample***Description**

Estimate components of relvariance for a sample design where primary sampling units (PSUs) are selected with *pps* and elements are selected via *srs*. The input is a sample selected in this way.

Usage

```
BW2stagePPSe(Ni, ni, X, psuID, w, m, pp)
```

Arguments

Ni	vector of number of elements in the population of each sample PSU; length is the number of PSUs in the sample.
ni	vector of number of sample elements in each sample PSU; length is the number of PSUs in the sample. PSUs must be in the same order in ni and in X.
X	data vector for sample elements; length is the number of elements in the sample. These must be in PSU order. PSUs must be in the same order in ni and in X.
psuID	vector of PSU identification numbers. This vector must be as long as X. Each element in a given PSU should have the same value in psuID.
w	vector of full sample weights. This vector must be as long as X. Vector must be in the same order as X.
m	number of sample PSUs
pp	vector of 1-draw probabilities for the PSUs. This vector must be as long as X. Each element in a given PSU should have the same value in pp. Vector must be in the same order as X.

Details

BW2stagePPSe computes the between and within population variance and relvariance components appropriate for a two-stage sample in which PSUs are selected with varying probabilities and with replacement. Elements within PSUs are selected by simple random sampling. The number of elements selected within each sample PSU can vary but must be at least two. The estimated components are appropriate for approximating the relvariance of the *pwr*-estimator of a total when the same number of elements are selected within each sample PSU. This function can also be used if PSUs are selected by *srswr* by appropriate definition of pp.

Value

List with values:

Vpsu	estimated between PSU unit variance
Vssu	estimated within PSU unit variance
B	estimated between PSU unit relvariance
W	estimated within PSU unit relvariance
delta	intraclass correlation estimated as $B/(B+W)$

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

- Cochran, W.G. (1977, pp.308-310). *Sampling Techniques*. New York: John Wiley & Sons.
- Valliant, R., Dever, J., Kreuter, F. (2013, sect. 9.4.1). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[BW2stagePPS](#), [BW2stageSRS](#), [BW3stagePPS](#), [BW3stagePPSe](#)

Examples

```
## Not run:
require(sampling)
require(reshape)      # has function that allows renaming variables
data(MDarea.pop)
Ni <- table(MDarea.pop$TRACT)
m <- 20
probi <- m*Ni / sum(Ni)
  # select sample of clusters
sam <- cluster(data=MDarea.pop, clustername="TRACT", size=m, method="systematic",
               pik=probi, description=TRUE)
  # extract data for the sample clusters
samclus <- getdata(MDarea.pop, sam)
samclus <- rename(samclus, c(Prob = "pi1"))

  # treat sample clusters as strata and select srswor from each
s <- strata(data = as.data.frame(samclus), stratanames = "TRACT",
            size = rep(50,m), method="srswor")
# extracts the observed data
samdat <- getdata(samclus,s)
samdat <- rename(samdat, c(Prob = "pi2"))

  # extract pop counts for PSUs in sample
pick <- names(Ni) %in% sort(unique(samdat$TRACT))
Ni.sam <- Ni[pick]
pp <- Ni.sam / sum(Ni)
wt <- 1/samdat$pi1/samdat$pi2

BW2stagePPSe(Ni = Ni.sam, ni = rep(50,20), X = samdat$y1,
              psuID = samdat$TRACT, w = wt,
              m = 20, pp = pp)

## End(Not run)
```

BW2stageSRS

*Relvariance components for 2-stage sample***Description**

Compute components of relvariance for a sample design where primary sampling units (PSUs) and elements are selected via *srs*. The input is an entire sampling frame.

Usage

```
BW2stageSRS(X, psuID)
```

Arguments

X	data vector; length is the number of elements in the population.
psuID	vector of PSU identification numbers. This vector must be as long as X. Each element in a given PSU should have the same value in psuID. PSUs must be in the same order as in X.

Details

BW2stageSRS computes the between and within population relvariance components appropriate for a two-stage sample in which PSUs are selected via *srs* (either with or without replacement). Elements within PSUs are assumed to be selected by *srswor*. The same number of elements is assumed to be selected within each sample PSU. The function requires that an entire frame of PSUs and elements be input. (Use [BW2stagePPSe](#) if only a sample of PSUs and elements are available.)

Value

List with values:

B2	between PSU unit relvariance
W2	within PSU unit relvariance
unit relvar	unit relvariance for population
B2+W2	$B^2 + W^2$
k	ratio of $B^2 + W^2$ to unit relvariance
delta full	intraclass correlation estimated as $B^2/(B^2 + W^2)$

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

Cochran, W.G. (1977, chap. 11). *Sampling Techniques*. New York: John Wiley & Sons.
 Valliant, R., Dever, J., Kreuter, F. (2013, sect. 9.2.1). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[BW2stagePPS](#), [BW2stagePPSe](#), [BW3stagePPS](#), [BW3stagePPSe](#)

Examples

```
data(MDarea.pop)
MDsub <- MDarea.pop[1:100000,]
# psu's are defined by PSU variable
BW2stageSRS(abs(MDsub$Hispanic-2), psuID=MDsub$PSU)
# psu's are defined by SSU variable
BW2stageSRS(abs(MDsub$Hispanic-2), psuID=MDsub$SSU)
```

BW3stagePPS

Relvariance components for 3-stage sample

Description

Compute components of relvariance for a sample design where primary sampling units (PSUs) are selected with *ppswr* and secondary sampling units (SSUs) and elements within SSUs are selected via *srs*. The input is an entire sampling frame.

Usage

```
BW3stagePPS(X, pp, psuID, ssuID)
```

Arguments

X	data vector; length is the number of elements in the population.
pp	vector of 1-draw probabilities for the PSUs. This vector must be as long as X and will contain the same value for the units that are in the same PSU. PSUs must be in the same order as in X.
psuID	vector of PSU identification numbers. This vector must be as long as X. Each element in a given PSU should have the same value in psuID. PSUs must be in the same order as in X.
ssuID	vector of SSU identification numbers. This vector must be as long as X. Each element in a given SSU should have the same value in ssuID. PSUs and SSUs must be in the same order as in X. ssuID should have the form psuID ssuID within PSU).

Details

BW3stagePPS computes the between and within population relvariance components appropriate for a three-stage sample in which PSUs are selected with varying probabilities and with replacement. SSUs and elements within SSUs are selected by simple random sampling. The components are appropriate for approximating the relvariance of the *pwr*-estimator of a total when the same number of SSUs are selected within each PSU, and the same number of elements are selected within each sample SSU. The function requires that an entire sampling frame of PSUs and elements be input. (Use [BW2stagePPSe](#) if only a sample of PSUs, SSUs, and elements is available.)

Value

List with values:

B	between PSU unit relvariance
W	within PSU unit relvariance computed as if the sample were two-stage
W2	unit relvariance among SSU totals
W3	unit relvariance among elements within PSU/SSUs
unit relvar	unit relvariance for population
k1	ratio of $B^2 + W^2$ to unit relvariance
k2	ratio of $W_2^2 + W_3^2$ to unit relvariance
delta1	homogeneity measure among elements within PSUs estimated as $B^2/(B^2 + W^2)$
delta2	homogeneity measure among elements within SSUs estimated as $W_2^2/(W_2^2 + W_3^2)$

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

- Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953, chap. 9, p.211). *Sample Survey Methods and Theory*, Vol.I. John Wiley & Sons.
- Saerndal, C.E., Swensson, B., and Wretman, J. (1992, p.149). *Model Assisted Survey Sampling*. Springer.
- Valliant, R., Dever, J., Kreuter, F. (2013, sect. 9.2.4). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[BW2stagePPS](#), [BW2stagePPSe](#), [BW2stageSRS](#), [BW3stagePPSe](#)

Examples

```
data(MDarea.pop)
MDsub <- MDarea.pop[1:100000,]
M <- length(unique(MDsub$PSU))
# srs/srs/srs design
pp.PSU <- rep(1/M,M)
BW3stagePPS(X=MDsub$y1, pp=pp.PSU, psuID=MDsub$PSU, ssuID=MDsub$SSU)
# ppswr/srs/srs design
pp.PSU <- table(MDsub$PSU) / nrow(MDsub)
BW3stagePPS(X=MDsub$y1, pp=pp.PSU, psuID=MDsub$PSU, ssuID=MDsub$SSU)
```


BW3stagePPSe

*Estimated relvariance components for 3-stage sample***Description**

Estimate components of relvariance for a sample design where primary sampling units (PSUs) are selected with probability proportional to size with replacement (*ppswr*) and secondary sampling units (SSUs) and elements within SSUs are selected via simple random sampling (*srs*). The input is a sample selected in this way.

Usage

```
BW3stagePPSe(dat, v, Ni, Qi, Qij, m)
```

Arguments

<code>dat</code>	data frame for sample elements with PSU and SSU identifiers, weights, and analysis variable(s). The data frame should be sorted in hierarchical order: by PSU and SSU within PSU. Required names for columns: <code>psuID</code> = PSU identifier; <code>ssuID</code> = SSU identifier. These must be unique, i.e., numbering should not restart within each PSU. Setting <code>ssuID = psuID ssuID</code> within PSU is a method of doing this. <code>w1i</code> = vector of weights for PSUs; <code>w2ij</code> = vector of weights for SSUs (PSU weight*SSU weight within PSU); <code>w</code> = full sample weight
<code>v</code>	Name or number of column in data frame <code>dat</code> with variable to be analyzed.
<code>Ni</code>	<code>m</code> -vector of number of SSUs in the population in the sample PSUs; <code>m</code> is number of sample PSUs.
<code>Qi</code>	<code>m</code> -vector of number of elements in the population in the sample PSUs
<code>Qij</code>	vector of numbers of elements in the population in the sample SSUs
<code>m</code>	number of sample PSUs

Details

BW3stagePPSe computes the between and within population relvariance components appropriate for a three-stage sample in which PSUs are selected with varying probabilities and with replacement. SSUs and elements within SSUs are selected by simple random sampling. The estimated components are appropriate for approximating the relvariance of the *pwr*-estimator of a total when the same number of SSUs are selected within each PSU, and the same number of elements are selected within each sample SSU.

Value

List with values:

<code>Vpsu</code>	estimated between PSU unit variance
<code>Vssu</code>	estimated second-stage unit variance among SSU totals
<code>Vtsu</code>	estimated third-stage unit variance
<code>B</code>	estimated between PSU unit relvariance
<code>W</code>	estimated within PSU unit relvariance computed as if the sample were two-stage
<code>W2</code>	estimated unit relvariance among SSU totals

W3	estimated third-stage unit relvariance among elements within PSU/SSUs
delta1	homogeneity measure among elements within PSUs estimated as $B^2/(B^2 + W^2)$
delta2	homogeneity measure among elements within SSUs estimated as $W_2^2/(W_2^2 + W_3^2)$

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

- Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953, chap. 9, sect. 10). *Sample Survey Methods and Theory*, Vol.II. New York: John Wiley & Sons.
- Valliant, R., Dever, J., Kreuter, F. (2013, sect. 9.4.2). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[BW2stagePPS](#), [BW2stagePPSe](#), [BW2stageSRS](#), [BW3stagePPS](#)

Examples

```
## Not run:
# select 3-stage sample from Maryland population
data(MDarea.pop)
MDpop <- MDarea.pop
require(sampling)
require(reshape) # has function that allows renaming variables
# make counts of SSUs and elements per PSU
xx <- do.call("rbind",list(by(1:nrow(MDpop),MDpop$PSU,head,1)))
pop.tmp <- MDpop[xx,]
Ni <- table(pop.tmp$PSU)
Qi <- table(MDarea.pop$PSU)
Qij <- table(MDpop$SSU)
m <- 30 # no. of PSUs to select
probi <- m*Qi / sum(Qi)
# select sample of clusters
sam <- cluster(data=MDpop, clustname="PSU", size=m, method="systematic",
              pik=probi, description=TRUE)
# extract data for the sample clusters
samclus <- getdata(MDpop, sam)
samclus <- rename(samclus, c(Prob = "p1i"))
samclus <- samclus[order(samclus$PSU),]
# treat sample clusters as strata and select srswor of block groups from each
# identify psu IDs for 1st instance of each ssuID
xx <- do.call("rbind",list(by(1:nrow(samclus),samclus$SSU,head,1)))
SSUs <- cbind(PSU=samclus$PSU[xx], SSU=samclus$SSU[xx])
# select 2 SSUs per tract
n <- 2
s <- strata(data = as.data.frame(SSUs), stratanames = "PSU",
           size = rep(n,m), method="srswor")
s <- rename(s, c(Prob = "p2i"))
# extract the SSU data
# s contains selection probs of SSUs, need to get those onto data file
```

```

SSUsam <- SSUs[s$ID_unit, ]
SSUsam <- cbind(SSUsam, s[, 2:3])
# identify rows in PSU sample that correspond to sample SSUs
tmp <- samclus$SSU %in% SSUsam$SSU
SSUdat <- samclus[tmp,]
SSUdat <- merge(SSUdat, SSUsam[, c("p2i", "SSU")], by="SSU")
# select srswor from each sample SSU
n.SSU <- m*n
s <- strata(data = as.data.frame(SSUdat), stratanames = "SSU",
           size = rep(50, n.SSU), method="srswor")
s <- rename(s, c(Prob = "p3i"))
samclus <- getdata(SSUdat, s)
del <- (1:ncol(samclus))[dimnames(samclus)[[2]] %in% c("ID_unit", "Stratum")]
samclus <- samclus[, -del]
# extract pop counts for PSUs in sample
pick <- names(Qi) %in% sort(unique(samclus$PSU))
Qi.sam <- Qi[pick]
# extract pop counts of SSUs for PSUs in sample
pick <- names(Ni) %in% sort(unique(samclus$PSU))
Ni.sam <- Ni[pick]
# extract pop counts for SSUs in sample
pick <- names(Qij) %in% sort(unique(samclus$SSU))
Qij.sam <- Qij[pick]
# compute full sample weight and wts for PSUs and SSUs
wt <- 1 / samclus$p1i / samclus$p2i / samclus$p3i
w1i <- 1 / samclus$p1i
w2ij <- 1 / samclus$p1i / samclus$p2i
samdat <- data.frame(psuID = samclus$PSU, ssuID = samclus$SSU,
                    w1i = w1i, w2ij = w2ij, w = wt,
                    samclus[, c("y1", "y2", "y3", "ins.cov", "hosp.stay")])
BW3stagePPSe(dat=samdat, v="y1", Ni=Ni.sam, Qi=Qi.sam, Qij=Qij.sam, m)

## End(Not run)

```

clusOpt2

*Compute optimal sample sizes for a two-stage sample***Description**

Compute the sample sizes that minimize the variance of the *pwr*-estimator of a total in a two-stage sample.

Usage

```
clusOpt2(C1, C2, delta, unit.rv, k=1, CV0=NULL, tot.cost=NULL, cal.sw)
```

Arguments

C1	unit cost per primary sampling unit (PSU)
C2	unit cost per element
delta	homogeneity measure δ
unit.rv	unit relvariance
k	ratio of $B^2 + W^2$ to unit relvariance

CV0	target CV
tot.cost	total budget for variable costs
cal.sw	specify type of optimum: 1 = find optimal m.opt for fixed total budget; 2 = find optimal m.opt for target CV0

Details

clusOpt2 will compute m_{opt} and \bar{n}_{opt} for a two-stage sample which uses simple random sampling at each stage or *ppswr* at the first stage and *srs* at the second.

Value

List with values:

C1	unit cost per PSU
C2	unit cost per element
delta	homogeneity measure
unit relvar	unit relvariance
k	ratio of $B^2 + W^2$ to unit relvariance
cost	total budget for variable costs, $C - C_0$ if cal.sw=1; or computed cost if cal.sw=2
m.opt	optimum number of sample PSUs
n.opt	optimum number of sample elements per PSU
CV	computed CV if cal.sw=1; or target CV if cal.sw=2

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953, chap. 6, sect. 16). *Sample Survey Methods and Theory*, Vol.I. John Wiley & Sons.

Valliant, R., Dever, J., Kreuter, F. (2013, sect. 9.3.1). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[clusOpt2fixedPSU](#), [clusOpt3](#), [clusOpt3fixedPSU](#)

Examples

```
# optimum for a fixed total budget
clusOpt2(C1=750, C2=100, delta=0.05, unit.rv=1, k=1, tot.cost=100000, cal.sw=1)
# optimum for a target CV
clusOpt2(C1=750, C2=100, delta=0.01, unit.rv=1, k=1, CV0=0.05, cal.sw=2)
```

clusOpt2fixedPSU	<i>Optimal number of sample elements per PSU in a two-stage sample when the sample of PSUs is fixed</i>
------------------	---

Description

Compute the optimum number of sample elements per primary sampling unit (PSU) for a fixed set of PSUs

Usage

```
clusOpt2fixedPSU(C1, C2, m, delta, unit.rv, k=1, CV0=NULL, tot.cost, cal.sw)
```

Arguments

C1	unit cost per PSU
C2	unit cost per element
m	number of sample PSU's (fixed)
delta	homogeneity measure
unit.rv	unit relvariance
k	ratio of $B^2 + W^2$ to unit relvariance
CV0	target CV
tot.cost	total budget for variable costs
cal.sw	specify type of optimum: 1 = find optimal \bar{n} for fixed total budget; 2 = find optimal \bar{n} for target CV0

Details

clusOpt2fixedPSU will compute \bar{n}_{opt} for a two-stage sample which uses simple random sampling at each stage or *ppswr* at the first stage and *srs* at the second. The PSU sample is fixed.

Value

List with values:

C1	unit cost per PSU
C2	unit cost per element
m	number of (fixed) sample PSUs
delta	homogeneity measure
unit relvar	unit relvariance
k	ratio of $B^2 + W^2$ to unit relvariance
cost	total budget for variable costs, $C - C_0$ if cal.sw=1; or computed cost if cal.sw=2
n	optimum number of sample elements per PSU
CV	computed CV if cal.sw=1; or target CV if cal.sw=2

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

Valliant, R., Dever, J., Kreuter, F. (2013, sect. 9.3.3). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[clusOpt2](#), [clusOpt3](#), [clusOpt3fixedPSU](#)

Examples

```
# optima for a vector of budgets
clusOpt2fixedPSU(C1=500, C2=100, m=100, delta=0.05, unit.rv=2, k=1, CV0=NULL,
  tot.cost=c(100000, 500000, 10^6), cal.sw=1)
# optima for a target CV and vector of PSU costs
clusOpt2fixedPSU(C1=c(500,1000,5000), C2=100, m=100, delta=0.05, unit.rv=2, k=1,
  CV0=0.05, tot.cost=NULL, cal.sw=2)
```

clusOpt3

Compute optimal sample sizes for a three-stage sample

Description

Compute the sample sizes that minimize the variance of the *pwr*-estimator of a total in a three-stage sample.

Usage

```
clusOpt3(unit.cost, delta1, delta2, unit.rv, k1=1, k2=1, CV0=NULL, tot.cost=NULL, cal.sw)
```

Arguments

unit.cost	vector with three components for unit costs: C1 = unit cost per primary sampling unit (PSU); C2 = unit cost per secondary sampling unit (SSU); C3 = unit cost per element
delta1	homogeneity measure among elements within PSUs
delta2	homogeneity measure among elements within SSUs
unit.rv	population unit relvariance
k1	ratio of $B^2 + W^2$ to the population unit relvariance
k2	ratio of $W_2^2 + W_3^2$ to the population unit relvariance
CV0	target CV
tot.cost	total budget for variable costs
cal.sw	specify type of optimum: 1 = find optimal m.opt for fixed total budget; 2 = find optimal m.opt for target CV0

Details

clusOpt3 will compute m_{opt} , \bar{n}_{opt} , and \bar{q}_{opt} for a three-stage sample which uses simple random sampling at each stage or *ppswr* at the first stage and *srs* at the second and third stages.

Value

List with values:

C1	unit cost per PSU
C2	unit cost per SSU
C3	unit cost per element
delta1	homogeneity measure among elements within PSUs
delta2	homogeneity measure among elements within SSUs
unit relvar	unit relvariance
k1	ratio of $B^2 + W^2$ to the population unit relvariance
k2	ratio of $W_2^2 + W_3^2$ to the population unit relvariance
cost	total budget for variable costs if cal.sw=1; or computed cost if cal.sw=2
m.opt	optimum number of sample PSUs
n.opt	optimum number of sample SSUs per PSU
q.opt	optimum number of sample elements per SSU
CV	achieved CV if cal.sw=1 or target CV if cal.sw=2

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

- Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953, p. 225). *Sample Survey Methods and Theory*, Vol.II. John Wiley & Sons.
- Valliant, R., Dever, J., Kreuter, F. (2013, sect. 9.3.2). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[clusOpt2](#), [clusOpt2fixedPSU](#), [clusOpt3fixedPSU](#)

Examples

```
# optima for a fixed total budget
clusOpt3(unit.cost=c(500, 100, 120), delta1=0.01, delta2=0.10, unit.rv=1,
  k1=1, k2=1, tot.cost=100000, cal.sw=1)

# optima for a target CV
clusOpt3(unit.cost=c(500, 100, 120), delta1=0.01, delta2=0.10, unit.rv=1,
  k1=1, k2=1, CV0=0.01, cal.sw=2)
```

clusOpt3fixedPSU	<i>Compute optimal number of sample secondary sampling units (SSUs) and elements per primary sampling unit (PSU) in a three-stage sample when the sample of PSUs is fixed</i>
------------------	---

Description

Compute the sample sizes that minimize the variance of the *pwr*-estimator of a total in a three-stage sample when the PSU sample is fixed.

Usage

```
clusOpt3fixedPSU(unit.cost, m, delta1, delta2, unit.rv, k1=1, k2=1, CV0=NULL,
  tot.cost=NULL, cal.sw)
```

Arguments

unit.cost	3-vector of unit costs: C1 = unit cost per PSU; C2 = unit cost per SSU; C3 = unit cost per element
m	number of sample PSUs (fixed)
delta1	homogeneity measure among elements within PSUs
delta2	homogeneity measure among elements within SSUs
unit.rv	unit relvariance
k1	ratio of $B^2 + W^2$ to unit relvariance
k2	ratio of $W_2^2 + W_3^2$ to unit relvariance
CV0	target CV
tot.cost	total budget for variable costs, including PSU costs
cal.sw	specify type of optimum: 1 = find optimal m.opt for fixed total budget; 2 = find optimal m.opt for target CV0

Details

clusOpt3 will compute \bar{n}_{opt} and \bar{q}_{opt} for a three-stage sample which uses simple random sampling at each stage or *ppswr* at the first stage and *srs* at the second and third stages. The set of sample PSUs is assumed to be fixed. "Variable costs" in tot.cost includes the budget for all costs that vary with the number of sample PSUs, SSUs, and elements, i.e., $C_1m + C_2m\bar{n} + C_3m\bar{n}\bar{q}$.

Value

List with values:

C1	unit cost per PSU
C2	unit cost per SSU
C3	unit cost per element
m	number of sample PSUs (fixed)
delta1	homogeneity measure among elements within PSUs
delta2	homogeneity measure among elements within SSUs

unit relvar	unit relvariance
k1	ratio of $B^2 + W^2$ to unit relvariance
k2	ratio of $W_2^2 + W_3^2$ to unit relvariance
cost	budget constraint, tot . cost if cal . sw=1; computed cost if cal . sw=2
n	optimum number of sample SSUs per PSU
q	optimum number of sample elements per SSU
CV	achieved CV, used if cal . sw=1; or target CV, used if cal . sw=2
CV check	computed CV based on optimal sample sizes; used only if cal . sw=2

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

- Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953, p. 225). *Sample Survey Methods and Theory*, Vol.II. John Wiley & Sons.
- Valliant, R., Dever, J., Kreuter, F. (2013, sect. 9.3.2). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[clusOpt2](#), [clusOpt2fixedPSU](#), [clusOpt3](#)

Examples

```
# optima for a fixed total budget
clusOpt3fixedPSU(unit.cost=c(500, 100, 120), m=100, delta1=0.01, delta2=0.05, unit.rv=1,
  k1=1, k2=1, tot.cost=500000, cal.sw=1)
# optima for a target CV
clusOpt3fixedPSU(unit.cost=c(500, 100, 120), m=100, delta1=0.01, delta2=0.05, unit.rv=1,
  k1=1, k2=1, CV0=0.05, cal.sw=2)
```

CVcalc2

Coefficient of variation of an estimated total in a 2-stage sample

Description

Compute the coefficient of variation of an estimated total in a two-stage design. Primary sampling units (PSUs) can be selected either with probability proportional to size (*pps*) or with equal probability. Elements are selected via simple random sampling (*srs*).

Usage

```
CVcalc2(V=NULL, m=NULL, nbar=NULL, k=1, delta=NULL, Bsqr=NULL, Wsq=NULL)
```

Arguments

V	unit relvariance of analysis variable in the population
m	number of sample PSUs
nbar	number of sample elements per PSU
k	ratio of $B^2 + W^2$ to V . Default value is 1.
delta	measure of homogeneity equal to $B^2 / (B^2 + W^2)$
Bsq	unit relvariance of PSU totals, equal to population variance of totals divided by \bar{t}_U^2 if PSUs are selected by simple random sampling; or, equal to $S_{U1(pwr)}^2$ divided by t_U^2 if PSUs are selected by <i>ppswr</i>
Wsqr	within PSU relvariance, equal to average element population variance divided by \bar{y}_U^2

Details

CVcalc2 computes the coefficient of variation of an estimated total for a two-stage sample. PSUs can be selected either with varying probabilities and with replacement or with equal probabilities and with replacement. Elements within PSUs are selected by simple random sampling. The *CV* formula is appropriate for approximating the relvariance of the probability-with-replacement (*pwr*)-estimator of a total when the same number of elements is selected within each sample PSU.

Value

Value of the coefficient of variation of an estimated total

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

- Cochran, W.G. (1977, pp.308-310). *Sampling Techniques*. New York: John Wiley & Sons.
- Saerndal, C.E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.
- Valliant, R., Dever, J., Kreuter, F. (2013, sect. 9.2.1). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[CVcalc3](#)

Examples

```
CVcalc2(V=1, m=20, nbar=5, k=1, delta=0.05)
CVcalc2(V=10, m=20, nbar=5, k=1, delta=0.5)
CVcalc2(V=2.5, m=20, nbar=5, k=2, Bsq=1, Wsq=4)
```

CVcalc3

*Coefficient of variation of an estimated total in a 3-stage sample***Description**

Compute the coefficient of variation of an estimated total in a three-stage design. Primary sampling units (PSUs) can be selected either with probability proportional to size (*pps*) or with equal probability. Secondary units and elements within SSUs are selected via simple random sampling (*srs*).

Usage

```
CVcalc3(V=NULL, m=NULL, nbar=NULL, qbar=NULL, k1=1, k2=1, delta1=NULL, delta2=NULL,
        Bsq=NULL, Wsq=NULL, W2sq=NULL, W3sq=NULL)
```

Arguments

V	unit relvariance of analysis variable in the population
m	number of sample PSUs
nbar	number of sample secondary units per PSU
qbar	number of sample elements per SSU
k1	ratio of $B^2 + W^2$ to V . Default value is 1.
k2	ratio of $W_2^2 + W_3^2$ to V . Default value is 1.
delta1	measure of homogeneity between PSUs equal to $B^2/(B^2 + W^2)$ if PSUs are selected by simple random sampling; or, equal to $S_{U1(pwr)}^2$ divided by t_U^2 if PSUs are selected by <i>ppswr</i>
delta2	measure of homogeneity between SSUs within PSUs, equal to $W_2^2/(W_2^2 + W_3^2)$
Bsq	unit relvariance of PSU totals, equal to population variance of totals divided by \bar{t}_U^2
Wsq	within PSU relvariance of elements
W2sq	unit SSU relvariance
W3sq	unit element relvariance

Details

CVcalc3 computes the coefficient of variation of an estimated total for a three-stage sample. PSUs can be selected either with varying probabilities and with replacement or with equal probabilities and with replacement. SSUs and elements within SSUs are selected by simple random sampling. The *CV* formula is appropriate for approximating the relvariance of the probability-with-replacement (*pwr*)-estimator of a total when the same number of SSUs is selected in each PSU and the same number of elements is selected within each sample SSU.

Value

Value of the coefficient of variation of an estimated total

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

- Cochran, W.G. (1977, pp.308-310). *Sampling Techniques*. New York: John Wiley & Sons.
- Saerndal, C.E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.
- Valliant, R., Dever, J., Kreuter, F. (2013, sect. 9.2.4). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[CVcalc3](#)

Examples

```
CVcalc3(V=1, m=20 , nbar=5, qbar=10, delta1=0.02, delta2=0.10)
CVcalc3(V=1, m=20 , nbar=5, qbar=10, delta1=0.02, delta2=0.09)
CVcalc3(V=2, m=20 , nbar=5, qbar=10, k1=5, k2=10, Bsqr=1, Wsq=9, W2sq=2 , W3sq=18 )
```

Domainy1y2

Domain data

Description

Domainy1y2 is a small data file used for an exercise in sample size calculations.

Usage

```
data(Domainy1y2)
```

Format

A data frame with 30 observations on 2 variables.

y1 an artificial variable

y2 an artificial variable

References

- Valliant, R., Dever, J., Kreuter, F. (2013, chap. 3). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

Examples

```
data(Domainy1y2)
str(Domainy1y2)
summary(Domainy1y2)
```

dub	<i>Sample sizes for a double sampling design</i>
-----	--

Description

Compute samples sizes at each phase of a two-phase design where strata are created using the first phase.

Usage

```
dub(c1, c2, Ctot, Nh, Sh, Yh.bar)
```

Arguments

c1	cost per unit in phase-1
c2	cost per unit in phase-2
Ctot	Total variable cost
Nh	Vector of stratum population counts or proportions
Sh	Vector of stratum population standard deviations
Yh.bar	Vector of stratum population means

Details

Compute the first and second phase sample sizes for a double sampling design. A first phase sample is selected by simple random sampling (*srs*). Strata are formed based on information collected in the first phase. The Neyman allocation to strata of the second phase sample is computed ignoring costs. Optimal total sample sizes are computed for the first and second phases, given per-unit costs for the first and second phases and a fixed total budget for both phases combined.

Value

A list object with elements:

V1	Variance component associated with phase-1
V2	Variance component associated with phase-2
n1	Phase-1 sample size
n2	Total phase-2 sample across all strata
"n2/n1"	Fraction that phase-2 is of phase-1
ney.alloc	Vector of stratum sample sizes for phase-2 sample
Vopt	Variance of mean with the calculated phase-1 and phase-2 sample sizes
nsrs	Size of an <i>srs</i> that has cost Ctot, assuming each unit costs c2
Vsrs	Variance of mean in an <i>srs</i> of cost Ctot, assuming each unit costs c2
Vratio	Ratio of Vopt to Vsrs
Ctot	Input value of total cost
cost.chk	Computed value of phase-1 plus phase-2 sample with optimal sample sizes; should agree with Ctot

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

Cochran W (1977, sect. 12.3) *Sampling Techniques*. New York: John Wiley & Sons, Inc.

Neyman J (1938) Contribution to the theory of sampling human populations. *Journal of the American Statistical Association*, 33(201), 101-116.

Valliant, R., Dever, J., Kreuter, F. (2013, sect. 17.5.2). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

Examples

```
Wh <- rep(0.25,4)
Ph <- c(0.02,0.12,0.37,0.54)
Sh <- sqrt(Ph*(1-Ph))
c1 <- 10
c2 <- 50
Ctot <- 20000
dub(c1, c2, Ctot, Nh=Wh, Sh, Yh.bar=Ph)
```

gamEst

Estimate variance model parameter γ

Description

Regresses a y on a set of covariates X where $Var_M(y) = \sigma^2 x^\gamma$ and then regresses the squared residuals on $\log(x)$ to estimate γ .

Usage

```
gamEst(X1, x1, y1, v1)
```

Arguments

X1	matrix of predictors in the linear model for y1
x1	vector of x 's for individual units in the assumed specification of $Var_M(y)$
y1	vector of dependent variables for individual units
v1	vector proportional to $Var_M(y)$

Details

The function `gamEst` estimates the power γ in a model where the variance of the errors is proportional to x^γ for some covariate x . Values of γ are typically in $[0,2]$. The function is iteratively called by `gammaFit`, which is normally the function that an analyst should use.

Value

The estimate of γ .

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

Valliant, R., Dever, J., Kreuter, F. (2013, chap. 3). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[gammaFit](#)

Examples

```
data(hospital)
x <- hospital$x
y <- hospital$y

X <- cbind(sqrt(x), x)
gamEst(X1 = X, x1 = x, y1 = y, v1 = x)
```

gammaFit

Iteratively estimate variance model parameter γ

Description

Iteratively computes estimate of γ in a model with $E_M(y) = x^T \beta$ and $Var_M(y) = \sigma^2 x^\gamma$.

Usage

```
gammaFit(X, x, y, maxiter = 100, show.iter = FALSE, tol = 0.001)
```

Arguments

X	matrix of predictors in the linear model for y
x	vector of x 's for individual units in the assumed specification of $Var_M(y)$
y	vector of dependent variables for individual units
maxiter	maximum number of iterations allowed
show.iter	should values of γ be printed of each iteration? TRUE or FALSE
tol	size of relative difference in $\hat{\gamma}$'s between consecutive iterations used to determine convergence. Algorithm terminates when relative difference is less than tol.

Details

The function `gammaFit` estimates the power γ in a model where the variance of the errors is proportional to x^γ for some covariate x . Values of γ are typically in $[0,2]$. The function calls [gamEst](#).

Value

A list with the components:

<code>g.hat</code>	estimate of γ when iterative procedure stopped
<code>converged</code>	TRUE or FALSE depending on whether convergence was obtained
<code>steps</code>	number of steps used by the algorithm

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

Valliant, R., Dever, J., Kreuter, F. (2013, chap. 3). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[gamEst](#)

Examples

```
data(hospital)
x <- hospital$x
y <- hospital$y

X <- cbind(sqrt(x), x)
gammaFit(X = X, x = x, y = y, maxiter=100, tol=0.001)
```

HMT

Generate an HMT population

Description

Generate a population that follows the model in Hansen, Madow, and Tepping (1983)

Usage

```
HMT(N=5000, H=10)
```

Arguments

N	population size
H	number of strata

Details

HMT generates a population based on the model: $E(y) = \alpha + \beta x$, $Var(y) = \sigma^2 x^{3/2}$. Both x and y have gamma distributions. Strata are formed to have approximately the same total of x .

Value

N x 3 matrix with columns:

strat	stratum ID
x	auxiliary variable <i>x</i>
y	analysis variable <i>y</i>

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

Hansen, M.H., Madow, W.G., and Tepping, B.J. (1983). An evaluation of model-dependent and probability sampling inferences in sample surveys. *Journal of the American Statistical Association*, 78, 776-793.

Examples

```
# generate HMT population with 1000 units and 5 strata and plot results
pop <- HMT(N=1000, H=5)
plot(pop[, "x"], pop[, "y"])
```

hospital

Hospital Data

Description

The hospital data file is a national sample of short-stay hospitals with fewer than 1000 beds.

Usage

```
data(hospital)
```

Format

A data frame with 393 observations on the following 2 variables.

y Number of patients discharged by the hospital in January 1968

x Number of inpatient beds in the hospital

Details

The hospital data are from the National Hospital Discharge Survey conducted by the U.S. National Center for Health Statistics. The survey collects characteristics of inpatients discharged from non-Federal short-stay hospitals in the United States. This population is from the January 1968 survey and contains observations on 393 hospitals.

Source

National Center for Health Statistics Hospital Discharge Survey of 1968.

References

Herson, J. (1976). An Investigation of Relative Efficiency of Least-Squares Prediction to Conventional Probability Sampling Plans. *Journal of the American Statistical Association*, 71, 700-703.

Examples

```
data(hospital)
str(hospital)
```

labor	<i>Labor force population</i>
-------	-------------------------------

Description

A clustered population of persons extracted from the September 1976 Current Population Survey (CPS)

Usage

```
data(labor)
```

Format

A data frame with 478 persons on the following variables:

h stratum

cluster cluster (or segment) number

person person number

age age of person

agecat age category (1 = 19 years and under; 2 = 20-24; 3 = 25-34; 4 = 35-64; 5 = 65 years and over)

race race (1 = non-Black; 2 = Black)

sex Gender (1=Male; 2=Female)

HoursPerWk Usual number of hours worked per week

WklyWage Usual amount of weekly wages (in 1976 U.S. dollars)

y An artificial variable generated to follow a model with a common mean. Persons in the same cluster are correlated. Persons in different clusters are uncorrelated under the model.

Details

This population is a clustered population of 478 persons extracted from the September 1976 Current Population Survey (CPS) in the United States. The clusters are compact geographic areas used as one of the stages of sampling in the CPS and are typically composed of about 4 nearby households. The elements within clusters for this illustrative population are individual persons.

Source

Current Population Survey, <http://www.census.gov/cps/>

Examples

```
data(labor)
str(labor)
table(labor$h)
hist(labor$WklyWage)
```

MDarea.pop

*Maryland area population***Description**

An artificial population of census tracts, block groups, and persons

Usage

```
data(MDarea.pop)
```

Format

A data frame with 403,997 persons on the following variables:

PSU A grouping of block groups (BLKGROUP) which has about 5050 persons

SSU A grouping of block groups which has about 1010 persons

TRACT A geographic area defined by the Census Bureau. Tracts generally have between 1,500 and 8,000 people but have a much wider range in Anne Arundel county.

BLKGROUP Block group. A geographic area defined by the Census Bureau. Block groups generally have between 600 and 3,000 people.

Hispanic Hispanic ethnicity (1=Hispanic; 2=Non-Hispanic)

Gender Gender (1=Male; 2=Female)

Age 23 level age category (1 = Under 5 years; 2 = 5 to 9 years; 3 = 10 to 14 years; 4 = 15 to 17 years; 5 = 18 and 19 years; 6 = 20 years; 7 = 21 years; 8 = 22 to 24 years; 9 = 25 to 29 years; 10 = 30 to 34 years; 11 = 35 to 39 years; 12 = 40 to 44 years; 13 = 45 to 49 years; 14 = 50 to 54 years; 15 = 55 to 59 years; 16 = 60 and 61 years; 17 = 62 to 64 years; 18 = 65 and 66 years; 19 = 67 to 69 years; 20 = 70 to 74 years; 21 = 75 to 79 years; 22 = 80 to 84 years; 23 = 85 years and over)

person Counter for person within tract/block group/Hispanic/Gender/Age combination

y1 Artificial continuous variable

y2 Artificial continuous variable

y3 Artificial continuous variable

ins.cov Medical coverage (0 = person does not have medical insurance coverage; 1 = person has medical insurance coverage)

hosp.stay Overnight hospital stay (0 = person did not have an overnight hospital stay in last 12 months; 1 = person did have an overnight hospital stay in last 12 months)

Details

A dataset of 403,997 persons based on the 2000 decennial U.S. Census for Anne Arundel County in the state of Maryland. Person records were generated based on counts from the 2000 census. Individual values for each person were generated using models. Groupings to form the variables PSU and SSU were done after sorting the census file by tract and block group within tract.

Source

2000 U.S. decennial census, <http://www.census.gov/main/www/cen2000.html>

Examples

```
data(MDarea.pop)
str(MDarea.pop)
table(MDarea.pop$TRACT)
table(MDarea.pop$TRACT, MDarea.pop$Hispanic)
```

nCont	<i>Compute a simple random sample size for an estimated mean</i>
-------	--

Description

Compute a simple random sample size using either a target coefficient of variation, CV_0 , or target variance, V_0 , for an estimated mean.

Usage

```
nCont(CV0=NULL, V0=NULL, S2=NULL, ybarU=NULL, N=Inf, CVpop=NULL)
```

Arguments

CV0	target value of coefficient of variation of \bar{y}_s
V0	target value of variance of \bar{y}_s
S2	unit (population) variance
ybarU	population mean of target variable
N	number of units in finite population
CVpop	unit (population) coefficient of variation

Details

If CV_0 is the desired target, then the unit CV, CVpop, or the population mean and variance, ybarU and S2, must also be provided. If V_0 is the constrained value, then S2 must be also be included in the function call.

Value

numeric sample size

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

Valliant, R., Dever, J., Kreuter, F. (2013, chap. 3). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[nLogOdds](#), [nProp](#), [nPropMoe](#), [nWilson](#)

Examples

```
nCont(CV0=0.05, CVpop=2)
nCont(CV0=0.05, CVpop=2, N=500)
nCont(CV0=0.10/1.645, CVpop=1)

# Compute sample size for a ratio estimator in smho98 population
# Identify large units to select with certainty first
data(smho98)
cert <- smho98[,"BEDS"] > 2000
tmp <- smho98[!cert, ]
tmp <- tmp[tmp[, "BEDS"] > 0, ]

x <- tmp[, "BEDS"]
y <- tmp[, "EXPTOTAL"]
m <- lm(y ~ 0 + x, weights = 1/x)
ybarU <- mean(y)
S2R <- sum(m$residuals^2/(length(x)-1))
nCont(CV0=0.15, S2=S2R, ybarU=ybarU)
```

nDep2sam

Simple random sample size for difference in means

Description

Compute a simple random sample size for estimating the difference in means when samples overlap

Usage

```
nDep2sam(S2x, S2y, g, r, rho, alt, del, sig.level=0.05, pow=0.80)
```

Arguments

S2x	unit variance of analysis variable x in sample 1
S2y	unit variance of analysis variable y in sample 2
g	proportion of sample 1 that is in the overlap with sample 2
r	ratio of the size of sample 1 to that of sample 2
rho	unit-level correlation between x and y
alt	should the test be 1-sided or 2-sided; allowable values are <code>alt="one.sided"</code> or <code>alt="two.sided"</code> .
del	size of the difference between the means to be detected
sig.level	significance level of the hypothesis test
pow	desired power of the test

Details

nDep2sam computes sample sizes in two groups that are required for testing whether the difference in group means is significant. The power of the test is one of the input parameters. The samples have a specified proportion of units in common. Both samples are assumed to be selected via simple random sampling.

Value

List with values:

n1	sample size in group 1
n2	sample size in group 2
S2x, S2y	unit variances in groups 1 and 2
delta	difference in group means to be detected
gamma	proportion of sample 1 that is in the overlap with sample 2
r	ratio of the size of sample 1 to that of sample 2
rho	unit-level correlation between analysis variables in groups 1 and 2
alt	type of test: one-sided or two-sided
sig.level	significance level of test
power	power of the test

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

- Valliant, R., Dever, J., Kreuter, F. (2013, chap. 4). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.
- Woodward, M. (1992). Formulas for Sample Size, Power, and Minimum Detectable Relative Risk in Medical Studies. *The Statistician*, 41, 185-196.

See Also

[nProp2sam](#)

Examples

```
nDep2sam(S2x=200, S2y=200,
          g=0.75, r=1, rho=0.9,
          alt="one.sided", del=5,
          sig.level=0.05, pow=0.80)
```

nDomain	<i>Compute a simple random sample size for an estimated mean or total for a domain</i>
---------	--

Description

Compute a simple random sample size using either a target coefficient of variation, $CV_0(d)$, or target variance, $V_0(d)$, for an estimated mean or total for a domain.

Usage

```
nDomain(CV0d=NULL, V0d=NULL, S2d=NULL, ybarUd=NULL, N=Inf, CVpopd=NULL, Pd, est.type)
```

Arguments

CV0d	target value of coefficient of variation of estimated domain mean or total
V0d	target value of variance of estimated domain mean or total
S2d	unit (population) variance for domain units
ybarUd	population mean of target variable for domain units
N	number of units in full finite population (not just the domain population)
CVpopd	unit (population) coefficient of variation for domain units
Pd	proportion of units in the population that are in the domain
est.type	type of estimate; allowable values are "mean" or "total"

Details

If CV0d is the desired target, then the unit CV, CVpopd, or the domain population mean and variance, ybarUd and S2d, must also be provided. If V0d is the constrained value, then ybarUd must be also be included in the function call. CV0d will then be computed as $\sqrt{V0d}/ybarUd$.

Value

numeric sample size

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

Valliant, R., Dever, J., Kreuter, F. (2013, sec. 3.5.2). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[nCont](#), [nLogOdds](#), [nProp](#), [nPropMoe](#), [nWilson](#)

Examples

```
nDomain(CV0d=0.05, N=Inf, CVpopd=1, Pd=0.5, est.type="total")
nDomain(CV0d=0.05, N=Inf, CVpopd=1, Pd=0.5, est.type="mean")
nDomain(V0d=50, ybarUd=50, S2d=100, N=Inf, Pd=0.5, est.type="total")
nDomain(CV0d=0.05, ybarUd=50, S2d=100, N=Inf, Pd=0.5, est.type="total")
nDomain(CV0d=0.05, ybarUd=50, S2d=100, N=Inf, Pd=0.5, est.type="mean")
```

nhis

National Health Interview Survey: Demographic variables

Description

Demographic variables from a U.S. national household survey

Usage

```
data(nhis)
```

Format

A data frame with 3,911 observations on the following 16 variables.

ID Identification variable

stratum Sample design stratum (1-100)

psu Primary sampling unit, numbered within each stratum (1,2)

svywt survey weight

sex Gender (1 = male; 2 = female)

age Age, continuous

age_r Recoded age (3 = 18-24 years; 4 = 25-44 years; 5 = 45-64 years; 6 = 65-69 years; 7 = 70-74 years; 8 = 75 years and older)

hisp Hispanic ethnicity (1 = Hispanic; 2 = Non-Hispanic)

marital Marital status (1 = Separated; 2 = Divorced; 3 = Married; 4 = Single/never married; 5 = Widowed; 9 = Unknown marital status)

parents Parent(s) of sample person present in the family (1 = Mother, no father; 2 = Father, no mother; 3 = Mother and father; 4 = Neither mother nor father)

parents_r Parent(s) of sample person present in the family recode (1 = Yes; 2 = No)

educ Education (1 = 8th grade or less; 2 = 9-12th grade, no high school diploma; 3 = High school graduate; 4 = General education development (GED) degree recipient; 5 = Some college, no degree; 6 = Associate's degree, technical or vocational; 7 = Associate's degree, academic program; 8 = Bachelor's degree (BA, BS, AB, BBA); 9 = Master's, professional, or doctoral degree)

educ_r Education recode (1 = High school, general education development degree (GED), or less; 2 = Some college; 3 = Bachelor's or associate's degree; 4 = Master's degree & higher)

race Race (1 = White; 2 = Black; 3 = Other)

resp Respondent (0 = nonrespondent; 1 = respondent)

Details

The National Health Interview Survey (NHIS) is used to monitor health conditions in the U.S. Data are collected through personal household interviews. Only demographic variables are included in this subset which was collected in 2003. The nhis data set contains observations for 3,911 persons. The file contains only persons 18 years and older.

Source

National Health Interview Survey of 2003 conducted by the U.S. National Center for Health Statistics. <http://www.cdc.gov/nchs/nhis.htm>

See Also

[nhis.large](#)

Examples

```
data(nhis)
str(nhis)
table(nhis$sex, nhis$age_r)
```

nhis.large

National Health Interview Survey: Demographic and health variables

Description

Demographic and health related variables from a U.S. national household survey

Usage

```
data(nhis.large)
```

Format

A data frame with 21,588 observations on the following 18 variables.

ID Identification variable

stratum Sample design stratum (1-100)

psu Primary sampling unit, numbered within each stratum (1,2)

svywt survey weight

sex Gender (1 = male; 2 = female)

age.grp Age group (1 = < 18 years; 2 = 18-24 years; 3 = 25-44 years; 4 = 45-64 years; 5 = 65+)

hisp Hispanic ethnicity (1 = Hispanic; 2 = Non-Hispanic White; 3 = Non-Hispanic Black; 4 = Non-Hispanic All other race groups)

parents Parents present in the household (1 = mother, father, or both present; 2 = neither present)

educ Highest level of education attained (1 = High school graduate, graduate equivalence degree, or less; 2 = Some college; 3 = Bachelor's or associate's degree; 4 = Master's degree or higher; NA = missing)

race Race (1 = White; 2 = Black; 3 = All other race groups)

`inc.grp` Family income group (1 = < \$20K; 2 = \$20000-\$24999; 3 = \$25000-\$34999; 4 = \$35000-\$44999; 5 = \$45000-\$54999; 6 = \$55000-\$64999; 7 = \$65000-\$74999; 8 = \$75K+; NA = missing)

`delay.med` Delayed medical care in last 12 months because of cost (1 = Yes; 2 = No; NA = missing)

`hosp.stay` Had an overnight hospital stay in last 12 months (1 = Yes; 2 = No; NA = missing)

`doc.visit` During 2 WEEKS before interview, did person see a doctor or other health care professional at a doctor's office, a clinic, an emergency room, or some other place? (excluding overnight hospital stay)? (1 = Yes; 2 = No)

`medicaid` Covered by medicaid, a governmental subsidy program for the poor (1 = Yes; 2 = No; NA = missing)

`notcov` Not covered by any type of health insurance (1 = Yes; 2 = No; NA = missing)

`doing.lw` What was person doing last week? (1 = Working for pay at a job or business; 2 = With a job or business but not at work; 3 = Looking for work; 4 = Working, but not for pay, at a job or business; 5 = Not working and not looking for work; NA = missing)

`limited` Is the person limited in any way in any activities because of physical, mental or emotional problems? (1 = Limited in some way; 2 = Not limited in any way; NA = missing)

Details

The National Health Interview Survey (NHIS) is used to monitor health conditions in the U.S. Data are collected through personal household interviews. Demographic variables and a few health related variables are included in this subset. The `nhis.large` data set contains observations on 21,588 persons. `nhis.large` is a set of 21,588 persons extracted from the 2003 U.S. survey. The file contains only persons 18 years and older.

Source

National Health Interview Survey of 2003 conducted by the U.S. National Center for Health Statistics. <http://www.cdc.gov/nchs/nhis.htm>

See Also

[nhis](#)

Examples

```
data(nhis.large)
str(nhis.large)
table(nhis.large$stratum, nhis.large$psu)
table(nhis.large$delay.med, useNA="always")
```

nLogOdds

Calculate simple random sample size for estimating a proportion

Description

Calculate the simple random sample size for estimating a proportion using the log-odds transformation.

Usage

```
nLogOdds(moe.sw, e, alpha=0.05, pU, N=Inf)
```

Arguments

moe.sw	switch for setting desired margin of error (1 = CI half-width on the proportion; 2 = CI half-width on a proportion divided by pU)
e	desired margin of error
alpha	1 - (confidence level)
pU	population proportion
N	number of units in finite population

Details

The function accepts five parameters, which are the same ones as accepted by [nPropMoe](#). The desired margin of error can be specified as the CI half-width on the proportion (moe.sw=1) or as the CI half-width as a proportion of the population value pU (moe.sw=2).

Value

numeric sample size

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

Valliant, R., Dever, J., Kreuter, F. (2013, chap. 3). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[nProp](#), [nPropMoe](#), [nWilson](#), [nCont](#)

Examples

```
nLogOdds(moe.sw=1, e=0.05, alpha=0.05, pU=0.2, N=Inf)
nLogOdds(moe.sw=2, e=0.05, alpha=0.05, pU=0.2, N=Inf)
```

nProp

Compute simple random sample size for estimating a proportion

Description

Compute the simple random sample size for estimating a proportion based on different precision requirements.

Usage

```
nProp(CV0 = NULL, V0 = NULL, pU = NULL, N = Inf)
```

Arguments

CV0	target value of coefficient of variation of the estimated proportion
V0	target value of variance of the estimated proportion
pU	population proportion
N	number of units in finite population

Details

The precision requirement of p_s can be set based on either a target coefficient of variation, CV_0 , or a target variance, V_0 . In either case, a value of p_U must be supplied.

Value

numeric sample size

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

Valliant, R., Dever, J., Kreuter, F. (2013, chap. 3). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[nCont](#), [nLogOdds](#), [nPropMoe](#), [nWilson](#)

Examples

```
# srs sample size so that CV of estimated proportion is 0.05
# assuming the population is large and pU=0.01
# Both examples below are equivalent
nProp(V0=0.0005^2, N=Inf, pU=0.01) #or
nProp(CV0=0.05, N=Inf, pU=0.01)

# srswor sample size so that half-width of 2-sided 95% CI is 0.005
nProp(V0=(0.005/1.96)^2, N=Inf, pU=0.01)
```

nProp2sam

Simple random sample size for difference in proportions

Description

Compute a simple random sample size for estimating the difference in proportions when samples overlap

Usage

```
nProp2sam(px, py, pxy, g, r, alt, sig.level=0.05, pow=0.80)
```

Arguments

px	proportion in group 1
py	proportion in group 2
pxy	proportion in the overlap has the characteristic in both samples
g	proportion of sample 1 that is in the overlap with sample 2
r	ratio of the size of sample 1 to that of sample 2
alt	should the test be 1-sided or 2-sided; allowable values are alt="one.sided" or alt="two.sided".
sig.level	significance level of the hypothesis test
pow	desired power of the test

Details

nProp2sam computes sample sizes in two groups that are required for testing whether the difference in group proportions is significant. The power of the test is one of the input parameters. The samples have a specified proportion of units in common.

Value

List with values:

n1	sample size in group 1
n2	sample size in group 2
px.py.pxy	input values of the px, py, pxy parameters
gamma	proportion of sample 1 that is in the overlap with sample 2
r	ratio of the size of sample 1 to that of sample 2
alt	type of test: one-sided or two-sided
sig.level	significance level of test
power	power of the test

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

- Valliant, R., Dever, J., Kreuter, F. (2013, chap. 4). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.
- Woodward, M. (1992). Formulas for Sample Size, Power, and Minimum Detectable Relative Risk in Medical Studies. *The Statistician*, 41, 185-196.

See Also

[nDep2sam](#)

Examples

```
nProp2sam(px=0.5, py=0.55, pxy=0.45, g=0.5, r=1, alt="two.sided")
```

nPropMoe

*Simple random sample size for a proportion based on margin of error***Description**

Calculates a simple random sample size based on a specified margin of error.

Usage

```
nPropMoe(moe.sw, e, alpha = 0.05, pU, N = Inf)
```

Arguments

moe.sw	switch for setting desired margin of error (1 = CI half-width on the proportion; 2 = CI half-width on a proportion divided by p_U)
e	desired margin of error; either $e = z_{1-\alpha/2} \sqrt{V(\bar{y}_s)}$ or $e = z_{1-\alpha/2} CV(\bar{y}_s)$
alpha	1 - (confidence level)
pU	population proportion
N	number of units in finite population

Details

The margin of error can be set as the half-width of a normal approximation confidence interval, $e = z_{1-\alpha/2} \sqrt{V(\bar{y}_s)}$, or as the half-width of a normal approximation confidence interval divided by the population proportion, $e = z_{1-\alpha/2} CV(\bar{y}_s)$. The type of margin of error is selected by the parameter `moe.sw` where `moe.sw=1` sets $e = z_{1-\alpha/2} \sqrt{V(\bar{y}_s)}$ and `moe.sw=2` sets i.e., $e = \frac{z_{1-\alpha/2} \sqrt{V(\bar{y}_s)}}{p_U}$.

Value

numeric sample size

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

Valliant, R., Dever, J., Kreuter, F. (2013, chap. 3). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

See Also

[nCont](#), [nLogOdds](#), [nProp](#), [nWilson](#)

Examples

```
# srs sample size so that half-width of a 95% CI is 0.01
# population is large and population proportion is 0.04
nPropMoe(moe.sw=1, e=0.01, alpha=0.05, pU=0.04, N=Inf)

# srswor sample size for a range of margins of error defined as
# half-width of a 95% CI
nPropMoe(moe.sw=1, e=seq(0.01,0.08,0.01), alpha=0.05, pU=0.5)

# srswor sample size for a range of margins of error defined as
# the proportion that the half-width of a 95% CI is of pU
nPropMoe(moe.sw=2, e=seq(0.05,0.1,0.2), alpha=0.05, pU=0.5)
```

NRFUopt

Sample sizes for a nonresponse follow-up study

Description

Compute optimal values of the first-phase sample size and the second-phase sampling fraction in a two-phase sample.

Usage

```
NRFUopt(Ctot=NULL, c1, c2, theta, CV0=NULL, CVpop=NULL, N=Inf, type.sw)
```

Arguments

Ctot	total variable cost
c1	cost per unit in phase-1
c2	cost per unit in phase-2
theta	probability of response for each unit
CV0	target coefficient of variation for the estimated total or mean
CVpop	Unit coefficient of variation
N	Population size; default is Inf
type.sw	type of allocation; "cost" = target total variable cost, "cv" = target coefficient of variation

Details

NRFUopt computes the optimal values of the first-phase sample size and the second-phase sampling fraction in a two-phase sample. Both stages are assumed to be selected using simple random sampling without replacement. If type.sw="cost", the optima are computed for a target total, expected cost across both phases. If type.sw="cv", the optima are computed for a target coefficient of variation for an estimated mean.

Value

List object with values:

allocation	type of allocation: either "fixed cost" or "fixed CV"
"Total variable cost"	expected total cost: fixed cost if type.sw="cost" or computed cost if type.sw="cv"
"Response rate"	first-phase response rate
CV	anticipated coefficient of variation (CV) if type.sw="cost" or target CV if type.sw="cv"
v.opt	optimal fraction of first-phase nonrespondents to select for second-phase follow-up
n1.opt	optimal number of units to sample at first-phase
"Expected n2"	expected number of respondents obtained at second-phase
"Expected total cases (2-phase)"	expected number of respondents across both phases
"srs sample for same cv"	size of single-phase simple random sample (srs) needed to obtain same CV as the two-phase sample
"Cost Ratio: Two phase to srs"	ratio of expected cost for two-phase sample to cost of single-phase srs

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

Saerndal, C.E., Swensson, B., and Wretman, J. (1992, examples 15.4.4 and 15.4.5). *Model Assisted Survey Sampling*. New York: Springer.

Valliant, R., Dever, J., Kreuter, F. (2013, chap.17). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

Examples

```
# optima for fixed target CV
NRFUopt(Ctot=NULL, c1=50, c2=200, theta=0.5, CV0=0.05, CVpop=1, type.sw = "cv")
# optima for fixed total cost
NRFUopt(Ctot=100000, c1=50, c2=200, theta=0.5, CV0=NULL, CVpop=1, type.sw = "cost")
```


nWilson

*Calculate a simple random sample size for estimating a proportion***Description**

Calculate a simple random sample size for estimating a proportion using the Wilson method.

Usage

```
nWilson(moe.sw, alpha = 0.05, pU, e)
```

Arguments

moe.sw	switch for setting desired margin of error (1 = CI half-width on the proportion; 2 = CI half-width on a proportion divided by p_U)
alpha	1 - (confidence level)
pU	population proportion
e	desired margin of error; either the value of CI half-width or the value of the half-width divided by p_U

Details

Calculate a simple random sample size using the Wilson (1927) method. A margin of error can be set as the CI half-width on the proportion (moe.sw=1) or as the CI half-width as a proportion of the population value p_U (moe.sw=2).

Value

n.sam	numeric sample size
"CI lower limit"	lower limit of Wilson confidence interval with computed sample size
"CI upper limit"	upper limit of Wilson confidence interval with computed sample size
"length of CI"	length of Wilson confidence interval with computed sample size

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

Valliant, R., Dever, J., Kreuter, F. (2013, chap. 3). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

Wilson, E.B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22, 209-212.

See Also

[nCont](#), [nLogOdds](#), [nProp](#), [nPropMoe](#)

Examples

```
# srs sample size using Wilson method so that half-width of a 95% CI
# is 0.01. Population proportion is 0.04
nWilson(moe.sw = 1, pU = 0.04, e = 0.01)
```

smho.N874

Survey of Mental Health Organizations Data

Description

Data from the 1998 Survey of Mental Health Organizations (SMHO)

Usage

```
data(smho.N874)
```

Format

A data frame with 874 observations on the following 6 variables.

EXPTOTAL Total expenditures in 1998

BEDS Total inpatient beds

SEENCNT Unduplicated client/patient count seen during year

EOYCNT End of year count of patients on the role

FINDIRCT Hospital receives money from the state mental health agency (1=Yes; 2=No)

hosp.type Hospital type (1 = Psychiatric; 2 = Residential or veterans; 3 = General; 4 = Outpatient, partial care; 5 = Multi-service, substance abuse)

Details

The 1998 SMHO was conducted by the U.S. Substance Abuse and Mental Health Services Administration. It collected data on mental health care organizations and general hospitals that provide mental health care services, with an objective to develop national and state level estimates for total expenditure, full time equivalent staff, bed count, and total caseload by type of organization. The population omits one extreme observation in the smho98 population and has fewer variables than smho98. smho.N874 contains observations on 874 facilities.

Source

Substance Abuse and Mental Health Services Administration

References

Manderscheid, R.W. and Henderson, M.J. (2002). Mental Health, United States, 2002.

DHHS Publication No. SMA04-3938. Rockville MD USA: Substance Abuse and Mental Health Services Administration.

Available at <http://mentalhealth.samhsa.gov/publications/allpubs/SMA04-3938/AppendixA.asp>

See Also[smho98](#)**Examples**

```
data(smho.N874)
str(smho.N874)
```

smho98

*Survey of Mental Health Organizations Data***Description**

Data from the 1998 Survey of Mental Health Organizations (SMHO)

Usage

```
data(smho98)
```

Format

A data frame with 875 observations on the following variables.

STRATUM Sample design stratum (1 = Psychiatric Hospital, private; 2 = Psychiatric Hospital, public; 3 = Residential, children; 4 = Residential, adults; 5 = General Hospital, public, inpatient or residential care; 6 = General Hospital, public, outpatient care only; 7 = General Hospital, private, inpatient or residential care; 8 = General Hospital, private, outpatient care only; 9 = Military Veterans, inpatient or residential care; 10 = Military Veterans, outpatient care only; 11 = Partial Care 12 = Outpatient care, private; 13 = Outpatient care, public; 14 = Multi-service, private; 15 = Multi-service, public; 16 = Substance Abuse)

BEDS Total inpatient beds

EXPTOTAL Total expenditures in 1998

SEENCNT Unduplicated client/patient count seen during year

EOYCNT End of year count of patients on the role

Y_IP Number of inpatient visits during year

OPCSFRST Number of outpatients on the rolls on the first day of the reporting year

OPCSADDS Number of outpatients admitted, readmitted, or transferred to the organization during the reporting year for less than a 24 hour period and not overnight

OPCSVIST Number of outpatient visits during the reporting year for less than a 24 hour period and not overnight

EMGWALK Number of emergency walk-ins during the reporting year

PSYREHAB Number of visits for psychiatric rehabilitation services

IPCSADDS Number of residential patients added during the reporting year or patients admitted for more than a 24 hour period

Details

The 1998 SMHO was conducted by the U.S. Substance Abuse and Mental Health Services Administration. It collected data on mental health care organizations and general hospitals that provide mental health care services, with an objective to develop national and state level estimates for total expenditure, full time equivalent staff, bed count, and total caseload by type of organization.

Source

Substance Abuse and Mental Health Services Administration

References

Manderscheid, R.W. and Henderson, M.J. (2002). Mental Health, United States, 2002.

DHHS Publication No. SMA04-3938. Rockville MD USA: Substance Abuse and Mental Health Services Administration.

Available at <http://mentalhealth.samhsa.gov/publications/allpubs/SMA04-3938/AppendixA.asp>

See Also

[smho.N874](#)

Examples

```
data(smho98)
str(smho98)
summary(smho98)
```

strAlloc

Allocate a sample to strata

Description

Compute the proportional, Neyman, cost-constrained, and variance-constrained allocations in a stratified simple random sample.

Usage

```
strAlloc(n.tot = NULL, Nh = NULL, Sh = NULL, cost = NULL, ch = NULL,
         V0 = NULL, CV0 = NULL, ybarU = NULL, alloc)
```

Arguments

n.tot	fixed total sample size
Nh	vector of population stratum sizes (N_h) or pop stratum proportions (W_h)
Sh	stratum unit standard deviations (S_h), required unless alloc = "prop"
cost	total variable cost
ch	vector of costs per unit in stratum h (c_h)
V0	fixed variance target for estimated mean

CV0	fixed CV target for estimated mean
ybarU	population mean of y (\bar{y}_U)
alloc	type of allocation; must be one of "prop", "neyman", "totcost", "totvar"

Details

alloc="prop" computes the proportional allocation of a fixed total sample size, n.tot, to the strata. alloc="neyman" computes the allocation of a fixed total sample size, n.tot, to the strata that minimizes the variance of an estimated mean. alloc="totcost" computes the allocation of a fixed total sample size, n.tot, to the strata that minimizes the variance of an estimated mean subject to the fixed total cost. alloc="totvar" computes the allocation that minimizes total cost subject to the target coefficient of variation, CV0, or the target variance, V0, of the estimated mean.

Value

List with values:

alloc	type of allocation: "prop", "neyman", "totcost", "totvar"
nh	vector of stratum sample sizes
"nh/n"	proportion of sample allocated to each stratum

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

- Cochran, W.G. (1977). *Sampling Techniques*. John Wiley & Sons.
- Valliant, R., Dever, J., Kreuter, F. (2013, chap. 3). *Practical Tools for Designing and Weighting Survey Samples*. Springer.

See Also

[nCont](#), [nLogOdds](#), [nProp](#), [nPropMoe](#), [nWilson](#)

Examples

```
# Neyman allocation
Nh <- c(215, 65, 252, 50, 149, 144)
Sh <- c(26787207, 10645109, 6909676, 11085034, 9817762, 44553355)
strAlloc(n.tot = 100, Nh = Nh, Sh = Sh, alloc = "neyman")

# cost constrained allocation
ch <- c(1400, 200, 300, 600, 450, 1000)
strAlloc(Nh = Nh, Sh = Sh, cost = 100000, ch = ch, alloc = "totcost")

# allocation with CV target of 0.05
strAlloc(Nh = Nh, Sh = Sh, CV0 = 0.05, ch = ch, ybarU = 11664181, alloc = "totvar")
```

wtdvar	<i>Compute weighted variance</i>
--------	----------------------------------

Description

Compute an estimate of a population unit variance from a complex sample with survey weights.

Usage

```
wtdvar(x,w)
```

Arguments

x	data vector
w	vector of survey weights; must be same length as x

Details

wtdvar is also used by [BW3stagePPSe](#) in estimating relvariance components.

Value

numeric estimate of population unit variance

Author(s)

Richard Valliant, Jill A. Dever, Frauke Kreuter

References

Valliant, R., Dever, J., Kreuter, F. (2013, chap. 9). *Practical Tools for Designing and Weighting Survey Samples*. Springer.

Examples

```
x <- c(1:3)
wts <- c(4, 6, 8)
wtdvar(x=x, w=wts)
```

Index

*Topic **datasets**

Domainy1y2, [20](#)
hospital, [25](#)
labor, [26](#)
MDarea.pop, [27](#)
nhis, [32](#)
nhis.large, [33](#)
smho.N874, [42](#)
smho98, [43](#)

*Topic **methods**

BW2stagePPS, [2](#)
BW2stagePPSe, [4](#)
BW2stageSRS, [6](#)
BW3stagePPS, [7](#)
BW3stagePPSe, [9](#)
clusOpt2, [11](#)
clusOpt2fixedPSU, [13](#)
clusOpt3, [14](#)
clusOpt3fixedPSU, [16](#)
CVcalc2, [17](#)
CVcalc3, [19](#)
dub, [21](#)
HMT, [24](#)
nCont, [28](#)
nDep2sam, [29](#)
nDomain, [31](#)
nLogOdds, [34](#)
nProp, [35](#)
nProp2sam, [36](#)
nPropMoe, [38](#)
NRFUopt, [39](#)
nWilson, [41](#)
strAlloc, [44](#)
wtdvar, [46](#)

*Topic **models**

gamEst, [22](#)
gammaFit, [23](#)

*Topic **survey**

BW2stagePPS, [2](#)
BW2stagePPSe, [4](#)
BW2stageSRS, [6](#)
BW3stagePPS, [7](#)
BW3stagePPSe, [9](#)

clusOpt2, [11](#)
clusOpt2fixedPSU, [13](#)
clusOpt3, [14](#)
clusOpt3fixedPSU, [16](#)
CVcalc2, [17](#)
CVcalc3, [19](#)
dub, [21](#)
HMT, [24](#)
nCont, [28](#)
nDep2sam, [29](#)
nDomain, [31](#)
nLogOdds, [34](#)
nProp, [35](#)
nProp2sam, [36](#)
nPropMoe, [38](#)
NRFUopt, [39](#)
nWilson, [41](#)
strAlloc, [44](#)
wtdvar, [46](#)

BW2stagePPS, [2](#), [5](#), [6](#), [8](#), [10](#)
BW2stagePPSe, [2](#), [3](#), [4](#), [6–8](#), [10](#)
BW2stageSRS, [3](#), [5](#), [6](#), [8](#), [10](#)
BW3stagePPS, [3](#), [5](#), [6](#), [7](#), [10](#)
BW3stagePPSe, [3](#), [5](#), [6](#), [8](#), [9](#), [46](#)

clusOpt2, [11](#), [14](#), [15](#), [17](#)
clusOpt2fixedPSU, [12](#), [13](#), [15](#), [17](#)
clusOpt3, [12](#), [14](#), [14](#), [17](#)
clusOpt3fixedPSU, [12](#), [14](#), [15](#), [16](#)
CVcalc2, [17](#)
CVcalc3, [18](#), [19](#), [20](#)

Domainy1y2, [20](#)
dub, [21](#)

gamEst, [22](#), [23](#), [24](#)
gammaFit, [22](#), [23](#), [23](#)

HMT, [24](#)
hospital, [25](#)

labor, [26](#)

MDarea.pop, [27](#)

nCont, [28](#), [31](#), [35](#), [36](#), [38](#), [41](#), [45](#)
nDep2sam, [29](#), [37](#)
nDomain, [31](#)
nhis, [32](#), [34](#)
nhis.large, [33](#), [33](#)
nLogOdds, [29](#), [31](#), [34](#), [36](#), [38](#), [41](#), [45](#)
nProp, [29](#), [31](#), [35](#), [35](#), [38](#), [41](#), [45](#)
nProp2sam, [30](#), [36](#)
nPropMoe, [29](#), [31](#), [35](#), [36](#), [38](#), [41](#), [45](#)
NRFUopt, [39](#)
nWilson, [29](#), [31](#), [35](#), [36](#), [38](#), [41](#), [45](#)

smho.N874, [42](#), [44](#)
smho98, [43](#), [43](#)
strAlloc, [44](#)

wtdvar, [46](#)