

# Package ‘spsurvey’

May 22, 2015

**Version** 3.0

**Date** 2015-05-21

**Title** Spatial Survey Design and Analysis

**Depends** R (>= 2.10), sp

**Imports** methods, deldir, foreign, MASS, rgeos

**Description** This group of functions implements algorithms for design and analysis of probability surveys. The functions are tailored for Generalized Random Tessellation Stratified survey designs.

**License** GPL (>= 2)

**URL** <http://www.epa.gov/nheerl/arm/>

**NeedsCompilation** yes

**Author** Tom Kincaid [aut, cre],  
Tony Olsen [aut],  
Don Stevens [ctb],  
Christian Platt [ctb],  
Denis White [ctb],  
Richard Remington [ctb]

**Maintainer** Tom Kincaid <Kincaid.Tom@epa.gov>

**Repository** CRAN

**Date/Publication** 2015-05-22 20:08:07

## R topics documented:

adjwgt . . . . .	3
albersgeod . . . . .	4
ash1.wgt . . . . .	5
attrisk.analysis . . . . .	6
attrisk.est . . . . .	8
cat.analysis . . . . .	12
category.est . . . . .	15
cdf.decon . . . . .	18

cdf.est . . . . .	22
cdf.plot . . . . .	26
cdf.test . . . . .	28
change.analysis . . . . .	33
change.est . . . . .	37
cont.analysis . . . . .	41
cont.cdfplot . . . . .	45
cont.cdfptest . . . . .	47
decon_data . . . . .	51
dsgnsum . . . . .	52
FL_lakes . . . . .	53
framesum . . . . .	54
geodalbers . . . . .	56
grts . . . . .	57
grtsarea . . . . .	60
grtslin . . . . .	61
grtspts . . . . .	62
IN_streams . . . . .	64
irs . . . . .	64
irsarea . . . . .	67
irslin . . . . .	68
irspts . . . . .	69
Luck_Ash_streams . . . . .	70
marinus . . . . .	70
NE_lakes . . . . .	71
NLA_2007 . . . . .	72
read.dbf . . . . .	73
read.sas . . . . .	74
read.shape . . . . .	75
relrisk.analysis . . . . .	76
relrisk.est . . . . .	79
SC_estuaries . . . . .	82
sp2shape . . . . .	83
SpatialDesign . . . . .	84
SpatialDesign-class . . . . .	84
spbalance . . . . .	86
spsurvey.analysis . . . . .	88
total.est . . . . .	93
uniqueID . . . . .	96
UT_ecoregions . . . . .	97
write.dbf . . . . .	97
write.object . . . . .	98

---

`adjwgt`*Adjust Initial Survey Design Weights*

---

## Description

This function adjusts initial survey design weights when implementation results in use of oversample sites or when it is desired to have final weights sum to a known frame size. Adjusted weights are equal to initial weight times the frame size divided by the sum of the initial weights. The adjustment is done separately for each category specified in argument `wtcat`.

## Usage

```
adjwgt(sites, wgt, wtcat, framesize)
```

## Arguments

<code>sites</code>	the logical value for each site, where <code>TRUE</code> = include the site and <code>FALSE</code> = do not include the site.
<code>wgt</code>	the initial weight (inverse of the sample inclusion probability) for each site.
<code>wtcat</code>	the weight adjustment category name for each site.
<code>framesize</code>	the known size of the frame for each category name in <code>wtcat</code> , which must have the names attribute set to match the category names used in <code>wtcat</code> .

## Value

A vector of adjusted weights, where the adjusted weight is set to zero for sites that have the logical value in the `sites` argument set to `FALSE`.

## Author(s)

Tony Olsen <Olsen.Tony@epa.gov>

## Examples

```
sites <- as.logical(rep(rep(c("TRUE", "FALSE"), c(9,1)), 5))
wgt <- runif(50, 10, 100)
wtcat <- rep(c("A", "B"), c(30, 20))
framesize <- c(1650, 1100)
names(framesize) <- c("A", "B")
adjwgt(sites, wgt, wtcat, framesize)
```

---

albersgeod                      *Project Albers Projection in Plane to Latitude and Longitude (Spheroid)*

---

### Description

Project Albers projection in the plane to spheroid models of the globe.

### Usage

```
albersgeod(x, y, sph="GRS80", clon=-96, clat=23, sp1=29.5, sp2=45.5)
```

### Arguments

x	Albers x-coordinate vector to be projected to latitude/longitude.
y	Albers y-coordinate vector to be projected to latitude/longitude.
sph	spheroid options: Clarke1866, GRS80, WGS84. The default is GRS80.
clon	center longitude (decimal degrees). The default is -96.
clat	origin latitude (decimal degrees). The default is 23.
sp1	standard parallel 1 (decimal degrees). The default is 29.5.
sp2	standard parallel 2 (decimal degrees). The default is 45.5.

### Details

Ask Denis White.

### Value

A data frame of latitude and longitude projections for Albers x-coordinates and y-coordinates.

### Author(s)

Tony Olsen <Olsen.Tony@epa.gov>

### References

J. Snyder, USGS Professional Paper 1395

---

`ash1.wgt`*Compute the Average Shifted Histogram for Weighted Data*

---

**Description**

This function computes the average shifted histogram (ASH) for weighted data.

**Usage**

```
ash1.wgt(x, wgt=rep(1,length(x)), m=5, nbin=50, ab=NULL, support="Continuous")
```

**Arguments**

<code>x</code>	vector of data to be used to estimate density. NAs are allowed.
<code>wgt</code>	vector of weights for each observation from a probability sample. The default is equal weights (equal probability).
<code>m</code>	the number of empty bins to add to the ends when the range is not completely specified. The default is 5.
<code>nbin</code>	the number of bins for density estimation. the default is 50.
<code>ab</code>	optional range for support associated with the density. Both values may be equal to NA. If equal to NA, then corresponding limit will be based on <code>nicerange()</code> . the default is NULL.
<code>support</code>	the type of support. If equal to "Continuous", then data are from a continuous distribution. If equal to "Ordinal", then data are from a discrete distribution defined for integers only. The default is "Continuous".

**Value**

A vector containing the ASH density estimate.

**Author(s)**

Tony Olsen <Olsen.Tony@epa.gov>

**Examples**

```
x <- rnorm(100, 10, sqrt(10))
wgt <- runif(100, 10, 100)
rslt <- ash1.wgt(x, wgt)
plot(rslt)
```

---

attrisk.analysis      *Attributable Risk Analysis for Probability Survey Data*

---

### Description

This function organizes input and output for attributable risk analysis of categorical data generated by a probability survey.

### Usage

```
attrisk.analysis(sites=NULL, subpop=NULL, design, data.ar, response.var,
  stressor.var, response.levels=rep(list(c("Poor", "Good")),
  length(response.var)), stressor.levels=rep(list(c("Poor", "Good")),
  length(stressor.var)), popcorrect=FALSE, pcfsiz=NULL, N.cluster=NULL,
  stage1size=NULL, sizeweight=FALSE, vartype="Local", conf=95)
```

### Arguments

sites	a data frame consisting of two variables: the first variable is site IDs, and the second variable is a logical vector indicating which sites to use in the analysis. The default is NULL.
subpop	a data frame describing sets of populations and subpopulations for which estimates will be calculated. The first variable is site IDs. Each subsequent variable identifies a Type of population, where the variable name is used to identify Type. A Type variable identifies each site with one of the subpopulations of that Type. The default is NULL.
design	a data frame consisting of design variables. Variables should be named as follows: siteID = site IDs wgt = final adjusted weights, which are either the weights for a single-stage sample or the stage two weights for a two-stage sample xcoord = x-coordinates for location, which are either the x-coordinates for a single-stage sample or the stage two x-coordinates for a two-stage sample ycoord = y-coordinates for location, which are either the y-coordinates for a single-stage sample or the stage two y-coordinates for a two-stage sample stratum = the stratum codes cluster = the stage one sampling unit (primary sampling unit or cluster) codes wgt1 = final adjusted stage one weights xcoord1 = the stage one x-coordinates for location ycoord1 = the stage one y-coordinates for location support = support values - the value one (1) for a site from a finite resource or the measure of the sampling unit associated with a site from an extensive resource, which is required for calculation of finite and continuous population correction factors swgt = size-weights, which is the stage two size-weight for a two- stage sample swgt1 = stage one size-weights

<code>data.ar</code>	data frame of categorical response and stressor variables, where each variable consists of two categories. If response or stressor variables include more than two categories, occurrences of those categories must be removed or replaced with missing values. The first column of this argument is site IDs. Subsequent columns are response and stressor variables. Missing data (NA) is allowed.
<code>response.var</code>	character vector providing names of columns in argument <code>data.ar</code> that contain a response variable, where names may be repeated. Each name in this argument is matched with the corresponding value in the <code>stressor.var</code> argument.
<code>stressor.var</code>	character vector providing names of columns in argument <code>data.ar</code> that contain a stressor variable, where names may be repeated. Each name in this argument is matched with the corresponding value in the <code>response.var</code> argument. This argument must be the same length as argument <code>response.var</code> .
<code>response.levels</code>	list providing the category values (levels) for each element in the <code>response.var</code> argument. This argument must be the same length as argument <code>response.var</code> . The default is a list containing the values "Poor" and "Good" for the first and second levels, respectively, of each element in the <code>response.var</code> argument.
<code>stressor.levels</code>	list providing the category values (levels) for each element in the <code>stressor.var</code> argument. This argument must be the same length as argument <code>response.var</code> . The default is a list containing the values "Poor" and "Good" for the first and second levels, respectively, of each element in the <code>stressor.var</code> argument.
<code>popcorrect</code>	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation, where TRUE = use the correction factors and FALSE = do not use the correction factors. The default is FALSE.
<code>pcfsize</code>	size of the resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample this argument must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
<code>N.cluster</code>	the number of stage one sampling units in the resource, which is required for calculation of finite and continuous population correction factors for a two-stage sample. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
<code>stage1size</code>	size of the stage one sampling units of a two-stage sample, which is required for calculation of finite and continuous population correction factors for a two-stage sample and must have the names attribute set to identify the stage one sampling unit codes. For a stratified sample, the names attribute must be set to identify both stratum codes and stage one sampling unit codes using a convention where the two codes are separated by the & symbol, e.g., "Stratum 1&Cluster 1". The default is NULL.
<code>sizeweight</code>	a logical value that indicates whether size-weights should be used in the analysis, where TRUE = use the size-weights and FALSE = do not use the size-weights. The default is FALSE.

vartype	the choice of variance estimator, where "Local" = local mean estimator and "SRS" = SRS estimator. The default is "Local".
conf	the confidence level. The default is 95%.

### Value

Value is a data frame of attributable risk estimates for all combinations of population Types, sub-populations within Types, and response variables. Standard error and confidence interval estimates also are provided.

### Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>

### References

Sarndal, C.E., B. Swensson, and J. Wretman. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

### See Also

[attrisk.est](#)

### Examples

```
mysiteID <- paste("Site", 1:100, sep="")
mysites <- data.frame(siteID=mysiteID, Active=rep(TRUE, 100))
mysubpop <- data.frame(siteID=mysiteID, All.Sites=rep("All Sites", 100),
  Resource.Class=rep(c("Agr", "Forest"), c(55,45)))
mydesign <- data.frame(siteID=mysiteID, wgt=runif(100, 10, 100),
  xcoord=runif(100), ycoord=runif(100), stratum=rep(c("Stratum1",
  "Stratum2"), 50))
mydata.ar <- data.frame(siteID=mysiteID, RespVar1=sample(c("Poor", "Good"),
  100, replace=TRUE), RespVar2=sample(c("Poor", "Good"), 100, replace=TRUE),
  StressVar=sample(c("Poor", "Good"), 100, replace=TRUE), wgt=runif(100, 10,
  100))
attrisk.analysis(sites=mysites, subpop=mysubpop, design=mydesign,
  data.ar=mydata.ar, response.var=c("RespVar1", "RespVar2"),
  stressor.var=rep("StressVar", 2))
```



## Description

This function calculates the attributable risk estimate for a 2x2 table of cell counts defined by a categorical response variable and a categorical explanatory (stressor) variable for an unequal probability design. The attributable risk of the stressor variable is the percent reduction in the first level of the response variable that would result from elimination of the stressor variable. The standard error of the base e log of the attributable risk estimate and confidence limits for the estimate also are calculated.

## Usage

```
attrisk.est(response, stressor, response.levels=c("Poor", "Good"),
  stressor.levels=c("Poor", "Good"), wgt, xcoord=NULL, ycoord=NULL,
  stratum=NULL, cluster=NULL, wgt1=NULL, xcoord1=NULL, ycoord1=NULL,
  popcorrect=FALSE, pcFSIZE=NULL, N.cluster=NULL, stage1size=NULL,
  support=NULL, sizeweight=FALSE, swgt=NULL, swgt1=NULL, vartype="Local",
  conf=95, check.ind=TRUE, warn.ind=NULL, warn.df=NULL, warn.vec=NULL)
```

## Arguments

response	the categorical response variable values.
stressor	the categorical explanatory (stressor) variable values.
response.levels	category values (levels) for the categorical response variable, where the first level is used for calculating the numerator and the denominator of the attributable risk estimate. If response.levels is not supplied, then values "Poor" and "Good" are used for the first level and second level of the response variable, respectively. The default is c("Poor", "Good").
stressor.levels	category values (levels) for the categorical stressor variable, where the first level is used for calculating the numerator of the attributable risk estimate and the second level is used for calculating the denominator of the estimate. If stressor.levels is not supplied, then values "Poor" and "Good" are used for the first level and second level of the stressor variable, respectively. The default is c("Poor", "Good").
wgt	the final adjusted weight (inverse of the sample inclusion probability) for each site, which is either the weight for a single-stage sample or the stage two weight for a two-stage sample.
xcoord	x-coordinate for location for each site, which is either the x-coordinate for a single-stage sample or the stage two x-coordinate for a two-stage sample. The default is NULL.
ycoord	y-coordinate for location for each site, which is either the y-coordinate for a single-stage sample or the stage two y-coordinate for a two-stage sample. The default is NULL.
stratum	the stratum for each site. The default is NULL.
cluster	the stage one sampling unit (primary sampling unit or cluster) code for each site. The default is NULL.

wgt1	the final adjusted stage one weight for each site. The default is NULL.
xcoord1	the stage one x-coordinate for location for each site. The default is NULL.
ycoord1	the stage one y-coordinate for location for each site. The default is NULL.
popcorrect	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation, where TRUE = use the correction factors and FALSE = do not use the correction factors. The default is FALSE.
pcfsize	size of the resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample this argument must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
N.cluster	the number of stage one sampling units in the resource, which is required for calculation of finite and continuous population correction factors for a two-stage sample. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
stage1size	size of the stage one sampling units of a two-stage sample, which is required for calculation of finite and continuous population correction factors for a two-stage sample and must have the names attribute set to identify the stage one sampling unit codes. For a stratified sample, the names attribute must be set to identify both stratum codes and stage one sampling unit codes using a convention where the two codes are separated by the & symbol, e.g., "Stratum 1&Cluster 1". The default is NULL.
support	the support value for each site - the value one (1) for a site from a finite resource or the measure of the sampling unit associated with a site from an extensive resource, which is required for calculation of finite and continuous population correction factors. The default is NULL.
sizeweight	a logical value that indicates whether size-weights should be used in the analysis, where TRUE = use the size-weights and FALSE = do not use the size-weights. The default is FALSE.
swgt	the size-weight for each site, which is the stage two size-weight for a two-stage sample. The default is NULL.
swgt1	the stage one size-weight for each site. The default is NULL.
vartype	the choice of variance estimator, where "Local" = local mean estimator and "SRS" = SRS estimator. The default is "Local".
conf	the confidence level. The default is 95%.
check.ind	a logical value that indicates whether compatibility checking of the input values is conducted, where TRUE = conduct compatibility checking and FALSE = do not conduct compatibility checking. The default is TRUE.
warn.ind	a logical value that indicates whether warning messages were generated, where TRUE = warning messages were generated and FALSE = warning messages were not generated. The default is NULL.
warn.df	a data frame for storing warning messages. The default is NULL.
warn.vec	a vector that contains names of the population type, the subpopulation, and an indicator. The default is NULL.

## Details

The attributable risk estimate is computed using the 2x2 table of cell counts defined by a categorical response variable and a categorical stressor variable (Van Sickle and Paulsen, 2008). Cell totals are estimated using the Horvitz-Thompson estimator. The standard error of the base e log of the attributable risk estimate is calculated using a first-order Taylor series linearization (Sarndal et al., 1992).

## Value

If the function was called by the `attrisk.analysis` function, then `value` is a list containing the following components:

- `Results` - a list containing estimates, confidence bounds, and associated values
- `warn.ind` - a logical value indicating whether warning messages were generated
- `warn.df` - a data frame containing warning messages

If the function was called directly, then `value` is the `Results` list, which contains the following components:

- `AttRisk` - the attributable risk estimate
- `ARlog.se` - standard error for the log of the attributable risk estimate
- `ConfLimits` - confidence limits for the attributable risk estimate
- `WeightTotal` - sum of the final adjusted weights
- `CellCounts` - cell and margin counts for the 2x2 table
- `CellProportions` - estimated cell proportions for the 2x2 table

## Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>

## References

Van Sickle, J. and S. G. Paulsen. (2008). Assessing the attributable risks, relative risks, and regional extent of aquatic stressors. *Journal of the North American Benthological Society* **27**, 920-931.

Sarndal, C.E., B. Swensson, and J. Wretman. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

## Examples

```
response <- sample(c("Poor", "Good"), 100, replace=TRUE)
stressor <- sample(c("Poor", "Good"), 100, replace=TRUE)
wgt <- runif(100, 10, 100)
attrisk.est(response, stressor, wgt=wgt, vartype="SRS")

xcoord <- runif(100)
ycoord <- runif(100)
attrisk.est(response, stressor, wgt=wgt, xcoord=xcoord, ycoord=ycoord)
```

cat.analysis

*Categorical Data Analysis for Probability Survey Data***Description**

This function organizes input and output for analysis of categorical data generated by a probability survey. Input can be either an object belonging to class `spsurvey.analysis` (see the documentation for function `spsurvey.analysis`) or through use of the other arguments to this function.

**Usage**

```
cat.analysis(sites=NULL, subpop=NULL, design=NULL, data.cat=NULL, popsize=NULL,
  popcorrect=FALSE, pcFSIZE=NULL, N.cluster=NULL, stage1size=NULL,
  sizeweight=FALSE, vartype="Local", conf=95, spsurvey.obj=NULL)
```

**Arguments**

sites	a data frame consisting of two variables: the first variable is site IDs, and the second variable is a logical vector indicating which sites to use in the analysis. If <code>spsurvey.obj</code> is not provided, then this argument is required. The default is <code>NULL</code> .
subpop	a data frame describing sets of populations and subpopulations for which estimates will be calculated. The first variable is site IDs. Each subsequent variable identifies a Type of population, where the variable name is used to identify Type. A Type variable identifies each site with one of the subpopulations of that Type. If <code>spsurvey.obj</code> is not provided, then this argument is required. The default is <code>NULL</code> .
design	a data frame consisting of design variables. If <code>spsurvey.obj</code> is not provided, then this argument is required. The default is <code>NULL</code> . Variables should be named as follows: siteID = site IDs wgt = final adjusted weights, which are either the weights for a single-stage sample or the stage two weights for a two-stage sample xcoord = x-coordinates for location, which are either the x-coordinates for a single-stage sample or the stage two x-coordinates for a two-stage sample ycoord = y-coordinates for location, which are either the y-coordinates for a single-stage sample or the stage two y-coordinates for a two-stage sample stratum = the stratum codes cluster = the stage one sampling unit (primary sampling unit or cluster) codes wgt1 = final adjusted stage one weights xcoord1 = the stage one x-coordinates for location ycoord1 = the stage one y-coordinates for location support = support values - the value one (1) for a site from a finite resource or the measure of the sampling unit associated with a site from an extensive resource, which is required for calculation of finite and continuous population correction factors

	swgt = size-weights, which is the stage two size-weight for a two- stage sample swgt1 = stage one size-weights
data.cat	a data frame of categorical response variables. The first variable is site IDs. Subsequent variables are response variables. Missing data (NA) is allowed. If spsurvey.obj is not provided, then this argument is required. The default is NULL.
popsiz	known size of the resource, which is used to perform ratio adjustment to estimators expressed using measurement units for the resource and to calculate strata proportions for calculating estimates for a stratified sample. For a finite resource, this argument is either the total number of sampling units or the known sum of size-weights. For an extensive resource, this argument is the measure of the resource, i.e., either known total length for a linear resource or known total area for an areal resource. The argument must be in the form of a list containing an element for each population Type in the subpop data frame, where NULL is a valid choice for a population Type. The list must be named using the column names for the population Types in subpop. If a population Type doesn't contain subpopulations, then each element of the list is either a single value for an unstratified sample or a vector containing a value for each stratum for a stratified sample, where elements of the vector are named using the stratum codes. If a population Type contains subpopulations, then each element of the list is a list containing an element for each subpopulation, where the list is named using the subpopulation names. The element for each subpopulation will be either a single value for an unstratified sample or a named vector of values for a stratified sample. The default is NULL.
	<p>Example popsiz for a stratified sample:</p> <pre>popsiz = list("Pop 1"=c("Stratum 1"=750, "Stratum 2"=500, "Stratum 3"=250), "Pop 2"=list("SubPop 1"=c("Stratum 1"=350, "Stratum 2"=250, "Stratum 3"=150), "SubPop 2"=c("Stratum 1"=250, "Stratum 2"=150, "Stratum 3"=100), "SubPop 3"=c("Stratum 1"=150, "Stratum 2"=150, "Stratum 3"=75)), "Pop 3"=NULL)</pre>
	<p>Example popsiz for an unstratified sample:</p> <pre>popsiz = list("Pop 1"=1500, "Pop 2"=list("SubPop 1"=750, "SubPop 2"=500, "SubPop 3"=375), "Pop 3"=NULL)</pre>
popcorrect	a logical value that indicates whether finite or continuous population correction

	factors should be employed during variance estimation, where TRUE = use the correction factors and FALSE = do not use the correction factors. The default is FALSE.
pcfsize	size of the resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample this argument must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
N.cluster	the number of stage one sampling units in the resource, which is required for calculation of finite and continuous population correction factors for a two-stage sample. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
stage1size	size of the stage one sampling units of a two-stage sample, which is required for calculation of finite and continuous population correction factors for a two-stage sample and must have the names attribute set to identify the stage one sampling unit codes. For a stratified sample, the names attribute must be set to identify both stratum codes and stage one sampling unit codes using a convention where the two codes are separated by the & symbol, e.g., "Stratum 1&Cluster 1". The default is NULL.
sizeweight	a logical value that indicates whether size-weights should be used in the analysis, where TRUE = use the size-weights and FALSE = do not use the size-weights. The default is FALSE.
vartype	the choice of variance estimator, where "Local" = local mean estimator and "SRS" = SRS estimator. The default is "Local".
conf	the confidence level. The default is 95%.
spsurvey.obj	a list of class spsurvey.analysis that was produced by the function spsurvey.analysis. Depending on input to that function, some elements of the list may be NULL. The default is NULL.

### Value

Value is a data frame of population estimates for all combinations of subpopulation Types, subpopulations within Types, response variables, and categories within each response variable. Estimates are calculated for proportion and size of the population. Standard error estimates and confidence interval estimates also are calculated.

### Author(s)

Tony Olsen <Olsen.Tony@epa.gov>  
Tom Kincaid <Kincaid.Tom@epa.gov>

### References

Diaz-Ramos, S., D.L. Stevens, Jr., and A.R. Olsen. (1996). *EMAP Statistical Methods Manual*. EPA/620/R-96/XXX. Corvallis, OR: U.S. Environmental Protection Agency, Office of Research and Development, National Health Effects and Environmental Research Laboratory, Western Ecology Division.

**See Also**[category.est](#)**Examples**

```
# Categorical variable example for two resource classes:
mysiteID <- paste("Site", 1:100, sep="")
mysites <- data.frame(siteID=mysiteID, Active=rep(TRUE, 100))
mysubpop <- data.frame(siteID=mysiteID, All.Sites=rep("All Sites", 100),
  Resource.Class=rep(c("Good", "Poor"), c(55,45)))
mydesign <- data.frame(siteID=mysiteID, wgt=runif(100, 10, 100),
  xcoord=runif(100), ycoord=runif(100), stratum=rep(c("Stratum1",
  "Stratum2"), 50))
mydata.cat <- data.frame(siteID=mysiteID, CatVar=rep(c("north", "south",
  "east", "west"), 25))
mypopsize <- list(All.Sites=c(Stratum1=3500, Stratum2=2000),
  Resource.Class=list(Good=c(Stratum1=2500, Stratum2=1500),
  Poor=c(Stratum1=1000, Stratum2=500)))
cat.analysis(sites=mysites, subpop=mysubpop, design=mydesign,
  data.cat=mydata.cat, popsize=mypopsize)

# Exclude category "south" from the analysis:
mysites <- data.frame(siteID=mysiteID, Active=rep(c(TRUE, FALSE, TRUE,
  TRUE), 25))
cat.analysis(sites=mysites, subpop=mysubpop, design=mydesign,
  data.cat=mydata.cat, popsize=mypopsize)
```

category.est

*Category Proportion and Size Estimates***Description**

This function estimates proportion (expressed as percent) and size of a resource in each of a set of categories and can also be used to estimate proportion and size for site status categories. Upper and lower confidence bounds also are estimated.

**Usage**

```
category.est(catvar, wgt, x=NULL, y=NULL, stratum=NULL, cluster=NULL, wgt1=NULL,
  x1=NULL, y1=NULL, popsize=NULL, popcorrect=FALSE, pcfsz=NULL,
  N.cluster=NULL, stage1size=NULL, support=NULL, sizeweight=FALSE, swgt=NULL,
  swgt1=NULL, vartype="Local", conf=95, check.ind=TRUE, warn.ind=NULL,
  warn.df=NULL, warn.vec=NULL)
```

**Arguments**

catvar            the value of the categorical response variable or the site status for each site.

wtg	the final adjusted weight (inverse of the sample inclusion probability) for each site, which is either the weight for a single-stage sample or the stage two weight for a two-stage sample.
x	x-coordinate for location for each site, which is either the x-coordinate for a single-stage sample or the stage two x-coordinate for a two-stage sample. The default is NULL.
y	y-coordinate for location for each site, which is either the y-coordinate for a single-stage sample or the stage two y-coordinate for a two-stage sample. The default is NULL.
stratum	the stratum for each site. The default is NULL.
cluster	the stage one sampling unit (primary sampling unit or cluster) code for each site. The default is NULL.
wtg1	the final adjusted stage one weight for each site. The default is NULL.
x1	the stage one x-coordinate for location for each site. The default is NULL.
y1	the stage one y-coordinate for location for each site. The default is NULL.
popsize	known size of the resource, which is used to perform ratio adjustment to estimators expressed using measurement units for the resource and to calculate strata proportions for calculating estimates for a stratified sample. For a finite resource, this argument is either the total number of sampling units or the known sum of size-weights. For an extensive resource, this argument is the measure of the resource, i.e., either known total length for a linear resource or known total area for an areal resource. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
popcorrect	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation, where TRUE = use the correction factors and FALSE = do not use the correction factors. The default is FALSE.
pcfsize	size of the resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample this argument must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
N.cluster	the number of stage one sampling units in the resource, which is required for calculation of finite and continuous population correction factors for a two-stage sample. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
stage1size	size of the stage one sampling units of a two-stage sample, which is required for calculation of finite and continuous population correction factors for a two-stage sample and must have the names attribute set to identify the stage one sampling unit codes. For a stratified sample, the names attribute must be set to identify both stratum codes and stage one sampling unit codes using a convention where the two codes are separated by the & symbol, e.g., "Stratum 1&Cluster 1". The default is NULL.



support	the support value for each site - the value one (1) for a site from a finite resource or the measure of the sampling unit associated with a site from an extensive resource, which is required for calculation of finite and continuous population correction factors. The default is NULL.
sizeweight	a logical value that indicates whether size-weights should be used in the analysis, where TRUE = use the size-weights and FALSE = do not use the size-weights. The default is FALSE.
swgt	the size-weight for each site, which is the stage two size-weight for a two-stage sample. The default is NULL.
swgt1	the stage one size-weight for each site. The default is NULL.
vartype	the choice of variance estimator, where "Local" = local mean estimator and "SRS" = SRS estimator. The default is "Local".
conf	the confidence level. The default is 95%.
check.ind	a logical value that indicates whether compatibility checking of the input values is conducted, where TRUE = conduct compatibility checking and FALSE = do not conduct compatibility checking. The default is TRUE.
warn.ind	a logical value that indicates whether warning messages were generated, where TRUE = warning messages were generated and FALSE = warning messages were not generated. The default is NULL.
warn.df	a data frame for storing warning messages. The default is NULL.
warn.vec	a vector that contains names of the population type, the subpopulation, and an indicator. The default is NULL.

## Details

Proportion estimates are calculated using the Horvitz-Thompson ratio estimator, i.e., the ratio of two Horvitz-Thompson estimators. The numerator of the ratio estimates the size of the category. The denominator of the ratio estimates the size of the resource. Variance estimates for the proportion estimates are calculated using either the local mean variance estimator or the simple random sampling (SRS) variance estimator. The choice of variance estimator is subject to user control. The local mean variance estimator requires the x-coordinate and the y-coordinate of each site. The SRS variance estimator uses the independent random sample approximation to calculate joint inclusion probabilities. Confidence bounds are calculated using a Normal distribution multiplier. For a finite resource size is the number of units in the resource. For an extensive resource size is the measure (extent) of the resource, i.e., length, area, or volume. Size estimates are calculated using the Horvitz-Thompson estimator. Variance estimates for the size estimates are calculated using either the local mean variance estimator or the SRS variance estimator. The function can accommodate a stratified sample. For a stratified sample, separate estimates and standard errors are calculated for each stratum, which are used to produce estimates and standard errors for all strata combined. Strata that contain a single value are removed. For a stratified sample, when either the size of the resource or the sum of the size-weights for the resource is provided for each stratum, those values are used as stratum weights for calculating the estimates and standard errors for all strata combined. In addition, when either of those known values is provided for each stratum, size estimates are obtained by multiplying the proportion estimate, i.e., the Horvitz-Thompson ratio estimator, by the known value for the stratum. For a stratified sample when neither the size of the resource nor the sum of the size-weights of the resource is provided for each stratum, estimated values are used

as stratum weights for calculating the estimates and standard errors for all strata combined. The function can accommodate single-stage and two-stage samples for both stratified and unstratified sampling designs. Finite population and continuous population correction factors can be utilized in variance estimation. The function checks for compatibility of input values and removes missing values.

### Value

If the function was called by the `cat.analysis` function, then value is a list containing the following components:

- `Results` - a data frame containing estimates and confidence bounds
- `warn.ind` - a logical value indicating whether warning messages were generated
- `warn.df` - a data frame containing warning messages

If the function was called directly, then value is a data frame containing estimates and confidence bounds.

### Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>

### References

Diaz-Ramos, S., D.L. Stevens, Jr., and A.R. Olsen. (1996). *EMAP Statistical Methods Manual*. EPA/620/R-96/XXX. Corvallis, OR: U.S. Environmental Protection Agency, Office of Research and Development, National Health Effects and Environmental Research Laboratory, Western Ecology Division.

### Examples

```
catvar <- rep(c("north", "south", "east", "west"), rep(25, 4))
wgt <- runif(100, 10, 100)
category.est(catvar, wgt, vartype="SRS")

x <- runif(100)
y <- runif(100)
category.est(catvar, wgt, x, y)
```

---

cdf.decon

*Cumulative Distribution Function - Deconvolution*

---

### Description

This function calculates an estimate of the deconvoluted cumulative distribution function (CDF) for the proportion (expressed as percent) and the total of a response variable, where the response variable may be defined for either a finite or an extensive resource. Optionally, for a finite resource, the size-weighted CDF can be calculated. In addition the standard error of the estimated CDF and confidence bounds are calculated. The simulation extrapolation deconvolution method (Stefanski and Bay, 1996) is used to deconvolute measurement error variance from the response.

**Usage**

```
cdf.decon(z, wgt, sigma, var.sigma=NULL, x=NULL, y=NULL, stratum=NULL,
  cluster=NULL, wgt1=NULL, x1=NULL, y1=NULL, popsize=NULL, popcorrect=FALSE,
  pcFSIZE=NULL, N.cluster=NULL, stage1size=NULL, support=NULL,
  sizeweight=FALSE, swgt=NULL, swgt1=NULL, vartype="Local", conf=95,
  cdfval=NULL, pctval=c(5,10,25,50,75,90,95), check.ind=TRUE, warn.ind=NULL,
  warn.df=NULL, warn.vec=NULL)
```

**Arguments**

z	the response value for each site.
wgt	the final adjusted weight (inverse of the sample inclusion probability) for each site, which is either the weight for a single-stage sample or the stage two weight for a two-stage sample.
sigma	measurement error variance.
var.sigma	variance of the measurement error variance. The default is NULL.
x	x-coordinate for location for each site, which is either the x-coordinate for a single-stage sample or the stage two x-coordinate for a two-stage sample. The default is NULL.
y	y-coordinate for location for each site, which is either the y-coordinate for a single-stage sample or the stage two y-coordinate for a two-stage sample. The default is NULL.
stratum	the stratum for each site. The default is NULL.
cluster	the stage one sampling unit (primary sampling unit or cluster) code for each site. The default is NULL.
wgt1	the final adjusted stage one weight for each site. The default is NULL.
x1	the stage one x-coordinate for location for each site. The default is NULL.
y1	the stage one y-coordinate for location for each site. The default is NULL.
popsize	known size of the resource, which is used to perform ratio adjustment to estimators expressed using measurement units for the resource and to calculate strata proportions for calculating estimates for a stratified sample. For a finite resource, this argument is either the total number of sampling units or the known sum of size-weights. For an extensive resource, this argument is the measure of the resource, i.e., either known total length for a linear resource or known total area for an areal resource. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
popcorrect	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation, where TRUE = use the correction factors and FALSE = do not use the correction factors. The default is FALSE.
pcFSIZE	size of the resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample this argument must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.

N.cluster	the number of stage one sampling units in the resource, which is required for calculation of finite and continuous population correction factors for a two-stage sample. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
stage1size	size of the stage one sampling units of a two-stage sample, which is required for calculation of finite and continuous population correction factors for a two-stage sample and must have the names attribute set to identify the stage one sampling unit codes. For a stratified sample, the names attribute must be set to identify both stratum codes and stage one sampling unit codes using a convention where the two codes are separated by the & symbol, e.g., "Stratum 1&Cluster 1". The default is NULL.
support	the support value for each site - the value one (1) for a site from a finite resource or the measure of the sampling unit associated with a site from an extensive resource, which is required for calculation of finite and continuous population correction factors. The default is NULL.
sizeweight	a logical value that indicates whether size-weights should be used in the analysis, where TRUE = use the size-weights and FALSE = do not use the size-weights. The default is FALSE.
swgt	the size-weight for each site, which is the stage two size-weight for a two-stage sample. The default is NULL.
swgt1	the stage one size-weight for each site. The default is NULL.
vartype	the choice of variance estimator, where "Local" = local mean estimator and "SRS" = SRS estimator. The default is "Local".
conf	the confidence level. The default is 95%.
cdfval	the set of values at which the CDF is estimated. If a set of values is not provided, then the sorted set of unique values of the response variable is used. The default is NULL.
pctval	the set of values at which percentiles are estimated. The default set is: {5, 10, 25, 50, 75, 90, 95}.
check.ind	a logical value that indicates whether compatibility checking of the input values is conducted, where TRUE = conduct compatibility checking and FALSE = do not conduct compatibility checking. The default is TRUE.
warn.ind	a logical value that indicates whether warning messages were generated, where TRUE = warning messages were generated and FALSE = warning messages were not generated. The default is NULL.
warn.df	a data frame for storing warning messages. The default is NULL.
warn.vec	a vector that contains names of the population type, the subpopulation, and an indicator. The default is NULL.

## Details

This function calculates an estimate of the deconvoluted cumulative distribution function (CDF) for the proportion (expressed as percent) and the total of a response variable, where the response variable may be defined for either a finite or an extensive resource. Optionally, for a finite resource,

the size-weighted CDF can be calculated. In addition the standard error of the estimated CDF and confidence bounds are calculated. The simulation extrapolation deconvolution method (Stefanski and Bay, 1996) is used to deconvolute measurement error variance from the response. The user can supply the set of values at which the CDF is estimated. For the CDF of a proportion, the Horvitz-Thompson ratio estimator, i.e., the ratio of two Horvitz-Thompson estimators, is used to calculate the CDF estimate. For the CDF of a total, the user can supply the known size of the resource or the known sum of the size-weights of the resource, as appropriate. For the CDF of a total when either the size of the resource or the sum of the size-weights of the resource is provided, the classic ratio estimator is used to calculate the CDF estimate, where that estimator is the product of the known value and the Horvitz-Thompson ratio estimator. For the CDF of a total when neither the size of the resource nor the sum of the size-weights of the resource is provided, the Horvitz-Thompson estimator is used to calculate the CDF estimate. Variance estimates for the estimated CDF are calculated using either the local mean variance estimator or the simple random sampling (SRS) variance estimator. The choice of variance estimator is subject to user control. The local mean variance estimator requires the x-coordinate and the y-coordinate of each site. The SRS variance estimator uses the independent random sample approximation to calculate joint inclusion probabilities. Confidence bounds are calculated using a Normal distribution multiplier. In addition the function uses the estimated CDF to calculate percentile estimates. Estimated confidence bounds for the percentile estimates are calculated. The user can supply the set of values for which percentiles estimates are desired. Optionally, the user can use the default set of percentiles. The function can accommodate a stratified sample. For a stratified sample, separate estimates and standard errors are calculated for each stratum, which are used to produce estimates and standard errors for all strata combined. Strata that contain a single value are removed. For a stratified sample, when either the size of the resource or the sum of the size- weights of the resource is provided for each stratum, those values are used as stratum weights for calculating the estimates and standard errors for all strata combined. For a stratified sample when neither the size of the resource nor the sum of the size-weights of the resource is provided for each stratum, estimated values are used as stratum weights for calculating the estimates and standard errors for all strata combined. The function can accommodate single-stage and two-stage samples for both stratified and unstratified sampling designs. Finite population and continuous population correction factors can be utilized in variance estimation. The function checks for compatibility of input values and removes missing values.

## Value

If the function was called by the `cont.analysis` function, then `value` is a list containing the following components:

- `Results` - a list composed of two objects:
  - `CDF` - a data frame that contains CDF estimates
  - `Pct` - a data frame that contains percentile estimates
- `warn.ind` - a logical value indicating whether warning messages were generated
- `warn.df` - a data frame containing warning messages

If the function was called directly, then `value` is a list containing the following components:

- `CDF` - a data frame that contains CDF estimates
- `Pct` - a data frame that contains percentile estimates

**Author(s)**

Tom Kincaid <Kincaid.Tom@epa.gov>

**References**

Diaz-Ramos, S., D.L. Stevens, Jr., and A.R. Olsen. (1996). *EMAP Statistical Methods Manual*. EPA/620/R-96/XXX. Corvallis, OR: U.S. Environmental Protection Agency, Office of Research and Development, National Health Effects and Environmental Research Laboratory, Western Ecology Division.

Stefanski, F.A. and J.M. Bay. (1996). Simulation extrapolation deconvolution of finite population cumulative distribution function estimators, *Biometrika* **83**, 407-417.

**Examples**

```
z <- rnorm(100, 10, 1)
wgt <- runif(100, 10, 100)
cdfval <- seq(min(z), max(z), length=20)
cdf.decon(z, wgt, sigma=0.25, var.sigma=0.1, vartype=
  "SRS", cdfval=cdfval)

x <- runif(100)
y <- runif(100)
cdf.decon(z, wgt, sigma=0.25, var.sigma=0.1, x, y, cdfval=
  cdfval)
```

---

cdf.est

*Cumulative Distribution Function - Estimation*

---

**Description**

This function calculates an estimate of the cumulative distribution function (CDF) for the proportion (expressed as percent) and the total of a response variable, where the response variable may be defined for either a finite or an extensive resource. Optionally, for a finite resource, the size-weighted CDF can be calculated. In addition the standard error of the estimated CDF and confidence bounds are calculated.

**Usage**

```
cdf.est(z, wgt, x=NULL, y=NULL, stratum=NULL, cluster=NULL, wgt1=NULL, x1=NULL,
  y1=NULL, popsize=NULL, popcorrect=FALSE, pcfsz=NULL, N.cluster=NULL,
  stage1size=NULL, support=NULL, sizeweight=FALSE, swgt=NULL, swgt1=NULL,
  vartype="Local", conf=95, cdfval=NULL, pctval=c(5,10,25,50,75,90,95),
  check.ind=TRUE, warn.ind=NULL, warn.df=NULL, warn.vec=NULL)
```

**Arguments**

<code>z</code>	the response value for each site.
<code>wgt</code>	the final adjusted weight (inverse of the sample inclusion probability) for each site, which is either the weight for a single-stage sample or the stage two weight for a two-stage sample.
<code>x</code>	x-coordinate for location for each site, which is either the x-coordinate for a single-stage sample or the stage two x-coordinate for a two-stage sample. The default is NULL.
<code>y</code>	y-coordinate for location for each site, which is either the y-coordinate for a single-stage sample or the stage two y-coordinate for a two-stage sample. The default is NULL.
<code>stratum</code>	the stratum for each site. The default is NULL.
<code>cluster</code>	the stage one sampling unit (primary sampling unit or cluster) code for each site. The default is NULL.
<code>wgt1</code>	the final adjusted stage one weight for each site. The default is NULL.
<code>x1</code>	the stage one x-coordinate for location for each site. The default is NULL.
<code>y1</code>	the stage one y-coordinate for location for each site. The default is NULL.
<code>popsiz</code>	known size of the resource, which is used to perform ratio adjustment to estimators expressed using measurement units for the resource and to calculate strata proportions for calculating estimates for a stratified sample. For a finite resource, this argument is either the total number of sampling units or the known sum of size-weights. For an extensive resource, this argument is the measure of the resource, i.e., either known total length for a linear resource or known total area for an areal resource. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
<code>popcorrect</code>	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation, where TRUE = use the correction factors and FALSE = do not use the correction factors. The default is FALSE.
<code>pcfsiz</code>	size of the resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample this argument must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
<code>N.cluster</code>	the number of stage one sampling units in the resource, which is required for calculation of finite and continuous population correction factors for a two-stage sample. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
<code>stage1siz</code>	size of the stage one sampling units of a two-stage sample, which is required for calculation of finite and continuous population correction factors for a two-stage sample and must have the names attribute set to identify the stage one sampling unit codes. For a stratified sample, the names attribute must be set to identify both stratum codes and stage one sampling unit codes using a convention where

	the two codes are separated by the & symbol, e.g., "Stratum 1&Cluster 1". The default is NULL.
support	the support value for each site - the value one (1) for a site from a finite resource or the measure of the sampling unit associated with a site from an extensive resource, which is required for calculation of finite and continuous population correction factors. The default is NULL.
sizeweight	a logical value that indicates whether size-weights should be used in the analysis, where TRUE = use the size-weights and FALSE = do not use the size-weights. The default is FALSE.
swgt	the size-weight for each site, which is the stage two size-weight for a two-stage sample. The default is NULL.
swgt1	the stage one size-weight for each site. The default is NULL.
vartype	the choice of variance estimator, where "Local" = local mean estimator and "SRS" = SRS estimator. The default is "Local".
conf	the confidence level. The default is 95%.
cdfval	the set of values at which the CDF is estimated. If a set of values is not provided, then the sorted set of unique values of the response variable is used. The default is NULL.
pctval	the set of values at which percentiles are estimated. The default set is: {5, 10, 25, 50, 75, 90, 95}.
check.ind	a logical value that indicates whether compatibility checking of the input values is conducted, where TRUE = conduct compatibility checking and FALSE = do not conduct compatibility checking. The default is TRUE.
warn.ind	a logical value that indicates whether warning messages were generated, where TRUE = warning messages were generated and FALSE = warning messages were not generated. The default is NULL.
warn.df	a data frame for storing warning messages. The default is NULL.
warn.vec	a vector that contains names of the population type, the subpopulation, and an indicator. The default is NULL.

## Details

This function calculates an estimate of the cumulative distribution function (CDF) for the proportion (expressed as percent) and the total of a response variable, where the response variable may be defined for either a finite or an extensive resource. Optionally, for a finite resource, the size-weighted CDF can be calculated. In addition the standard error of the estimated CDF and confidence bounds are calculated. The user can supply the set of values at which the CDF is estimated. For the CDF of a proportion, the Horvitz-Thompson ratio estimator, i.e., the ratio of two Horvitz-Thompson estimators, is used to calculate the CDF estimate. For the CDF of a total, the user can supply the known size of the resource or the known sum of the size-weights of the resource, as appropriate. For the CDF of a total when either the size of the resource or the sum of the size-weights of the resource is provided, the classic ratio estimator is used to calculate the CDF estimate, where that estimator is the product of the known value and the Horvitz-Thompson ratio estimator. For the CDF of a total when neither the size of the resource nor the sum of the size-weights of the resource is provided, the Horvitz-Thompson estimator is used to calculate the CDF estimate. Variance estimates for the



estimated CDF are calculated using either the local mean variance estimator or the simple random sampling (SRS) variance estimator. The choice of variance estimator is subject to user control. The local mean variance estimator requires the x-coordinate and the y-coordinate of each site. The SRS variance estimator uses the independent random sample approximation to calculate joint inclusion probabilities. Confidence bounds are calculated using a Normal distribution multiplier. In addition the function uses the estimated CDF to calculate percentile estimates. Estimated confidence bounds for the percentile estimates are calculated. The user can supply the set of values for which percentiles estimates are desired. Optionally, the user can use the default set of percentiles. The function can accommodate a stratified sample. For a stratified sample, separate estimates and standard errors are calculated for each stratum, which are used to produce estimates and standard errors for all strata combined. Strata that contain a single value are removed. For a stratified sample, when either the size of the resource or the sum of the size- weights of the resource is provided for each stratum, those values are used as stratum weights for calculating the estimates and standard errors for all strata combined. For a stratified sample when neither the size of the resource nor the sum of the size-weights of the resource is provided for each stratum, estimated values are used as stratum weights for calculating the estimates and standard errors for all strata combined. The function can accommodate single-stage and two-stage samples for both stratified and unstratified sampling designs. Finite population and continuous population correction factors can be utilized in variance estimation. The function checks for compatibility of input values and removes missing values.

### Value

If the function was called by the `cont.analysis` function, then `value` is a list containing the following components:

- `Results` - a list composed of two objects:
  - `CDF` - a data frame that contains CDF estimates
  - `Pct` - a data frame that contains percentile estimates
- `warn.ind` - a logical value indicating whether warning messages were generated
- `warn.df` - a data frame containing warning messages

If the function was called directly, then `value` is a list containing the following components:

- `CDF` - a data frame that contains CDF estimates
- `Pct` - a data frame that contains percentile estimates

### Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>

### References

Diaz-Ramos, S., D.L. Stevens, Jr., and A.R. Olsen. (1996). *EMAP Statistical Methods Manual*. EPA/620/R-96/XXX. Corvallis, OR: U.S. Environmental Protection Agency, Office of Research and Development, National Health Effects and Environmental Research Laboratory, Western Ecology Division.

**Examples**

```
z <- rnorm(100, 10, 1)
wgt <- runif(100, 10, 100)
cdfval <- seq(min(z), max(z), length=20)
cdf.est(z, wgt, vartype="SRS", cdfval=cdfval)
```

```
x <- runif(100)
y <- runif(100)
cdf.est(z, wgt, x, y, cdfval=cdfval)
```

cdf.plot

*Plot the CDF and Associated Confidence Limits***Description**

This function creates a CDF plot. Input data for the plots is provided by a data frame utilizing the same structure as the data frame named "CDF" that is included in the output object produced by function `cont.analysis`, but the data frame includes only the values for a single CDF. Confidence limits for the CDF also are plotted.

**Usage**

```
cdf.plot(cdfest, units.cdf="Percent", type.cdf="Continuous", logx="", xlbl=NULL,
  ylbl="Percent", ylbl.r=NULL, figlab=NULL, legloc="BR", confcut=5, conflev=95,
  cex.main=1.2, ...)
```

**Arguments**

<code>cdfest</code>	data frame utilizing the same structure as the data frame named "CDF" that is included in the output object produced by function <code>cont.analysis</code> . The data frame must contain only a single cdf estimate.
<code>units.cdf</code>	indicator for the type of units in which the CDF is plotted, where "Percent" means the plot is in terms of percent of the population, and "Units" means the plot is in terms of units of the population. The default is "Percent".
<code>type.cdf</code>	character string consisting of the value "Continuous" or "Ordinal" that controls the type of CDF plot for each indicator. The default is "Continuous".
<code>logx</code>	character string consisting of the value "" or "x" that controls whether the x axis uses the original scale ("" ) or the base 10 logarithmic scale ("x"). The default is "".
<code>xlbl</code>	character string providing the x-axis label. If this argument equals NULL, then the indicator name is used as the label. The default is NULL.
<code>ylbl</code>	character string providing the the y-axis label. The default is "Percent".
<code>ylbl.r</code>	character string providing the label for the right side y-axis, where NULL means a label is not created, and "Same" means the label is the same as the left side label (i.e., argument <code>ylbl</code> ). The default is NULL.

figlab	character string providing the plot title. The default is NULL.
legloc	indicator for location of the plot legend, where "BR" means bottom right, "BL" means bottom left, "TR" means top right, and "TL" means top left. The default is "BR".
confcut	numeric value that controls plotting confidence limits at the CDF extremes. Confidence limits for CDF values (percent scale) less than confcut or greater than 100 minus confcut are not plotted. A value of zero means confidence limits are plotted for the complete range of the CDF. The default is 5.
conflev	numeric value of the confidence level used for confidence limits. The default is 95.
cex.main	expansion factor for the plot title. The default is 1.2.
...	additional arguments passed to the plot function.

**Value**

A plot of the CDF and its associated confidence limits.

**Author(s)**

Tony Olsen <Olsen.Tony@epa.gov>  
Tom Kincaid <Kincaid.Tom@epa.gov>

**References**

Diaz-Ramos, S., D.L. Stevens, Jr., and A.R. Olsen. (1996). *EMAP Statistical Methods Manual*. EPA/620/R-96/XXX. Corvallis, OR: U.S. Environmental Protection Agency, Office of Research and Development, National Health Effects and Environmental Research Laboratory, Western Ecology Division.

**See Also**

[interp.cdf](#), [interp.axis](#)

**Examples**

```
mysiteID <- paste("Site", 1:100, sep="")
mysites <- data.frame(siteID=mysiteID, Active=rep(TRUE, 100))
mysubpop <- data.frame(siteID=mysiteID, All.Sites=rep("All Sites",100),
  Resource.Class=rep(c("Good","Poor"), c(55,45)))
mydesign <- data.frame(siteID=mysiteID, wgt=runif(100, 10, 100),
  xcoord=runif(100), ycoord=runif(100), stratum=rep(c("Stratum1",
  "Stratum2"), 50))
ContVar <- rnorm(100, 10, 1)
mydata.cont <- data.frame(siteID=mysiteID, ContVar=ContVar)
mypopsize <- list(All.Sites=c(Stratum1=3500, Stratum2=2000),
  Resource.Class=list(Good=c(Stratum1=2500, Stratum2=1500),
  Poor=c(Stratum1=1000, Stratum2=500)))
myanalysis <- cont.analysis(sites=mysites, subpop=mysubpop, design=mydesign,
  data.cont=mydata.cont, popsize=mypopsize)
```

```

par(mfrow=c(2,1))
keep <- myanalysis$CDF$Type == "Resource.Class" &
  myanalysis$CDF$Subpopulation == "Good"
cdf.plot(myanalysis$CDF[keep,], xlbl="ContVar", ylbl="Percent of Stream Length",
  ylbl.r="Stream Length (km)", figlab="Estimates for Resource Class: Good")
keep <- myanalysis$CDF$Type == "Resource.Class" &
  myanalysis$CDF$Subpopulation == "Poor"
cdf.plot(myanalysis$CDF[keep,], xlbl="ContVar", ylbl="Percent of Stream Length",
  ylbl.r="Same", figlab="Estimates for Resource Class: Poor")

```

cdf.test

*Cumulative Distribution Function - Inference***Description**

This function tests for differences between cumulative distribution functions (CDFs) generated by probability surveys. The function returns a variety of test statistics along with their degrees of freedom and p values.

**Usage**

```

cdf.test(bounds, z_1, wgt_1, x_1=NULL, y_1=NULL, z_2, wgt_2, x_2=NULL, y_2=NULL,
  stratum_1=NULL, stratum_2=NULL, cluster_1=NULL, cluster_2=NULL, wgt1_1=NULL,
  x1_1=NULL, y1_1=NULL, wgt1_2=NULL, x1_2=NULL, y1_2=NULL, popsize_1=NULL,
  popsize_2=NULL, popcorrect_1=FALSE, pcfsz_1=NULL, N.cluster_1=NULL,
  stage1size_1=NULL, support_1=NULL, popcorrect_2=FALSE, pcfsz_2=NULL,
  N.cluster_2=NULL, stage1size_2=NULL, support_2=NULL, sizeweight_1=FALSE,
  swgt_1=NULL, swgt1_1=NULL, sizeweight_2=FALSE, swgt_2=NULL, swgt1_2=NULL,
  vartype_1="Local", vartype_2="Local", check.ind=TRUE, warn.ind=NULL,
  warn.df=NULL, warn.vec=NULL)

```

**Arguments**

bounds	vector of upper bounds that define classes for the CDFs, which must contain at least three values.
z_1	response value for each sample one site.
wgt_1	final adjusted weight (inverse of the sample inclusion probability) for each sample one site, which is either the weight for a single-stage sample or the stage two weight for a two-stage sample.
x_1	x-coordinate for location for each sample one site, which is either the x-coordinate for a single-stage sample or the stage two x-coordinate for a two-stage sample. The default is NULL.
y_1	y-coordinate for location for each sample one site, which is either the y-coordinate for a single-stage sample or the stage two y-coordinate for a two-stage sample. The default is NULL.
z_2	response value for each sample two site.

wtg_2	final adjusted weight (inverse of the sample inclusion probability) for each sample two site, which is either the weight for a single-stage sample or the stage two weight for a two-stage sample.
x_2	x-coordinate for location for each sample two site, which is either the x-coordinate for a single-stage sample or the stage two x-coordinate for a two-stage sample. The default is NULL.
y_2	y-coordinate for location for each sample two site, which is either the y-coordinate for a single-stage sample or the stage two y-coordinate for a two-stage sample. The default is NULL.
stratum_1	the stratum for each sample one site. The default is NULL.
stratum_2	the stratum for each sample two site. The default is NULL.
cluster_1	the stage one sampling unit (primary sampling unit or cluster) code for each sample one site. The default is NULL.
cluster_2	the stage one sampling unit (primary sampling unit or cluster) code for each sample two site. The default is NULL.
wtg1_1	the final adjusted stage one weight for each sample one site. The default is NULL.
x1_1	the stage one x-coordinate for location for each sample one site. The default is NULL.
y1_1	the stage one y-coordinate for location for each sample one site. The default is NULL.
wtg1_2	the final adjusted stage one weight for each sample two site. The default is NULL.
x1_2	the stage one x-coordinate for location for each sample two site. The default is NULL.
y1_2	the stage one y-coordinate for location for each sample two site. The default is NULL.
popsize_1	the known size of the sample one resource - the total number of sampling units of a finite resource or the measure of an extensive resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample, this variable also is used to calculate strata weights. For a stratified sample, this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
popsize_2	the known size of the sample two resource - the total number of sampling units of a finite resource or the measure of an extensive resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample, this variable also is used to calculate strata weights. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
popcorrect_1	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation for sample one, where TRUE = use the correction factors and FALSE = do not use the correction factors. The default is FALSE.

pcfsize_1	size of the sample one resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample this argument must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
N.cluster_1	the number of stage one sampling units in the sample one resource, which is required for calculation of finite and continuous population correction factors for a two-stage sample. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
stage1size_1	size of the stage one sampling units of a two-stage sample for sample one, which is required for calculation of finite and continuous population correction factors for a two-stage sample and must have the names attribute set to identify the stage one sampling unit codes. For a stratified sample, the names attribute must be set to identify both stratum codes and stage one sampling unit codes using a convention where the two codes are separated by the & symbol, e.g., "Stratum 1&Cluster 1". The default is NULL.
support_1	the support value for each sample one site - the value one (1) for a site from a finite resource or the measure of the sampling unit associated with a site from an extensive resource, which is required for calculation of finite and continuous population correction factors. The default is NULL.
popcorrect_2	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation for sample two, where TRUE = use the correction factors and FALSE = do not use the correction factors. The default is FALSE.
pcfsize_2	size of the sample two resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample this argument must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
N.cluster_2	the number of stage two sampling units in the sample one resource, which is required for calculation of finite and continuous population correction factors for a two-stage sample. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
stage1size_2	size of the stage one sampling units of a two-stage sample for sample two, which is required for calculation of finite and continuous population correction factors for a two-stage sample and must have the names attribute set to identify the stage one sampling unit codes. For a stratified sample, the names attribute must be set to identify both stratum codes and stage one sampling unit codes using a convention where the two codes are separated by the & symbol, e.g., "Stratum 1&Cluster 1". The default is NULL.
support_2	the support value for each sample two site - the value one (1) for a site from a finite resource or the measure of the sampling unit associated with a site from an extensive resource, which is required for calculation of finite and continuous population correction factors. The default is NULL.

sizeweight_1	a logical value that indicates whether size-weights should be used in the analysis for sample one, where TRUE = use the size-weights and FALSE = do not use the size-weights. The default is FALSE.
swgt_1	the size-weight for each sample one site, which is the stage two size-weight for a two-stage sample. The default is NULL.
swgt1_1	the stage one size-weight for each sample one site. The default is NULL.
sizeweight_2	a logical value that indicates whether size-weights should be used in the analysis for sample two, where TRUE = use the size-weights and FALSE = do not use the size-weights. The default is FALSE.
swgt_2	the size-weight for each sample two site, which is the stage two size-weight for a two-stage sample. The default is NULL.
swgt1_2	the stage one size-weight for each sample two site. The default is NULL.
vartype_1	the choice of variance estimator for sample one, where "Local" = local mean estimator and "SRS" = SRS estimator. The default is "Local".
vartype_2	the choice of variance estimator for sample two, where "Local" = local mean estimator and "SRS" = SRS estimator. The default is "Local".
check.ind	a logical value that indicates whether compatibility checking of the input values is conducted, where TRUE = conduct compatibility checking and FALSE = do not conduct compatibility checking. The default is TRUE.
warn.ind	a logical value that indicates whether warning messages were generated, where TRUE = warning messages were generated and FALSE = warning messages were not generated. The default is NULL.
warn.df	a data frame for storing warning messages. The default is NULL.
warn.vec	a vector that contains names of the population type, the subpopulation, and an indicator. The default is NULL.

## Details

The inferential procedures divide the CDFs into a discrete set of intervals (classes) and then utilize procedures that have been developed for analysis of categorical data from probability surveys. The function calculates the Wald, Rao-Scott first order corrected (mean eigenvalue corrected), and Rao-Scott second order corrected (Satterthwaite corrected) test statistics. Both standard versions of the three statistics, which are distributed as Chi-squared random variables, and alternate version of the statistics, which are distributed as F random variables, are available. The user supplies the set of upper bounds that define the intervals (classes) into which the CDFs are divided (binned). The minimum number of classes is three. The Horvitz-Thompson ratio estimator, i.e., the ratio of two Horvitz-Thompson estimators, is used to calculate estimates of the class proportions for the CDFs. Variance estimates for the test statistics are calculated using either the local mean variance estimator or the simple random sampling (SRS) variance estimator. The choice of variance estimator is subject to user control. The SRS variance estimator uses the independent random sample approximation to calculate joint inclusion probabilities. The function can accommodate a stratified sample. For a stratified sample, separate class proportion estimates and associated covariance estimates are calculated for each stratum, which are used to produce estimates for all strata combined. Strata that contain a single value are removed. For a stratified sample, when either the size of the resource or the sum of the size-weights of the resource is provided for each stratum, those values are used as

stratum weights for calculating the estimates for all strata combined. For a stratified sample when neither the size of the resource nor the sum of the size-weights of the resource is provided for each stratum, estimated values are used as stratum weights for calculating the estimates for all strata combined. The function can accommodate single-stage and two-stage samples for both stratified and unstratified sampling designs. Finite population and continuous population correction factors can be utilized in variance estimation. The function checks for compatibility of input values and removes missing values.

### Value

Value is a data frame containing the test statistic, degrees of freedom (two values labeled Degrees of Freedom\1 and Degrees of Freedom\2), and p value for the Wald, mean eigenvalue, and Satterthwaite test procedures, which includes both Chi-squared distribution and F distribution versions of the procedures. For the Chi-squared versions of the test procedures, Degrees of Freedom\1 contains the relevant value and Degrees of Freedom\2 is set to missing (NA). For the F-based versions of the test procedures Degrees of Freedom\1 contains the numerator degrees of freedom and Degrees of Freedom\2 contains the denominator degrees of freedom.

### Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>

### References

Kincaid, T.M. (2000). Testing for differences between cumulative distribution functions from complex environmental sampling surveys. In *2000 Proceeding of the Section on Statistics and the Environment*, American Statistical Association, Alexandria, VA.

### Examples

```
n <- 100
resp <- rnorm(n, 10, 1)
wgt <- runif(n, 10, 100)
sample1 <- list(z=resp, wgt=wgt)
sample2 <- list(z=resp+0.5, wgt=wgt)
bounds <- sort(c(sample1$z, sample2$z))[floor(seq((2*n)/3, (2*n),
length=3))]
cdf.test(bounds=bounds, z_1=sample1$z, wgt_1=sample1$wgt, z_2=sample2$z,
wgt_2=sample2$wgt, vartype_1="SRS", vartype_2="SRS")

xcoord <- runif(n)
ycoord <- runif(n)
sample1 <- list(z=resp, wgt=wgt, x=xcoord, y=ycoord)
sample2 <- list(z=1.05*resp, wgt=wgt, x=xcoord, y=ycoord)
cdf.test(bounds=bounds, z_1=sample1$z, wgt_1=sample1$wgt, x_1=sample1$x,
y_1=sample1$y, z_2=sample2$z, wgt_2=sample2$wgt, x_2=sample2$x,
y_2=sample2$y)
```



---

change.analysis      *Change Analysis for Probability Survey Data*

---

### Description

This function organizes input and output for analysis of change between two probability surveys.

### Usage

```
change.analysis(sites, repeats=NULL, subpop=NULL, design, data.cat=NULL,
  data.cont=NULL, revisitwgt=FALSE, popsize_1=NULL, popsize_2=NULL,
  popcorrect_1=FALSE, popcorrect_2=FALSE, pcfsz_1=NULL, pcfsz_2=NULL,
  N.cluster_1=NULL, N.cluster_2=NULL, stage1size_1=NULL, stage1size_2=NULL,
  sizeweight_1=FALSE, sizeweight_2=FALSE, vartype_1="Local", vartype_2="Local",
  conf=95)
```

### Arguments

sites	a data frame consisting of three variables: the first variable is site IDs, and the other variables are logical vectors indicating which sites to use in the analysis. The first logical vector indicates the complete set of sites for the first survey. The second logical vector indicates the complete set of sites for the second survey.
repeats	a data frame that identifies site IDs for repeat visit sites from the two surveys. The first variable is site IDs for survey one. The second variable is site IDs for survey two. For each row of the data frame, the two site IDs must correspond to the same site. This argument should equal NULL when repeat visit sites are not present. The default is NULL.
subpop	a data frame describing sets of populations and subpopulations for which estimates will be calculated. The first variable is site IDs. Each subsequent variable identifies a Type of population, where the variable name is used to identify Type. A Type variable identifies each site with one of the subpopulations of that Type. The default is NULL.
design	a data frame consisting of design variables. Variables should be named as follows: siteID = site IDs wgt = final adjusted weights, which are either the weights for a single-stage sample or the stage two weights for a two-stage sample xcoord = x-coordinates for location, which are either the x-coordinates for a single-stage sample or the stage two x-coordinates for a two-stage sample ycoord = y-coordinates for location, which are either the y-coordinates for a single-stage sample or the stage two y-coordinates for a two-stage sample stratum = the stratum codes cluster = the stage one sampling unit (primary sampling unit or cluster) codes wgt1 = final adjusted stage one weights xcoord1 = the stage one x-coordinates for location ycoord1 = the stage one y-coordinates for location

	<p>support = support values - the value one (1) for a site from a finite resource or the measure of the sampling unit associated with a site from an extensive resource, which is required for calculation of finite and continuous population correction factors</p> <p>swgt = size-weights, which is the stage two size-weight for a two- stage sample</p> <p>swgt1 = stage one size-weights</p>
data.cat	a data frame of categorical response variables. The first variable is site IDs. Subsequent variables are response variables. Missing data (NA) is allowed. The default is NULL.
data.cont	a data frame of continuous response variables. The first variable is site IDs. Subsequent variables are response variables. Missing data (NA) is allowed. The default is NULL.
revisitwgt	a logical value that indicates whether the repeat visit sites in the two surveys have the same survey design weights, where TRUE = the weights are the same and FALSE = the weights are not the same. When this argument is FALSE, the repeat visit sites are assigned equal weights when calculating the covariance component of the change estimate variance. The default is FALSE.
popsiz_1	known size of the resource for survey one, which is used to perform ratio adjustment to estimators expressed using measurement units for the resource and to calculate strata proportions for calculating estimates for a stratified sample. For a finite resource, this argument is either the total number of sampling units or the known sum of size-weights. For an extensive resource, this argument is the measure of the resource, i.e., either known total length for a linear resource or known total area for an areal resource. The argument must be in the form of a list containing an element for each population Type in the subpop data frame, where NULL is a valid choice for a population Type. The list must be named using the column names for the population Types in subpop. If a population Type doesn't contain subpopulations, then each element of the list is either a single value for an unstratified sample or a vector containing a value for each stratum for a stratified sample, where elements of the vector are named using the stratum codes. If a population Type contains subpopulations, then each element of the list is a list containing an element for each subpopulation, where the list is named using the subpopulation names. The element for each subpopulation will be either a single value for an unstratified sample or a named vector of values for a stratified sample. The default is NULL.

Example popsize for a stratified sample:

```
popsiz = list("Pop 1"=c("Stratum 1"=750,
"Stratum 2"=500,
"Stratum 3"=250),
"Pop 2"=list("SubPop 1"=c("Stratum 1"=350,
"Stratum 2"=250,
"Stratum 3"=150),
"SubPop 2"=c("Stratum 1"=250,
"Stratum 2"=150,
"Stratum 3"=100),
"SubPop 3"=c("Stratum 1"=150,
"Stratum 2"=150,
```

```
"Stratum 3"=75)),
"Pop 3"=NULL)
```

Example popsize for an unstratified sample:

```
popsiz = list("Pop 1"=1500,
"Pop 2"=list("SubPop 1"=750,
"SubPop 2"=500,
"SubPop 3"=375),
"Pop 3"=NULL)
```

popsiz_2	known size of the resource for survey two. The default is NULL.
popcorrect_1	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation for survey one, where TRUE = use the correction factors and FALSE = do not use the correction factors. The default is FALSE.
popcorrect_2	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation for survey two. The default is FALSE.
pcfsiz_1	size of the resource for survey one, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample this argument must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
pcfsiz_2	size of the resource for survey two. The default is NULL.
N.cluster_1	the number of stage one sampling units in the resource for survey one, which is required for calculation of finite and continuous population correction factors for a two-stage sample. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
N.cluster_2	the number of stage one sampling units in the resource for survey two. The default is NULL.
stage1siz_1	size of the stage one sampling units of a two-stage sample for survey one, which is required for calculation of finite and continuous population correction factors for a two-stage sample and must have the names attribute set to identify the stage one sampling unit codes. For a stratified sample, the names attribute must be set to identify both stratum codes and stage one sampling unit codes using a convention where the two codes are separated by the & symbol, e.g., "Stratum 1&Cluster 1". The default is NULL.
stage1siz_2	size of the stage one sampling units of a two-stage sample for survey two. The default is NULL.
sizeweight_1	a logical value that indicates whether size-weights should be used in the analysis of survey one, where TRUE = use the size-weights and FALSE = do not use the size-weights. The default is FALSE.
sizeweight_2	a logical value that indicates whether size-weights should be used in the analysis of survey two. The default is FALSE.

vartype_1	the choice of variance estimator for survey one, where "Local" = local mean estimator and "SRS" = SRS estimator. The default is "Local".
vartype_2	the choice of variance estimator for survey two. The default is "Local".
conf	the confidence level. The default is 95%.

### Value

Value is a data frame of change estimates for all combinations of population Types, subpopulations within Types, response variables, and categories within each response variable (for categorical variables only). Estimates provided plus standard error and confidence interval estimates.

### Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>

### References

Diaz-Ramos, S., D.L. Stevens, Jr., and A.R. Olsen. (1996). *EMAP Statistical Methods Manual*. EPA/620/R-96/XXX. Corvallis, OR: U.S. Environmental Protection Agency, Office of Research and Development, National Health Effects and Environmental Research Laboratory, Western Ecology Division.

### See Also

[change.est](#)

### Examples

```
# Categorical variable example for three resource classes:
mysiteID <- paste("Site", 1:200, sep="")
mysites <- data.frame(siteID=mysiteID,
                     Survey1=rep(c(TRUE, FALSE), c(100,100)),
                     Survey2=rep(c(FALSE, TRUE), c(100,100)))
myrepeats <- data.frame(siteID_1=paste("Site", 1:40, sep=""),
                       siteID_2=paste("Site", 101:140, sep=""))
mysubpop <- data.frame(siteID=mysiteID,
                      All_Sites=rep("All Sites", 200),
                      Region=rep(c("North", "South"), 100))
mydesign <- data.frame(siteID=mysiteID,
                     wgt=runif(200, 10, 100),
                     xcoord=runif(200),
                     ycoord=runif(200),
                     stratum=rep(rep(c("Stratum1", "Stratum2"), c(2,2)), 50))
mydata.cat <- data.frame(siteID=mysiteID,
                        Resource_Class=sample(c("Good", "Fair", "Poor"),
                                              200, replace=TRUE))
change.analysis(sites=mysites, repeats=myrepeats, subpop=mysubpop,
               design=mydesign, data.cat=mydata.cat, data.cont=NULL)
```

change.est

*Category Proportion and Size Estimates***Description**

This function estimates change between two probability surveys. Upper and lower confidence bounds also are estimated.

**Usage**

```
change.est(resp.ind, z_1, wgt_1, x_1=NULL, y_1=NULL, repeat_1, z_2,
           wgt_2, x_2=NULL, y_2=NULL, repeat_2, revisitwgt=FALSE, stratum_1=NULL,
           stratum_2=NULL, cluster_1=NULL, cluster_2=NULL, wgt1_1=NULL, x1_1=NULL,
           y1_1=NULL, wgt1_2=NULL, x1_2=NULL, y1_2=NULL, popsize_1=NULL, popsize_2=NULL,
           popcorrect_1=FALSE, pcfsiz_1=NULL, N.cluster_1=NULL, stage1size_1=NULL,
           support_1=NULL, popcorrect_2=FALSE, pcfsiz_2=NULL, N.cluster_2=NULL,
           stage1size_2=NULL, support_2=NULL, sizeweight_1=FALSE, swgt_1=NULL,
           swgt1_1=NULL, sizeweight_2=FALSE, swgt_2=NULL, swgt1_2=NULL,
           vartype_1="Local", vartype_2="Local", conf=95, check.ind=TRUE, warn.ind=NULL,
           warn.df=NULL, warn.vec=NULL)
```

**Arguments**

resp.ind	a character value that indicates the type of response variable, where "cat" indicates a categorical variable and "cont" indicates a continuous variable.
z_1	response value for each survey one site.
wgt_1	the final adjusted weight (inverse of the sample inclusion probability) for each survey one site, which is either the weight for a single-stage sample or the stage two weight for a two-stage sample.
x_1	x-coordinate for location for each survey one site, which is either the x-coordinate for a single-stage sample or the stage two x-coordinate for a two-stage sample. The default is NULL.
y_1	y-coordinate for location for each survey one site, which is either the y-coordinate for a single-stage sample or the stage two y-coordinate for a two-stage sample. The default is NULL.
repeat_1	a logical variable that identifies repeat visit sites for survey one.
z_2	response value for each survey two site.
wgt_2	the final adjusted weight for each survey two site.
x_2	x-coordinate for location for each survey two site. The default is NULL.
y_2	y-coordinate for location for each survey two site. The default is NULL.
repeat_2	a logical variable that identifies repeat visit sites for survey two.

revisitwgt	a logical value that indicates whether the repeat visit sites in the two surveys have the same survey design weights, where TRUE = the weights are the same and FALSE = the weights are not the same. When this argument is FALSE, the repeat visit sites are assigned equal weights when calculating the covariance component of the change estimate variance. The default is FALSE.
stratum_1	the stratum for each survey one site. The default is NULL.
stratum_2	the stratum for each survey two site. The default is NULL.
cluster_1	the stage one sampling unit (primary sampling unit or cluster) code for each survey one site. The default is NULL.
cluster_2	the stage one sampling unit (primary sampling unit or cluster) code for each survey two site. The default is NULL.
wgt1_1	the final adjusted stage one weight for each survey one site. The default is NULL.
x1_1	the stage one x-coordinate for location for each survey one site. The default is NULL.
y1_1	the stage one y-coordinate for location for each survey one site. The default is NULL.
wgt1_2	the final adjusted stage one weight for each survey two site. The default is NULL.
x1_2	the stage one x-coordinate for location for each survey two site. The default is NULL.
y1_2	the stage one y-coordinate for location for each survey two site. The default is NULL.
popsize_1	known size of the survey one resource, which is used to perform ratio adjustment to estimators expressed using measurement units for the resource. For a finite resource, this argument is either the total number of sampling units or the known sum of size-weights. For an extensive resource, this argument is the measure of the resource, i.e., either known total length for a linear resource or known total area for an areal resource. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
popsize_2	known size of the survey two resource. The default is NULL.
popcorrect_1	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation for survey one, where TRUE = use the correction factors and FALSE = do not use the correction factors. The default is FALSE.
pcfsize_1	size of the survey one resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample this argument must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
N.cluster_1	the number of stage one sampling units in the survey one resource, which is required for calculation of finite and continuous population correction factors for a two-stage sample. For a stratified sample this variable must be a vector

	containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
stage1size_1	size of the stage one sampling units of a two-stage sample for survey one, which is required for calculation of finite and continuous population correction factors for a two-stage sample and must have the names attribute set to identify the stage one sampling unit codes. For a stratified sample, the names attribute must be set to identify both stratum codes and stage one sampling unit codes using a convention where the two codes are separated by the & symbol, e.g., "Stratum 1&Cluster 1". The default is NULL.
support_1	the support value for each survey one site - the value one (1) for a site from a finite resource or the measure of the sampling unit associated with a site from an extensive resource, which is required for calculation of finite and continuous population correction factors. The default is NULL.
popcorrect_2	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation for survey two. The default is FALSE.
pcfsize_2	size of the survey two resource. The default is NULL.
N.cluster_2	the number of stage one sampling units in the survey two resource. The default is NULL.
stage1size_2	size of the stage one sampling units of a two-stage sample for survey two. The default is NULL.
support_2	the support value for each survey two site. The default is NULL.
sizeweight_1	a logical value that indicates whether size-weights should be used in the analysis for survey one, where TRUE = use the size-weights and FALSE = do not use the size-weights. The default is FALSE.
swgt_1	the size-weight for each survey one site, which is the stage two size-weight for a two-stage sample. The default is NULL.
swgt1_1	the stage one size-weight for each survey one site. The default is NULL.
sizeweight_2	a logical value that indicates whether size-weights should be used in the analysis for survey two. The default is FALSE.
swgt_2	the size-weight for each survey two site. The default is NULL.
swgt1_2	the stage one size-weight for each survey two site. The default is NULL.
vartype_1	the choice of variance estimator for survey one, where "Local" = local mean estimator and "SRS" = SRS estimator. The default is "Local".
vartype_2	the choice of variance estimator for survey two. The default is "Local".
conf	the confidence level. The default is 95%.
check.ind	a logical value that indicates whether compatibility checking of the input values is conducted, where TRUE = conduct compatibility checking and FALSE = do not conduct compatibility checking. The default is TRUE.
warn.ind	a logical value that indicates whether warning messages were generated, where TRUE = warning messages were generated and FALSE = warning messages were not generated. The default is NULL.
warn.df	a data frame for storing warning messages. The default is NULL.
warn.vec	a vector that contains names of the population type, the subpopulation, and an indicator. The default is NULL.

## Details

Change estimates are calculated using the Horvitz-Thompson ratio estimator, i.e., the ratio of two Horvitz-Thompson estimators. The numerator of the ratio estimates the size of the category (for categorical variables) or the variable total (for a continuous variable). The denominator of the ratio estimates the size of the resource. Variance estimates for the proportion estimates are calculated using either the local mean variance estimator or the simple random sampling (SRS) variance estimator. The choice of variance estimator is subject to user control. The local mean variance estimator requires the x-coordinate and the y-coordinate of each site. The SRS variance estimator uses the independent random sample approximation to calculate joint inclusion probabilities. Confidence bounds are calculated using a Normal distribution multiplier. For a finite resource size is the number of units in the resource. For an extensive resource size is the measure (extent) of the resource, i.e., length, area, or volume. Size estimates are calculated using the Horvitz-Thompson estimator. Variance estimates for the size estimates are calculated using either the local mean variance estimator or the SRS variance estimator. The function can accommodate a stratified sample. For a stratified sample, separate estimates and standard errors are calculated for each stratum, which are used to produce estimates and standard errors for all strata combined. Strata that contain a single value are removed. For a stratified sample, when either the size of the resource or the sum of the size-weights for the resource is provided for each stratum, those values are used as stratum weights for calculating the estimates and standard errors for all strata combined. In addition, when either of those known values is provided for each stratum, size estimates are obtained by multiplying the proportion estimate, i.e., the Horvitz-Thompson ratio estimator, by the known value for the stratum. For a stratified sample when neither the size of the resource nor the sum of the size-weights of the resource is provided for each stratum, estimated values are used as stratum weights for calculating the estimates and standard errors for all strata combined. The function can accommodate single-stage and two-stage samples for both stratified and unstratified sampling designs. Finite population and continuous population correction factors can be utilized in variance estimation. The function checks for compatibility of input values and removes missing values.

## Value

If the function was called by the `change.analysis` function, then `value` is a list containing the following components:

- `Results` - a data frame containing estimates and confidence bounds
- `warn.ind` - a logical value indicating whether warning messages were generated
- `warn.df` - a data frame containing warning messages

If the function was called directly, then `value` is a data frame containing estimates and confidence bounds.

## Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>

## References

Diaz-Ramos, S., D.L. Stevens, Jr., and A.R. Olsen. (1996). *EMAP Statistical Methods Manual*. EPA/620/R-96/XXX. Corvallis, OR: U.S. Environmental Protection Agency, Office of Research and



Development, National Health Effects and Environmental Research Laboratory, Western Ecology Division.

### Examples

```
# Categorical variable example:
z_1 <- sample(c("Good","Fair","Poor"), 100, replace=TRUE)
z_2 <- sample(c("Good","Fair","Poor"), 100, replace=TRUE)
wgt_1 <- runif(100, 10, 100)
wgt_2 <- runif(100, 10, 100)
repeat_1 <- rep(c(TRUE,FALSE), c(20,80))
repeat_2 <- rep(c(TRUE,FALSE), c(20,80))
stratum_1 <- rep(c("Stratum1", "Stratum2"), 50)
stratum_2 <- rep(c("Stratum1", "Stratum2"), 50)
change.est(resp.ind="cat", z_1=z_1, wgt_1=wgt_1, repeat_1=repeat_1,
           z_2=z_2, wgt_2=wgt_2, repeat_2=repeat_2, stratum_1=stratum_1,
           stratum_2=stratum_2, vartype_1="SRS", vartype_2="SRS")

# Continuous variable example:
z_1 <- rnorm(100, 10,10)
z_2 <- rnorm(100, 12, 10)
stratum_1 <- rep(c("Stratum1", "Stratum2"), 50)
stratum_2 <- rep(c("Stratum1", "Stratum2"), 50)
change.est(resp.ind="cont", z_1=z_1, wgt_1=wgt_1, repeat_1=repeat_1,
           z_2=z_2, wgt_2=wgt_2, repeat_2=repeat_2, stratum_1=stratum_1,
           stratum_2=stratum_2, vartype_1="SRS", vartype_2="SRS")
```

---

cont.analysis

*Continuous Data Analysis for Probability Survey Data*

---

### Description

This function organizes input and output for analysis of continuous data generated by a probability survey. Input can be either an object belonging to class `spsurvey.analysis` (see the documentation for function `spsurvey.analysis`) or through use of the other arguments to this function.

### Usage

```
cont.analysis(sites=NULL, subpop=NULL, design=NULL, data.cont=NULL, sigma=NULL,
             var.sigma=NULL, popsize=NULL, popcorrect=FALSE, pcFSIZE=NULL, N.cluster=NULL,
             stage1size=NULL, sizeweight=FALSE, total=FALSE, vartype="Local", conf=95,
             pctval=c(5,10,25,50,75,90,95), spsurvey.obj=NULL)
```

### Arguments

`sites` a data frame consisting of two variables: the first variable is site IDs, and the second variable is a logical vector indicating which sites to use in the analysis. If `spsurvey.obj` is not provided, then this argument is required. The default is `NULL`.

subpop	a data frame describing sets of populations and subpopulations for which estimates will be calculated. The first variable is site IDs. Each subsequent variable identifies a Type of population, where the variable name is used to identify Type. A Type variable identifies each site with one of the subpopulations of that Type. If spsurvey.obj is not provided, then this argument is required. The default is NULL.
design	a data frame consisting of design variables. If spsurvey.obj is not provided, then this argument is required. The default is NULL. Variables should be named as follows: siteID = site IDs wgt = final adjusted weights, which are either the weights for a single-stage sample or the stage two weights for a two-stage sample xcoord = x-coordinates for location, which are either the x-coordinates for a single-stage sample or the stage two x-coordinates for a two-stage sample ycoord = y-coordinates for location, which are either the y-coordinates for a single-stage sample or the stage two y-coordinates for a two-stage sample stratum = the stratum codes cluster = the stage one sampling unit (primary sampling unit or cluster) codes wgt1 = final adjusted stage one weights xcoord1 = the stage one x-coordinates for location ycoord1 = the stage one y-coordinates for location support = support values - the value one (1) for a site from a finite resource or the measure of the sampling unit associated with a site from an extensive resource, which is required for calculation of finite and continuous population correction factors swgt = size-weights, which is the stage two size-weight for a two-stage sample swgt1 = stage one size-weights
data.cont	a data frame of continuous response variables. The first variable is site IDs. Subsequent variables are response variables. Missing data (NA) is allowed. If spsurvey.obj is not provided, then this argument is required. The default is NULL.
sigma	measurement error variance. This variable must be a vector containing a value for each response variable and must have the names attribute set to identify the response variable names. Missing data (NA) is allowed. The default is NULL.
var.sigma	variance of the measurement error variance. This variable must be a vector containing a value for each response variable and must have the names attribute set to identify the response variable names. Missing data (NA) is allowed. The default is NULL.
popsiz	known size of the resource, which is used to perform ratio adjustment to estimators expressed using measurement units for the resource and to calculate strata proportions for calculating estimates for a stratified sample. For a finite resource, this argument is either the total number of sampling units or the known sum of size-weights. For an extensive resource, this argument is the measure of the resource, i.e., either known total length for a linear resource or known total area for an areal resource. The argument must be in the form of a list containing an element for each population Type in the subpop data frame, where NULL is a valid choice for a population Type. The list must be named using the column

names for the population Types in subpop. If a population Type doesn't contain subpopulations, then each element of the list is either a single value for an unstratified sample or a vector containing a value for each stratum for a stratified sample, where elements of the vector are named using the stratum codes. If a population Type contains subpopulations, then each element of the list is a list containing an element for each subpopulation, where the list is named using the subpopulation names. The element for each subpopulation will be either a single value for an unstratified sample or a named vector of values for a stratified sample. The default is NULL.

Example popsize for a stratified sample:

```
popsize = list("Pop 1"=c("Stratum 1"=750,
"Stratum 2"=500,
"Stratum 3"=250),
"Pop 2"=list("SubPop 1"=c("Stratum 1"=350,
"Stratum 2"=250,
"Stratum 3"=150),
"SubPop 2"=c("Stratum 1"=250,
"Stratum 2"=150,
"Stratum 3"=100),
"SubPop 3"=c("Stratum 1"=150,
"Stratum 2"=150,
"Stratum 3"=75)),
"Pop 3"=NULL)
```

Example popsize for an unstratified sample:

```
popsize = list("Pop 1"=1500,
"Pop 2"=list("SubPop 1"=750,
"SubPop 2"=500,
"SubPop 3"=375),
"Pop 3"=NULL)
```

popcorrect	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation, where TRUE = use the correction factors and FALSE = do not use the correction factors. The default is FALSE.
pcfsize	size of the resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample this argument must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
N.cluster	the number of stage one sampling units in the resource, which is required for calculation of finite and continuous population correction factors for a two-stage sample. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
stage1size	size of the stage one sampling units of a two-stage sample, which is required for calculation of finite and continuous population correction factors for a two-stage sample and must have the names attribute set to identify the stage one sampling

	unit codes. For a stratified sample, the names attribute must be set to identify both stratum codes and stage one sampling unit codes using a convention where the two codes are separated by the & symbol, e.g., "Stratum 1&Cluster 1". The default is NULL.
sizeweight	a logical value that indicates whether size-weights should be used in the analysis, where TRUE = use the size-weights and FALSE = do not use the size-weights. The default is FALSE.
total	a logical value that indicates whether the population total estimate should be included in the output Pct data frame, where TRUE = include the total estimate and FALSE = do not include the estimate. The default is FALSE.
vartype	the choice of variance estimator, where "Local" = local mean estimator and "SRS" = SRS estimator. The default is "Local".
conf	the confidence level. The default is 95%.
pctval	the set of values at which percentiles are estimated. The default set is: {5, 10, 25, 50, 75, 90, 95}.
spsurvey.obj	A list of class spsurvey.analysis that was produced by the function spsurvey.analysis. Depending on input to that function, some elements of the list may be NULL. The default is NULL.

### Details

CDF estimates are calculated for all unique values of a response variable.

### Value

Value is a list containing either two or four data frames of population estimates for all combinations of population Types, subpopulations within Types, and response variables. The data frames containing deconvoluted CDF estimates and deconvoluted percentile estimates are only included in the output list when an input value for measurement error variance is provided to the function. CDF estimates are calculated for both proportion and size of the population. Standard error estimates and confidence interval estimates also are calculated. The four data frames are:

- CDF - a data frame containing the CDF estimates
- Pct - a data frame containing the percentile estimates plus population mean, standard deviation, and variance estimates
- CDF.D - a data frame containing the deconvoluted CDF estimates
- Pct.D - a data frame containing the deconvoluted percentile estimates

### Author(s)

Tony Olsen <Olsen.Tony@epa.gov>  
Tom Kincaid <Kincaid.Tom@epa.gov>

### References

Diaz-Ramos, S., D.L. Stevens, Jr., and A.R. Olsen. (1996). *EMAP Statistical Methods Manual*. EPA/620/R-96/XXX. Corvallis, OR: U.S. Environmental Protection Agency, Office of Research and Development, National Health Effects and Environmental Research Laboratory, Western Ecology Division.

**See Also**

[cdf.est](#), [total.est](#)

**Examples**

```
# Continuous variable example:
mysiteID <- paste("Site", 1:100, sep="")
mysites <- data.frame(siteID=mysiteID, Active=rep(TRUE, 100))
mysubpop <- data.frame(siteID=mysiteID, All.Sites=rep("All Sites",100),
  Resource.Class=rep(c("Good","Poor"), c(55,45)))
mydesign <- data.frame(siteID=mysiteID, wgt=runif(100, 10, 100),
  xcoord=runif(100), ycoord=runif(100), stratum=rep(c("Stratum1",
  "Stratum2"), 50))
ContVar <- rnorm(100, 10, 1)
mydata.cont <- data.frame(siteID=mysiteID, ContVar=ContVar)
mypopsize <- list(All.Sites=c(Stratum1=3500, Stratum2=2000),
  Resource.Class=list(Good=c(Stratum1=2500, Stratum2=1500),
  Poor=c(Stratum1=1000, Stratum2=500)))
cont.analysis(sites=mysites, subpop=mysubpop, design=mydesign,
  data.cont=mydata.cont, popsize=mypopsize)

# Include deconvolution estimates:
mydata.cont <- data.frame(siteID=mysiteID, ContVar=ContVar,
  ContVar.1=ContVar + rnorm(100, 0, sqrt(0.25)),
  ContVar.2=ContVar + rnorm(100, 0, sqrt(0.50)))
mysigma <- c(NA, 0.25, 0.50)
names(mysigma) <- c("ContVar", "ContVar.1", "ContVar.2")
cont.analysis(sites=mysites, subpop=mysubpop[,1:2], design=mydesign,
  data.cont=mydata.cont, sigma=mysigma, popsize=mypopsize[1])
```

---

 cont.cdfplot

---

*Create CDF Plots from the cont.analysis Output Object*


---

**Description**

This function creates CDF plots. Input data for the plots is provided by a data frame utilizing the same structure as the data frame named "CDF" that is included in the output object produced by function `cont.analysis`. Plots are produced for every combination of Type of population, subpopulation within Type, and indicator. Output from the function is placed in a PDF file.

**Usage**

```
cont.cdfplot(pdffile="cdf2x2.pdf", cdfest, units.cdf="Percent",
  ind.type=rep("Continuous", nind), logx=rep("", nind), xlabel=NULL,
  ylabel="Percent", ylabel.r=NULL, legloc="BR", cdf.page=4, width=10, height=8,
  confcut=5, cex.main=1.2, ...)
```

**Arguments**

pdffile	name of the PDF file. The default is "cdf2x2.pdf".
cdfest	data frame utilizing the same structure as the data frame named "CDF" that is included in the output object produced by function cont.analysis.
units.cdf	indicator for the type of units in which the CDF is plotted, where "Percent" means the plot is in terms of percent of the population, and "Units" means the plot is in terms of units of the population. The default is "Percent".
ind.type	character vector consisting of the values "Continuous" or "Ordinal" that controls the type of CDF plot for each indicator. The default is "Continuous" for every indicator.
logx	character vector consisting of the values "" or "x" that controls whether the x axis uses the original scale ("") or the base 10 logarithmic scale ("x") for each indicator. The default is "" for every indicator.
xlbl	character vector consisting of the x-axis label for each indicator. If this argument equals NULL, then indicator names are used as the labels. The default is NULL.
ylbl	character string providing the the y-axis label. The default is "Percent".
ylbl.r	character string providing the label for the right side y-axis, where NULL means a label is not created, and "Same" means the label is the same as the left side label (i.e., argument ylbl). The default is NULL.
legloc	indicator for location of the plot legend, where "BR" means bottom right, "BL" means bottom left, "TR" means top right, and "TL" means top left. The default is "BR".
cdf.page	number of CDF plots on each page, which must be chosen from the values: 1, 2, 4, or 6. The default is 4.
width	width of the graphic region in inches. The default is 10.
height	height of the graphic region in inches. The default is 8.
confcut	numeric value that controls plotting confidence limits at the CDF extremes. Confidence limits for CDF values (percent scale) less than confcut or greater than 100 minus confcut are not plotted. A value of zero means confidence limits are plotted for the complete range of the CDF. The default is 5.
cex.main	expansion factor for the plot title. The default is 1.2.
...	additional arguments passed to the cdf.plot function.

**Value**

A PDF file containing the CDF plots.

**Author(s)**

Tony Olsen <Olsen.Tony@epa.gov>  
Tom Kincaid <Kincaid.Tom@epa.gov>

## References

Diaz-Ramos, S., D.L. Stevens, Jr., and A.R. Olsen. (1996). *EMAP Statistical Methods Manual*. EPA/620/R-96/XXX. Corvallis, OR: U.S. Environmental Protection Agency, Office of Research and Development, National Health Effects and Environmental Research Laboratory, Western Ecology Division.

## See Also

[cdf.plot](#)

## Examples

```
mysiteID <- paste("Site", 1:100, sep="")
mysites <- data.frame(siteID=mysiteID, Active=rep(TRUE, 100))
mysubpop <- data.frame(siteID=mysiteID, All.Sites=rep("All Sites",100),
  Resource.Class=rep(c("Good", "Poor"), c(55,45)))
mydesign <- data.frame(siteID=mysiteID, wgt=runif(100, 10, 100),
  xcoord=runif(100), ycoord=runif(100), stratum=rep(c("Stratum1",
  "Stratum2"), 50))
ContVar <- rnorm(100, 10, 1)
mydata.cont <- data.frame(siteID=mysiteID, ContVar=ContVar)
mypopsize <- list(All.Sites=c(Stratum1=3500, Stratum2=2000),
  Resource.Class=list(Good=c(Stratum1=2500, Stratum2=1500),
  Poor=c(Stratum1=1000, Stratum2=500)))
myanalysis <- cont.analysis(sites=mysites, subpop=mysubpop, design=mydesign,
  data.cont=mydata.cont, popsize=mypopsize)
cont.cdfplot("myanalysis.pdf", myanalysis$CDF, ylbl.r="Stream Length (km)")
```

---

cont.cdfctest

*Cumulative Distribution Function Inference for Probability Survey Data*

---

## Description

This function organizes input and output for conducting inference regarding cumulative distribution functions (CDFs) generated by a probability survey. Input can be either an object of class `spsurvey.analysis` (see the documentation for function `spsurvey.analysis`) or through use of the other arguments to this function.

## Usage

```
cont.cdfctest(sites=NULL, subpop=NULL, design=NULL, data.cont=NULL, popsize=NULL,
  popcorrect=FALSE, pcfsz=pcfsz, N.cluster=NULL, stage1size=NULL,
  sizeweight=FALSE, vartype="Local", testname="Wald_F", nclass=3,
  spsurvey.obj=NULL)
```

**Arguments**

sites	a data frame consisting of two variables: the first variable is site IDs, and the second variable is a logical vector indicating which sites to use in the analysis. If <code>spsurvey.obj</code> is not provided, then this argument is required. The default is <code>NULL</code> .
subpop	a data frame describing sets of populations and subpopulations for which estimates will be calculated. The first variable is site IDs. Each subsequent variable identifies a Type of population, where the variable name is used to identify Type. A Type variable identifies each site with one of the subpopulations of that Type. If <code>spsurvey.obj</code> is not provided, then this argument is required. The default is <code>NULL</code> .
design	a data frame consisting of design variables. If <code>spsurvey.obj</code> is not provided, then this argument is required. The default is <code>NULL</code> . Variables should be named as follows: siteID = site IDs wgt = final adjusted weights, which are either the weights for a single-stage sample or the stage two weights for a two-stage sample xcoord = x-coordinates for location, which are either the x-coordinates for a single-stage sample or the stage two x-coordinates for a two-stage sample ycoord = y-coordinates for location, which are either the y-coordinates for a single-stage sample or the stage two y-coordinates for a two-stage sample stratum = the stratum codes cluster = the stage one sampling unit (primary sampling unit or cluster) codes wgt1 = final adjusted stage one weights xcoord1 = the stage one x-coordinates for location ycoord1 = the stage one y-coordinates for location support = support values - the value one (1) for a site from a finite resource or the measure of the sampling unit associated with a site from an extensive resource, which is required for calculation of finite and continuous population correction factors swgt = size-weights, which is the stage two size-weight for a two-stage sample swgt1 = stage one size-weights
data.cont	a data frame of continuous response variables. The first variable is site IDs. Subsequent variables are response variables. Missing data ( <code>NA</code> ) is allowed. If <code>spsurvey.obj</code> is not provided, then this argument is required. The default is <code>NULL</code> .
popsize	known size of the resource, which is used to perform ratio adjustment to estimators expressed using measurement units for the resource. For a finite resource, this argument is either the total number of sampling units or the known sum of size-weights. For an extensive resource, this argument is the measure of the resource, i.e., either known total length for a linear resource or known total area for an areal resource. The argument must be in the form of a list containing an element for each population Type in the <code>subpop</code> data frame, where <code>NULL</code> is a valid choice for a population Type. The list must be named using the column names for the population Types in <code>subpop</code> . If a population Type doesn't contain subpopulations, then each element of the list is either a single value for an unstratified sample or a vector containing a value for each stratum for a stratified



sample, where elements of the vector are named using the stratum codes. If a population Type contains subpopulations, then each element of the list is a list containing an element for each subpopulation, where the list is named using the subpopulation names. The element for each subpopulation will be either a single value for an unstratified sample or a named vector of values for a stratified sample. The default is NULL.

Example popsize for a stratified sample:

```
popsize = list("Pop 1"=c("Stratum 1"=750,
"Stratum 2"=500,
"Stratum 3"=250),
"Pop 2"=list("SubPop 1"=c("Stratum 1"=350,
"Stratum 2"=250,
"Stratum 3"=150),
"SubPop 2"=c("Stratum 1"=250,
"Stratum 2"=150,
"Stratum 3"=100),
"SubPop 3"=c("Stratum 1"=150,
"Stratum 2"=150,
"Stratum 3"=75)),
"Pop 3"=NULL)
```

Example popsize for an unstratified sample:

```
popsize = list("Pop 1"=1500,
"Pop 2"=list("SubPop 1"=750,
"SubPop 2"=500,
"SubPop 3"=375),
"Pop 3"=NULL)
```

popcorrect	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation, where TRUE = use the correction factors and FALSE = do not use the correction factors. The default is FALSE.
pcfsize	size of the resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample this argument must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
N.cluster	the number of stage one sampling units in the resource, which is required for calculation of finite and continuous population correction factors for a two-stage sample. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
stage1size	size of the stage one sampling units of a two-stage sample, which is required for calculation of finite and continuous population correction factors for a two-stage sample and must have the names attribute set to identify the stage one sampling unit codes. For a stratified sample, the names attribute must be set to identify both stratum codes and stage one sampling unit codes using a convention where

	the two codes are separated by the & symbol, e.g., "Stratum 1&Cluster 1". The default is NULL.
sizeweight	a logical value that indicates whether size-weights should be used in the analysis, where TRUE = use the size-weights and FALSE = do not use the size-weights. The default is FALSE.
vartype	the choice of variance estimator, where "Local" = local mean estimator and "SRS" = SRS estimator. The default is "Local".
testname	name of the test statistic to be reported in the output data frame. Choices for the name are: "Wald", "Wald_F", "Mean_Eigenvalue", "Mean_Eigenvalue_F", "Satterthwaite", and "Satterthwaite_F". The default is "Wald_F".
nclass	number of classes into which the CDFs will be divided (binned), which must equal at least three. The default is 3.
spsurvey.obj	A list of class spsurvey.analysis that was produced by the function spsurvey.analysis. Depending on input to that function, some elements of the list may be NULL. The default is NULL.

### Details

For every response variable and every population Type, differences between CDFs are tested for every pair of subpopulations within a Type. The inferential procedures divide the CDFs into a discrete set of intervals (classes) and then utilize procedures that have been developed for analysis of categorical data from probability surveys. Choices for inference are the Wald, Rao-Scott first order corrected (mean eigenvalue corrected), and Rao-Scott second order corrected (Satterthwaite corrected) test statistics. Both standard versions of the three statistics, which are distributed as Chi-squared random variables, and alternate version of the statistics, which are distributed as F random variables, are available. The default test statistic is the F distribution version of the Wald statistic.

### Value

A data frame of CDF test results for all pairs of subpopulations within each population Type for every response variable. The data frame includes the test statistic specified by argument testname plus its degrees of freedom and p-value.

### Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>

### References

Kincaid, T.M. (2000). Testing for differences between cumulative distribution functions from complex environmental sampling surveys. In *2000 Proceeding of the Section on Statistics and the Environment*, American Statistical Association, Alexandria, VA.

### See Also

[cdf.test](#)

**Examples**

```

n <- 200
mysiteID <- paste("Site", 1:n, sep="")
mysites <- data.frame(siteID=mysiteID, Active=rep(TRUE, n))
mysubpop <- data.frame(siteID=mysiteID, Resource_Class=sample(c("Agr",
  "Forest", "Urban"), n, replace=TRUE))
mydesign <- data.frame(siteID=mysiteID, wgt=runif(n, 10, 100),
  xcoord=runif(n), ycoord=runif(n), stratum=rep(c("Stratum1",
  "Stratum2"), n/2))
mypopsize <- list(Resource_Class=list(Agr=c(Stratum1=2500, Stratum2=1500),
  Forest=c(Stratum1=1000, Stratum2=500), Urban=c(Stratum1=600, Stratum2=450)))
ContVar <- numeric(n)
tst <- mysubpop$Resource_Class == "Agr"
ContVar[tst] <- rnorm(sum(tst), 10, 1)
tst <- mysubpop$Resource_Class == "Forest"
ContVar[tst] <- rnorm(sum(tst), 10.1, 1)
tst <- mysubpop$Resource_Class == "Urban"
ContVar[tst] <- rnorm(sum(tst), 10.5, 1)
mydata.cont <- data.frame(siteID=mysiteID, ContVar=ContVar)
cont.cdfctest(sites=mysites, subpop=mysubpop, design=mydesign,
  data.cont=mydata.cont, popsize=mypopsize, testname="Mean_Eigenvalue")

```

decon\_data

*Small Lakes in Florida***Description**

A data frame containing attributes for small lakes in Florida.

**Usage**

```
data(decon_data)
```

**Format**

A data frame with 930 rows and the following 6 columns:

**xcoord** simulated x-coordinate value.

**ycoord** simulated y-coordinate value.

**Richness** simulated species richness value.

**Richness\_25** species richness value plus 25% measurement error variance.

**Richness\_50** species richness value plus 50% measurement error variance.

**Richness\_100** species richness value plus 100% measurement error variance.

**Examples**

```
data(decon_data)
head(decon_data)
```

---

`dsgnsum`*Summarize the Sites Selected for a Survey Design*

---

**Description**

This function summarizes the sites selected for a survey design by producing contingency tables containing the cross-tabulation of number of sites for survey design variables and, optionally, for auxiliary variables.

**Usage**

```
dsgnsum(spsample, auxvar = NULL)
```

**Arguments**

<code>spsample</code>	an object of class <code>SpatialDesign</code> produced by either the <code>grts</code> or <code>irs</code> functions that contains survey design information and additional attribute (auxiliary) variables.
<code>auxvar</code>	a vector containing the names of columns in the data slot of the <code>SpatialDesign</code> object that identify auxiliary variables to be used to summarize the survey design.

**Value**

A list containing the following components:

<code>DesignSum</code>	a list of contingency tables containing the cross-tabulation of number of sites for the following combinations of survey design variables: (1) multidensity category ( <code>mdcaty</code> ) and <code>stratum</code> (2) <code>stratum</code> and <code>panel</code> (3) <code>mdcaty</code> , <code>panel</code> , and <code>stratum</code>
<code>AuxVarSum</code>	a list of contingency tables containing the cross-tabulation of number of sites for each auxiliary variable and the design variables <code>mdcaty</code> , <code>panel</code> , and <code>stratum</code>

In addition, the output list plus labeling information is printed to the console.

**Author(s)**

Tony Olsen <Olsen.Tony@epa.gov>  
Tom Kincaid <Kincaid.Tom@epa.gov>

**References**

Stevens, D.L., Jr., and A.R. Olsen. (2004). Spatially-balanced sampling of natural resources. *Journal of the American Statistical Association* **99**: 262-278.

**See Also**

[grts framesum](#)

**Examples**

```
## Not run:
design <- list(Stratum1=list(panel=c(PanelOne=50), seltype="Equal", over=10),
             Stratum2=list(panel=c(PanelOne=50, PanelTwo=50), seltype="Unequal",
                             caty.n=c(CatyOne=25, CatyTwo=25, CatyThree=25, CatyFour=25), over=75))
attframe <- read.dbf("shapefile")
samp <- grts(design=design, DesignID="Test.Site", type.frame="area",
            src.frame="shapefile", in.shape="shapefile", att.frame=attframe,
            stratum="stratum", mdcaty="mdcaty", shapefile=TRUE,
            shapefilename="sample")
dsgnsum(samp, auxvar=c("ecoregion", "state"))

## End(Not run)
```

---

FL\_lakes

*Small Lakes in Florida*


---

**Description**

A data frame containing attributes for small lakes in Florida.

**Usage**

```
data(FL_lakes)
```

**Format**

A data frame with 930 rows and the following 11 columns:

**siteID** site ID value.

**xcoord** Albers projection x-coordinate.

**ycoord** Albers projection y-coordinate.

**wgt** survey design weight.

**Basin** stream basin code.

**Status** site evaluation status code.

**TNT** target or nontarget category for the site evaluation status code.

**pH\_Cat** IBI (pH category).

**Coliform\_Cat** fecal coliform count category.

**Oxygen** dissolved oxygen value.

**Turbidity** turbidity value.

**Examples**

```
data(FL_lakes)
head(FL_lakes)
```

framesum

*Summarize Frame Size for a Survey Design***Description**

This function summarizes the frame for a survey design. When `type.frame` equals "finite", summary is a count of number of units in `att.frame` for cross-tabulation of `stratum`, `mdcaty`, and `auxvar`. When `type.frame` equals "linear" or "area", summary is the sum of length or area for units for cross-tabulation of `stratum`, `mdcaty`, and `auxvar`. Note that length and area are taken from `length_mdm` and `area_mdm`, which are calculated by the function `read.dbf` and added to `att.frame`. If argument `mdcaty` or argument `stratum` equals NULL or if both arguments equal NULL, then the cross-tabulation is performed without use of the design variable(s).

**Usage**

```
framesum(att.frame, design, type.frame="finite", stratum=NULL, mdcaty=NULL,
         auxvar=NULL, units.in="Number", scale=1, units.out="Number")
```

**Arguments**

<code>att.frame</code>	a data frame composed of attributes associated with elements in the frame, which must contain the columns used for <code>stratum</code> and <code>mdcaty</code> (if required by the survey design).
<code>design</code>	named list of <code>stratum</code> design specifications which are also lists. <code>Stratum</code> names must be subset of values in <code>stratum</code> argument. Each <code>stratum</code> list has four list components: <code>panel</code> = named vector of sample sizes for each panel in <code>stratum</code> ; <code>seltype</code> = the type of random selection, which must be one of following: "Equal" - equal probability selection, "Unequal" - unequal probability selection by the categories specified in <code>caty.n</code> and <code>mdcaty</code> , or "Continuous" - unequal probability selection proportional to auxiliary variable <code>mdcaty</code> ; <code>caty.n</code> = if <code>seltype</code> equals "Unequal", a named vector of sample sizes for each category specified by <code>mdcaty</code> , where sum of the sample sizes must equal sum of the panel sample sizes, and names must be a subset of values in <code>mdcaty</code> ; <code>over</code> = number of replacement sites ("oversample" sites) for the entire design, which is set equal to 0 if none are required.
<code>type.frame</code>	the type of frame, which must be one of following: "finite", "linear", or "area". The default is "finite"
<code>stratum</code>	name of the column from <code>att.frame</code> that identifies <code>stratum</code> membership for each element in the frame. If <code>stratum</code> equals NULL, the design is unstratified. The default is NULL.
<code>mdcaty</code>	name of the column from <code>att.frame</code> that identifies the unequal probability category for each element in the frame. The default is NULL.
<code>auxvar</code>	a vector containing the names of columns from sites that identify auxiliary variables to be used to summarize frame size. The default is NULL.

units.in	a character string giving the name of units used to measure size in the frame. The default is "Number".
scale	the scale factor used to change units.in to units.out. For example, use 1000 to change "Meters" to "Kilometers". The default is 1.
units.out	a character string giving the name of units used to measure size in the results. The default is "Number".

### Value

A list containing the following components:

DesignSize	a table (for type.frame equals "finite") or an array (for type.frame equals "linear" or "area") that contains the cross-tabulation of frame extent for design variables multidensity category (mdcaty) and stratum, where extent of the frame is the number of sites for type.frame equals "finite", the sum of site length for type.frame equals "linear", or the sum of site area for type.frame equals "area".
AuxVarSize	a list containing a component for each auxiliary variable, where each component of the list is one of the following: (1) if the type of random selection does not equal "Continuous" for any stratum, each component is either a table (for type.frame equals "finite") or an array (for type.frame equals "linear" or "area") that contains the cross-tabulation of frame extent for mdcaty, stratum, and the auxiliary variable or (2) if the type of random selection equals "Continuous" for all strata, each component is either a table (finite frame) or an array (linear or area frame) containing the cross-tabulation of frame extent for stratum and the auxiliary variable.

In addition the output list plus labeling information is printed to the console.

### Author(s)

Tony Olsen <Olsen.Tony@epa.gov>  
Tom Kincaid <Kincaid.Tom@epa.gov>

### References

Stevens, D.L., Jr., and A.R. Olsen. (2004). Spatially-balanced sampling of natural resources. *Journal of the American Statistical Association* **99**: 262-278.

### See Also

[grts dsgnsum](#)

### Examples

```
## Not run:
attframe <- read.dbf("shapefile")
design <- list(Stratum1=list(panel=c(PanelOne=50), seltype="Equal", over=10),
  Stratum2=list(panel=c(PanelOne=50, PanelTwo=50), seltype="Unequal",
  caty.n=c(CatyOne=25, CatyTwo=25, CatyThree=25, CatyFour=25), over=75))
framesum(att.frame=attframe, design=design, type.frame="area",
```

```

stratum="stratum", mdcaty="mdcaty", auxvar=c("ecoregion",
"state"), units.in="Meters", scale=1000, units.out="Kilometers")

## End(Not run)

```

---

geodalbers                      *Project Latitude and Longitude (Spheroid) to Albers Projection in Plane*

---

### Description

Project spheroid models of the globe to Albers projection in the plane.

### Usage

```
geodalbers(lon, lat, sph="GRS80", clon=-96, clat=23, sp1=29.5, sp2=45.5)
```

### Arguments

lon	longitude (decimal degrees) vector to be projected using Albers.
lat	latitude (decimal degrees) vector to be projected using Albers.
sph	spheroid options: Clarke1866, GRS80, WGS84. The default is GRS80.
clon	center longitude (decimal degrees). The default is -96.
clat	origin latitude (decimal degrees). The default is 23.
sp1	standard parallel 1 (decimal degrees). The default is 29.5.
sp2	standard parallel 2 (decimal degrees). The default is 45.5.

### Details

Ask Denis White.

### Value

A data frame of Albers x-coordinate and y-coordinate projections for latitude and longitude.

### Author(s)

Tony Olsen <Olsen.Tony@epa.gov>

### References

J. Snyder, USGS Professional Paper 1395



**Description**

Selects a sample using a generalized random-tessellation stratified (GRTS) survey design. The GRTS survey design may include stratification, unequal probability using categories, unequal selection proportional to an auxiliary variable, survey over time structures, and provision for an over-sample.

**Usage**

```
grts(design, DesignID="Site", SiteBegin=1, type.frame="finite",
     src.frame="shapefile", in.shape=NULL, sp.object=NULL, att.frame=NULL,
     id=NULL, xcoord=NULL, ycoord=NULL, stratum=NULL, mdcaty=NULL, startlev=NULL,
     maxlev=11, maxtry=1000, shift.grid=TRUE, do.sample=rep(TRUE, length(design)),
     shapefile=TRUE, prjfilename=NULL, out.shape="sample")
```

**Arguments**

`design` named list of stratum design specifications, where each element of `design` is a list containing the design specifications for a stratum. For an unstratified sample, `design` contains a single list. If the sample is stratified, the names in `design` must occur among the strata names in the stratum column of the attributes data frame (`att.frame`). If the sample is unstratified, the name of the single list in `design` is arbitrary. Each list in `design` has four components:

- `panel` = named vector of sample sizes for each panel in stratum
- `seltype` = the type of random selection, which must be one of following: "Equal" - equal probability selection, "Unequal" - unequal probability selection by the categories specified in `caty.n` and `mdcaty`, or "Continuous" - unequal probability selection proportional to auxiliary variable `mdcaty`
- `caty.n` = if `seltype` equals "Unequal", a named vector of sample sizes for each category specified by `mdcaty`, where sum of the sample sizes must equal sum of the panel sample sizes, and names must be a subset of values in `mdcaty`
- `over` = number of replacement sites ("oversample" sites) for the entire design, which is set equal to 0 if none are required

Example design for a stratified sample:

```
design <- list("Stratum 1"=list(panel=c(Panel=50), seltype="Equal", over=10),
             "Stratum 2"=list(panel=c("Panel One"=50, "Panel Two"=50), seltype="Unequal",
                               caty.n=c(CatyOne=25, CatyTwo=25, CatyThree=25, CatyFour=25), over=75))
```

Example design for an unstratified sample:

```
design <- list(None=list(panel=c(Panel1=50, Panel2=100, Panel3=50), seltype="Unequal",
                              caty.n=c("Caty 1"=50, "Caty 2"=25, "Caty 3"=25, "Caty 4"=25, "Caty 5"=75),
                              over=100))
```

DesignID	name for the design, which is used to create a site ID for each site. The default is "Site".
SiteBegin	number to use for first site in the design. The default is 1.
type.frame	the type of frame, which must be one of following: "finite", "linear", or "area". The default is "finite".
src.frame	source of the frame, which equals "shapefile" if the frame is to be read from a shapefile, "sp.object" if the frame is obtained from an sp package object, or "att.frame" if type.frame equals "finite" and the frame is included in att.frame. The default is "shapefile".
in.shape	name (without any extension) of the input shapefile. If src.frame equal "shapefile" and in.shape equals NULL, then the shapefile or shapefiles in the working directory are used. The default is NULL.
sp.object	name of the sp package object when src.frame equals "sp.object". The default is NULL.
att.frame	a data frame composed of attributes associated with elements in the frame, which must contain the columns used for stratum and mdcaty (if required). If src.frame equals "shapefile" and att.frame equals NULL, then att.frame is created from the dbf file(s) in the working directory. If src.frame equals "sp.object" and att.frame equals NULL, then att.frame is created from the sp object. If src.frame equals "att.frame", then att.frame must include columns that contain x-coordinates and y-coordinates for each element in the frame. The default is NULL.
id	name of the column from att.frame that identifies the ID value for each element in the frame. If id equals NULL, a column named "id" that contains values from one through the number of rows in att.frame is added to att.frame. The default is NULL.
xcoord	name of the column from att.frame that identifies x-coordinates when src.frame equals "att.frame". If xcoord equals NULL, then xcoord is given the value "x". The default is NULL.
ycoord	name of the column from att.frame that identifies y-coordinates when src.frame equals "att.frame". If ycoord equals NULL, then ycoord is given the value "y". The default is NULL.
stratum	name of the column from att.frame that identifies stratum membership for each element in the frame. If stratum equals NULL, the design is unstratified, and a column named "stratum" (with all its elements equal to the stratum name specified in design) is added to att.frame. The default is NULL.
mdcaty	name of the column from att.frame that identifies the unequal probability category for each element in the frame. The default is NULL.
startlev	initial number of hierarchical levels to use for the GRTS grid, which must be less than or equal to maxlev (if maxlev is specified) and cannot be greater than 11. The default is NULL.
maxlev	maximum number of hierarchical levels to use for the GRTS grid, which cannot be greater than 11. The default is 11.
maxtry	maximum number of iterations for randomly generating a point within a grid cell to select a site when type.frame equals "area". The default is 1000.

<code>shift.grid</code>	option to randomly shift the hierarchical grid, where TRUE means shift the grid and FALSE means do not shift the grid, which is useful if one desires strict spatial stratification by hierarchical grid cells. The default is TRUE.
<code>do.sample</code>	named vector that provides the option controlling sample selection for each stratum, where TRUE means select a sample from a stratum and FALSE means return the sample frame for a stratum in reverse hierarchical order. Note that FALSE can only be used when <code>type.frame</code> equals "points" and <code>seltype</code> equals "Equal". Names for the vector must match the names in design. If the vector is not named, then the names in design are used. The default is TRUE for each stratum.
<code>shapefile</code>	option to create a shapefile containing the survey design information, where TRUE equals create a shapefile and FALSE equals do not create a shapefile. The default is TRUE.
<code>prjfilename</code>	name (without any extension) of the projection file for the input shapefile, which is use to name the projection file for the output shapefile. The default is NULL.
<code>out.shape</code>	name (without any extension) of the output shapefile containing the survey design information. The default is "sample".

### Details

The GRTS survey design process selects a spatially balanced sample based on the survey design specification.

Function `dsgnsum()`, can be used to summarize the sites selected for a survey design.

### Value

An `sp` package object containing the survey design information and any additional attribute variables that were provided. The object is assigned class "SpatialPointsDataFrame". For further information regarding the output object, see documentation for the `sp` package. Optionally, a shapefile can be created that contains the survey design information.

### Author(s)

Tony Olsen <Olsen.Tony@epa.gov>  
Tom Kincaid <Kincaid.Tom@epa.gov>

### References

Stevens, D.L., Jr., and A.R. Olsen. (2004). Spatially-balanced sampling of natural resources. *Journal of the American Statistical Association* **99**, 262-278.

### See Also

[grtspts](#) [grtslin](#) [grtsarea](#) [albersgeod](#) [dsgnsum](#)

## Examples

```
## Not run:
The following example will select a sample from an area resource. The design
includes two strata. For Stratum 1, an equal probability sample of size 50
will be selected for a single panel. For Stratum 2, an unequal probability
sample of size 50 will be selected for each of two panels. The sample for
Stratum 2 will be apportioned into samples of size 25 for each of four
unequal probability categories. In addition both strata will include
oversamples (size 10 for Stratum 1 and size 75 for Stratum 2). It is assumed
that a shapefile defining the polygons for the area resource is located in the
folder from which R is started. Attribute data for the design will be read
from the dbf file of the shapefile, which is assumed to have variables named
"test.stratum" and "test.mdcaty" that specify stratum membership value and
unequal probability category, respectively, for each record in the shapefile.
A shapefile named "test.sample" containing the survey design information will
be created.
test.design <- list("Stratum 1"=list(panel=c(Panel=50), seltype="Equal",
  over=10), "Stratum 2"=list(panel=c("Panel One"=50, "Panel Two"=50),
  seltype="Unequal", caty.n=c(CatyOne=25, CatyTwo=25, CatyThree=25,
  CatyFour=25), over=75))
test.attframe <- read.dbf("test.shapefile")
test.sample <- grts(design=test.design, DesignID="Test.Site", type.frame="area",
  src.frame="shapefile", in.shape="test.shapefile", att.frame=test.attframe,
  stratum="test.stratum", mdcaty="test.mdcaty", shapefile=TRUE,
  out.shape="test.sample")

## End(Not run)
```

---

grtsarea

*Select GRTS Sample of an Area Resource*

---

## Description

This function select a GRTS sample of an area resource. This function uses hierarchical randomization to ensure that the sample will include no more than one point per cell and then picks a point in selected cells.

## Usage

```
grtsarea(shapefilename=NULL, areaframe, samplesize=100, SiteBegin=1,
  shift.grid=TRUE, startlev=NULL, maxlev=11, maxtry=1000)
```

## Arguments

shapefilename	name of the input shapefile. If shapefilename equals NULL, then the shapefile or shapefiles in the working directory are used. The default is NULL.
areaframe	a data frame containing id, mdcaty and mdm.
samplesize	number of points to select in the sample. The default is 100.

SiteBegin	number to use for first site in the design. The default is 1.
shift.grid	option to randomly shift the hierarchical grid, where TRUE means shift the grid and FALSE means do not shift the grid, which is useful if one desires strict spatial stratification by hierarchical grid cells. The default is TRUE.
startlev	initial number of hierarchical levels to use for the GRTS grid, which must be less than or equal to maxlev (if maxlev is specified) and cannot be greater than 11. The default is NULL.
maxlev	maximum number of hierarchical levels to use for the GRTS grid, which cannot be greater than 11. The default is 11.
maxtry	maximum number of iterations for randomly generating a point The default is 1000.

**Value**

A data frame of GRTS sample points containing: SiteID, id, x, y, mdcaty, and weight.

**Author(s)**

Tony Olsen <Olsen.Tony@epa.gov>  
Tom Kincaid <Kincaid.Tom@epa.gov>

**References**

Stevens, D.L., Jr., and A.R. Olsen. (2004). Spatially-balanced sampling of natural resources. *Journal of the American Statistical Association* **99**, 262-278.

**See Also**

[grts](#)

---

grtslin

*Select GRTS Sample of a Linear Resource*


---

**Description**

This function select a GRTS sample of a linear resource. This function uses hierarchical randomization to ensure that the sample will include no more than one point per cell and then picks a point in selected cells.

**Usage**

```
grtslin(shapefilename=NULL, linframe, samplesize=100, SiteBegin=1,
        shift.grid=TRUE, startlev=NULL, maxlev=11)
```

**Arguments**

shapefilename	name of the input shapefile. If shapefilename equals NULL, then the shapefile or shapefiles in the working directory are used. The default is NULL.
linframe	a data frame containing id, mdcaty and mdm.
samplesize	number of points to select in the sample. The default is 100.
SiteBegin	number to use for first site in the design. The default is 1.
shift.grid	option to randomly shift the hierarchical grid, where TRUE means shift the grid and FALSE means do not shift the grid, which is useful if one desires strict spatial stratification by hierarchical grid cells. The default is TRUE.
startlev	initial number of hierarchical levels to use for the GRTS grid, which must be less than or equal to maxlev (if maxlev is specified) and cannot be greater than 11. The default is NULL.
maxlev	maximum number of hierarchical levels to use for the GRTS grid, which cannot be greater than 11. The default is 11.

**Value**

A data frame of GRTS sample points containing: SiteID, id, x, y, mdcaty, and weight.

**Author(s)**

Tony Olsen <Olsen.Tony@epa.gov>  
Tom Kincaid <Kincaid.Tom@epa.gov>

**References**

Stevens, D.L., Jr., and A.R. Olsen. (2004). Spatially-balanced sampling of natural resources. *Journal of the American Statistical Association* **99**, 262-278.

**See Also**

[grts](#)

---

grtspts

*Select GRTS Sample of a Finite Resource*

---

**Description**

This function select a GRTS sample of a finite resource. This function uses hierarchical randomization to ensure that the sample will include no more than one point per cell and then picks a point in selected cells.

**Usage**

```
grtspts(src.frame="shapefile", shapefilename=NULL, ptsframe, samplesize=100,
        SiteBegin=1, shift.grid=TRUE, do.sample=TRUE, startlev=NULL, maxlev=11)
```

**Arguments**

src.frame	source of the frame, which equals "shapefile" if the frame is to be read from a shapefile, or "att.frame" if the frame is included in ptsframe. The default is "shapefile".
shapefilename	name of the input shapefile. If src.frame equal "shapefile" and shapefilename equals NULL, then the shapefile or shapefiles in the working directory are used. The default is NULL.
ptsframe	a data frame containing id, x, y, mdcaty, and mdm.
samplesize	number of points to select in the sample. The default is 100.
SiteBegin	number to use for first site in the design. The default is 1.
shift.grid	option to randomly shift the hierarchical grid, where TRUE means shift the grid and FALSE means do not shift the grid, which is useful if one desires strict spatial stratification by hierarchical grid cells. The default is TRUE.
do.sample	option to select a sample, where TRUE means select a sample and FALSE means return the entire sample frame in reverse hierarchical order. The default is TRUE.
startlev	initial number of hierarchical levels to use for the GRTS grid, which must be less than or equal to maxlev (if maxlev is specified) and cannot be greater than 11. The default is NULL.
maxlev	maximum number of hierarchical levels to use for the GRTS grid, which cannot be greater than 11. The default is 11.

**Value**

A data frame of GRTS sample points containing: SiteID, id, x, y, mdcaty, and weight.

**Author(s)**

Tony Olsen <Olsen.Tony@epa.gov>  
Tom Kincaid <Kincaid.Tom@epa.gov>

**References**

Stevens, D.L., Jr., and A.R. Olsen. (2004). Spatially-balanced sampling of natural resources. *Journal of the American Statistical Association* **99**, 262-278.

**See Also**

[grts](#)

---

 IN\_streams

*Streams in the Upper Wabash Basin in Indiana*


---

### Description

A data frame containing attributes for streams in Indiana.

### Usage

```
data(IN_streams)
```

### Format

A data frame with 100 rows and the following 11 columns:

**siteID** site ID value.

**xcoord** Albers projection x-coordinate.

**ycoord** Albers projection y-coordinate.

**wgt** survey design weight.

**Strahler\_Cat** Strahler order category.

**Status** site evaluation status code.

**TNT** target or nontarget category for the site evaluation status code.

**IBI\_Score** IBI (index of biotic integrity) score.

**IBI\_Status** status category of the IBI score.

**QHEI\_Score** QHEI ((qualitative habitat evaluation index) score.

**QHEI\_Status** status category of the QHEI score.

### Examples

```
data(IN_streams)
head(IN_streams)
```

---

 irs

*Independent Random Sample (IRS) Survey Design*


---

### Description

Selects an independent random sample (IRS) survey design. The IRS survey design may include stratification, unequal probability using categories, unequal selection proportional to an auxiliary variable, survey over time structures, and provision for an oversample.



**Usage**

```
irs(design, DesignID="Site", SiteBegin=1, type.frame="finite",
    src.frame="shapefile", in.shape=NULL, sp.object=NULL, att.frame=NULL,
    id=NULL, xcoord=NULL, ycoord=NULL, stratum=NULL, mdcaty=NULL, maxtry=1000,
    shapefile=TRUE, prjfilename=NULL, out.shape="sample")
```

**Arguments**

**design**            named list of stratum design specifications, where each element of design is a list containing the design specifications for a stratum. For an unstratified sample, design contains a single list. If the sample is stratified, the names in design must occur among the strata names in the stratum column of the attributes data frame (att.frame). If the sample is unstratified, the name of the single list in design is arbitrary. Each list in design has four components:  
panel = named vector of sample sizes for each panel in stratum  
seltype = the type of random selection, which must be one of following: "Equal" - equal probability selection, "Unequal" - unequal probability selection by the categories specified in caty.n and mdcaty, or "Continuous" - unequal probability selection proportional to auxiliary variable mdcaty  
caty.n = if seltype equals "Unequal", a named vector of sample sizes for each category specified by mdcaty, where sum of the sample sizes must equal sum of the panel sample sizes, and names must be a subset of values in mdcaty  
over = number of replacement sites ("oversample" sites) for the entire design, which is set equal to 0 if none are required

Example design for a stratified sample:

```
design = list("Stratum 1"=list(panel=c(Panel=50), seltype="Equal", over=10),
"Stratum 2"=list(panel=c("Panel One"=50, "Panel Two"=50), seltype="Unequal",
caty.n=c(CatyOne=25, CatyTwo=25, CatyThree=25, CatyFour=25), over=75))
```

Example design for an unstratified sample:

```
design = list(None=list(panel=c(Panel1=50, Panel2=100, Panel3=50), seltype="Unequal",
caty.n=c("Caty 1"=50, "Caty 2"=25, "Caty 3"=25, "Caty 4"=25, "Caty 5"=75),
over=100))
```

**DesignID**        name for the design, which is used to create a site ID for each site. The default is "Site".

**SiteBegin**       number to use for first site in the design. The default is 1.

**type.frame**      the type of frame, which must be one of following: "finite", "linear", or "area". The default is "finite".

**src.frame**       source of the frame, which equals "shapefile" if the frame is to be read from a shapefile, "sp.object" if the frame is obtained from an sp package object, or "att.frame" if type.frame equals "finite" and the frame is included in att.frame. The default is "shapefile".

**in.shape**        name (without any extension) of the input shapefile. If src.frame equal "shapefile" and in.shape equals NULL, then the shapefile or shapefiles in the working directory are used. The default is NULL.

<code>sp.object</code>	name of the sp package object when <code>src.frame</code> equals "sp.object". The default is NULL.
<code>att.frame</code>	a data frame composed of attributes associated with elements in the frame, which must contain the columns used for <code>stratum</code> and <code>mdcaty</code> (if required). If <code>src.frame</code> equals "shapefile" and <code>att.frame</code> equals NULL, then <code>att.frame</code> is created from the dbf file(s) in the working directory. If <code>src.frame</code> equals "att.frame", then <code>att.frame</code> includes columns that contain x-coordinates and y-coordinates for each element in the frame. The default is NULL.
<code>id</code>	name of the column from <code>att.frame</code> that identifies the ID value for each element in the frame. If <code>id</code> equals NULL, a column named "id" that contains values from one through the number of rows in <code>att.frame</code> is added to <code>att.frame</code> . The default is NULL.
<code>xcoord</code>	name of the column from <code>att.frame</code> that identifies x-coordinates when <code>src.frame</code> equals "att.frame". If <code>xcoord</code> equals NULL, then <code>xcoord</code> is given the value "x". The default is NULL.
<code>ycoord</code>	name of the column from <code>att.frame</code> that identifies y-coordinates when <code>src.frame</code> equals "att.frame". If <code>ycoord</code> equals NULL, then <code>ycoord</code> is given the value "y". The default is NULL.
<code>stratum</code>	name of the column from <code>att.frame</code> that identifies stratum membership for each element in the frame. If <code>stratum</code> equals NULL, the design is unstratified, and a column named "stratum" (with all its elements equal to the stratum name specified in design) is added to <code>att.frame</code> . The default is NULL.
<code>mdcaty</code>	name of the column from <code>att.frame</code> that identifies the unequal probability category for each element in the frame. The default is NULL.
<code>maxtry</code>	maximum number of iterations for randomly generating a point within the frame to select a site when <code>type.frame</code> equals "area". The default is 1000.
<code>shapefile</code>	option to create a shapefile containing the survey design information, where TRUE equals create a shapefile and FALSE equals do not create a shapefile. The default is TRUE.
<code>prjfilename</code>	name (without any extension) of the project file for an input shapefile. The default is NULL.
<code>out.shape</code>	name (without any extension) of the output shapefile containing the survey design information. The default is "sample".

### Details

The IRS survey design process selects a sample based on the survey design specification.

Function `dsgnsum()`, can be used to summarize the sites selected for a survey design.

### Value

An sp package object containing the survey design information and any additional attribute variables that were provided. The object is assigned class "SpatialPointsDataFrame". For further information regarding the output object, see documentation for the sp package. Optionally, a shapefile can be created that contains the survey design information.

**Author(s)**

Tom Kincaid <Kincaid.Tom@epa.gov>

**See Also**

[irspts](#) [irslin](#) [irsarea](#) [dsgnsum](#)

**Examples**

```
## Not run:
# The following example will select a sample from an area resource. The design
# includes two strata. For Stratum 1, a sample of size 50 will be selected for
# a single panel. For Stratum 2, a sample of size 50 will be selected for each
# of two panels. In addition both strata will include oversamples (size 10 for
# Stratum 1 and size 75 for Stratum 2). It is assumed that a shapefile defining
# the polygons for the area resource is located in the folder from which R is
# started. Attribute data for the design will be read from the dbf file of the
# shapefile, which is assumed to have a variable named "test.stratum" that
# specifies stratum membership value for each record in the shapefile. A
# shapefile named "test.sample" containing the survey design information will be
# created.
test.design <- list("Stratum 1"=list(panel=c(Panel=50), seltype="Equal",
  over=10), "Stratum 2"=list(panel=c("Panel One"=50, "Panel Two"=50),
  seltype="Unequal", caty.n=c(CatyOne=25, CatyTwo=25, CatyThree=25,
  CatyFour=25), over=75))
test.attframe <- read.dbf("test.shapefile")
test.sample <- irs(design=test.design, DesignID="Test.Site", type.frame="area",
  src.frame="shapefile", in.shape="test.shapefile", att.frame=test.attframe,
  stratum="test.stratum", mdcaty="test.mdcaty", shapefile=TRUE,
  out.shape="test.sample")

## End(Not run)
```

---

 irsarea

---

*Select an Independent Random Sample (IRS) of an Area Resource*


---

**Description**

This function selects an independent random sample (IRS) of an area resource.

**Usage**

```
irsarea(shapefilename=NULL, areaframe, samplesize=100, SiteBegin=1, maxtry=1000)
```

**Arguments**

shapefilename	name of the input shapefile. If shapefilename equals NULL, then the shapefile or shapefiles in the working directory are used. The default is NULL.
areaframe	a data frame containing id, mdcaty and mdm.
samplesize	number of points to select in the sample. The default is 100.
SiteBegin	number to use for first site in the design. The default is 1.
maxtry	maximum number of iterations for randomly generating a point within the frame to select a site when type.frame equals "area". The default is 1000.

**Value**

A data frame of IRS sample points containing: SiteID, id, x, y, mdcaty, and weight.

**Author(s)**

Tom Kincaid <Kincaid.Tom@epa.gov>

**See Also**

[irs](#)

---

 irslin

*Select an Independent Random Sample (IRS) of a Linear Resource*

---

**Description**

This function selects an independent random sample (IRS) of a linear resource.

**Usage**

```
irslin(shapefilename=NULL, linframe, samplesize=100, SiteBegin=1)
```

**Arguments**

shapefilename	name of the input shapefile. If shapefilename equals NULL, then the shapefile or shapefiles in the working directory are used. The default is NULL.
linframe	a data frame containing id, mdcaty and mdm.
samplesize	number of points to select in the sample. The default is 100.
SiteBegin	number to use for first site in the design. The default is 1.

**Value**

A data frame of IRS sample points containing: SiteID, id, x, y, mdcaty, and weight.

**Author(s)**

Tom Kincaid <Kincaid.Tom@epa.gov>

**See Also**

[irs](#)

---

irspts

*Select an Independent Random Sample (IRS) of a Finite Resource*

---

**Description**

This function selects an independent random sample (IRS) of a finite resource.

**Usage**

```
irspts(ptsframe, samplesize=100, SiteBegin=1)
```

**Arguments**

ptsframe	a data frame containing id, x, y, mdcaty, and mdm.
samplesize	number of points to select in the sample. The default is 100.
SiteBegin	number to use for first site in the design. The default is 1.

**Value**

A data frame of IRS sample points containing: SiteID, id, x, y, mdcaty, and weight.

**Author(s)**

Tom Kincaid <Kincaid.Tom@epa.gov>

**See Also**

[irs](#)

---

Luck_Ash_streams	<i>Streams in the Luckiamute Watershed in Oregon</i>
------------------	--

---

**Description**

An object of class "SpatialLinesDataFrame" that was created from a shapefile of streams in the Luckiamute watershed in Oregon.

**Usage**

```
data(Luck_Ash_streams)
```

**Format**

The format is: Formal class "SpatialLinesDataFrame" [package "sp"].  
The data slot contains 429 rows and the following 3 columns:

**Per\_Int** stream type for the line segment.

**Level3\_Nam** Strahler order category for the line segment.

**Length\_km** length of the line segment in kilometers.

**Examples**

```
data(Luck_Ash_streams)
plot(Luck_Ash_streams)
```

---

marinus	<i>Convert Coordinates from Latitude/Longitude to the Equidistant, Cylindric Map Projection</i>
---------	---

---

**Description**

This function converts x,y coordinates measured in units of latitude and longitude, i.e., geographic coordinates measured in decimal degrees, to coordinates in the equidistant, cylindric map projection measured in units of kilometers. The projection center is defined as the midpoint in latitude-longitude space. The map projection is named after Marinus of Tyre.

**Usage**

```
marinus(lat, lon)
```

**Arguments**

lat                vector of latitudes.

lon                vector of longitudes.

**Value**

A data frame with column names "x" and "y" containing the x and y coordinates in the equidistant, cylindrical map projection measured in units of kilometers.

**Author(s)**

Denis White <White.Denis@epa.gov>

**References**

J.P. Snyder. USGS Prof. Paper 1395, p. 90.

**Examples**

```
lat <- 45 + runif(100, -5, 5)
lon <- 120 + runif(100, -10, 10)
marinus(lat, lon)
```

---

NE\_lakes

*Lakes in the Southern New England Region of the U.S.*

---

**Description**

An object of class "SpatialPointsDataFrame" that was created from a shapefile of lakes in the southern New England region of the U.S.

**Usage**

```
data(NE_lakes)
```

**Format**

The format is: Formal class "SpatialPointsDataFrame" [package "sp"].  
The data slot contains 6,121 rows and the following 4 columns:

**xcoord** Albers projection x-coordinate of the lake centroid.

**ycoord** Albers projection y-coordinate of the lake centroid.

**State** state code for the lake.

**Area\_Cat** lake surface area category in hectares.

**Examples**

```
data(NE_lakes)
plot(NE_lakes)
```

---

NLA\_2007

*Western Mountains Ecoregion Lakes in the EPA National Lakes Assessment (NLA) for 2007*

---

## Description

A data frame containing attributes for lakes sampled by the EPA National Lakes Assessment (NLA) in the Western Mountains Ecoregion during 2007.

## Usage

```
data(NLA_2007)
```

## Format

A data frame with 236 rows and the following 15 columns:

**siteID** site ID value.

**xcoord** Albers projection x-coordinate.

**ycoord** Albers projection y-coordinate.

**wgt** survey design weight.

**Lake\_Origin** lake origin category.

**Chla** chlorophyll-a concentration.

**OE5** value of the index of macroinvertebrate taxa loss.

**PTL** total phosphorus concentration.

**NTL** total nitrogen concentration.

**Turbidity** turbidity value.

**Chla\_cond** condition class category ("Good", "Fair", or "Poor") of the chlorophyll-a value.

**OE5\_cond** condition class category of the macroinvertebrate taxa loss index.

**PTL\_cond** condition class category of the total phosphorus value.

**NTL\_cond** condition class category of the total nitrogen value.

**Turbidity\_cond** condition class category of the turbidity value.

## Examples

```
data(NLA_2007)
```

```
head(NLA_2007)
```



---

read.dbf	<i>Read the Attribute (dbf) File of an ESRI Shapefile</i>
----------	---

---

### Description

This function reads either a single dbf file or multiple dbf files. For multiple dbf files, all of the dbf files must have the same variable names.

### Usage

```
read.dbf(filename=NULL)
```

### Arguments

filename	name of the dbf file without any extension. If filename equals a dbf file name, then that dbf file is read. If filename equals NULL, then all of the dbf files in the working directory are read. The default is NULL.
----------	--

### Details

Function `summary()`, i.e., `summary.SurveyFrame()`, can be used to summarize the the frame for a survey design.

### Value

A data frame composed of either the contents of the single dbf file, when filename is provided, or the contents of the dbf file(s) in the working directory, when filename is NULL. The data frame is assigned class "SurveyFrame".

### Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>

### References

ESRI Shapefile Technical Description: <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>

### See Also

[read.shape.framesum](#)

### Examples

```
## Not run:  
read.shape("my.dbffile")  
  
## End(Not run)
```

---

`read.sas`*Read SAS Datasets or a SAS XPORT (Transport) File*

---

**Description**

This function reads either a SAS dataset or a SAS XPORT (transport) file and creates a data frame.

**Usage**

```
read.sas(filename, libname=NULL, xport=FALSE, sascmd="/Program Files/SAS/SAS 9.1/sas.exe")
```

**Arguments**

<code>filename</code>	if <code>xport</code> equals <code>TRUE</code> , a character string giving the full path to the SAS XPORT file, which must include the file extension. If <code>xport</code> equals <code>FALSE</code> , either a character string giving the the name of a dataset in the SAS library or a vector of character strings giving the names of datasets in the SAS library, where the dataset names cannot exceed eight characters in length and do not include the file extension.
<code>libname</code>	a character string defining the SAS library, which is usually a directory reference. If <code>xport</code> equals <code>FALSE</code> and the dataset(s) named in argument <code>filename</code> do not reside in the working directory, then this argument is required. The default value is <code>NULL</code> .
<code>xport</code>	a logical value indicating whether the input file is a SAS XPORT file. The default value is <code>FALSE</code> .
<code>sascmd</code>	a character string giving the full path to SAS executable. This argument is required only when <code>xport</code> equals <code>FALSE</code> . The default value is <code>"C:/Program Files/SAS/SAS 9.1/sas.exe"</code> .

**Value**

Either a single data frame or a list of data frames.

**Author(s)**

Tom Kincaid <Kincaid.Tom@epa.gov>

**Examples**

```
## Not run:  
MySasFile <- read.sas("mysasfil", "C:/Documents and Settings/auser/My Documents/My Project")  
  
## End(Not run)
```

---

read.shape	<i>Read an ESRI Shapefile</i>
------------	-------------------------------

---

### Description

This function reads either a single shapefile or multiple shapefiles. For multiple shapefiles, all of the shapefiles must be the same type, i.e., point, polyline, or polygon.

### Usage

```
read.shape(filename=NULL)
```

### Arguments

filename	name of the shapefile without any extension. If filename equals a shapefile name, than that shapefile is read. If filename equals NULL, then all of the shapefiles in the working directory are read. The default is NULL.
----------	--

### Value

An sp package object containing information in the shapefile. The object is assigned class "SpatialPointsDataFrame", "SpatialLinesDataFrame", or "SpatialPolygonsDataFrame" corresponding to the shapefile type, i.e., point, polyline, or polygon, respectively. For further information regarding the output object, see documentation for the sp package.

### Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>

### References

ESRI Shapefile Technical Description: <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>

### See Also

[read.dbf](#)

### Examples

```
## Not run:  
read.shape("my.shapefile")  
  
## End(Not run)
```

---

relrisk.analysis      *Relative Risk Analysis for Probability Survey Data*

---

### Description

This function organizes input and output for relative risk analysis of categorical data generated by a probability survey.

### Usage

```
relrisk.analysis(sites=NULL, subpop=NULL, design, data.rr, response.var,
  stressor.var, response.levels=rep(list(c("Poor","Good")),
  length(response.var)), stressor.levels=rep(list(c("Poor","Good")),
  length(stressor.var)), popcorrect=FALSE, pcfsiz=NULL, N.cluster=NULL,
  stage1size=NULL, sizeweight=FALSE, vartype="Local", conf=95)
```

### Arguments

sites	a data frame consisting of two variables: the first variable is site IDs, and the second variable is a logical vector indicating which sites to use in the analysis. The default is NULL.
subpop	a data frame describing sets of populations and subpopulations for which estimates will be calculated. The first variable is site IDs. Each subsequent variable identifies a Type of population, where the variable name is used to identify Type. A Type variable identifies each site with one of the subpopulations of that Type. The default is NULL.
design	a data frame consisting of design variables. Variables should be named as follows: siteID = site IDs wgt = final adjusted weights, which are either the weights for a single-stage sample or the stage two weights for a two-stage sample xcoord = x-coordinates for location, which are either the x-coordinates for a single-stage sample or the stage two x-coordinates for a two-stage sample ycoord = y-coordinates for location, which are either the y-coordinates for a single-stage sample or the stage two y-coordinates for a two-stage sample stratum = the stratum codes cluster = the stage one sampling unit (primary sampling unit or cluster) codes wgt1 = final adjusted stage one weights xcoord1 = the stage one x-coordinates for location ycoord1 = the stage one y-coordinates for location support = support values - the value one (1) for a site from a finite resource or the measure of the sampling unit associated with a site from an extensive resource, which is required for calculation of finite and continuous population correction factors swgt = size-weights, which is the stage two size-weight for a two- stage sample swgt1 = stage one size-weights

<code>data.rr</code>	data frame of categorical response and stressor variables, where each variable consists of two categories. If response or stressor variables include more than two categories, occurrences of those categories must be removed or replaced with missing values. The first column of this argument is site IDs. Subsequent columns are response and stressor variables. Missing data (NA) is allowed.
<code>response.var</code>	character vector providing names of columns in argument <code>data.rr</code> that contain a response variable, where names may be repeated. Each name in this argument is matched with the corresponding value in the <code>stressor.var</code> argument.
<code>stressor.var</code>	character vector providing names of columns in argument <code>data.rr</code> that contain a stressor variable, where names may be repeated. Each name in this argument is matched with the corresponding value in the <code>response.var</code> argument. This argument must be the same length as argument <code>response.var</code> .
<code>response.levels</code>	list providing the category values (levels) for each element in the <code>response.var</code> argument. This argument must be the same length as argument <code>response.var</code> . The first level for each element in the list is used for calculating the numerator and the denominator of the relative risk estimate. The default is a list containing the values "Poor" and "Good" for the first and second levels, respectively, of each element in the <code>response.var</code> argument.
<code>stressor.levels</code>	list providing the category values (levels) for each element in the <code>stressor.var</code> argument. This argument must be the same length as argument <code>response.var</code> . The first level for each element in the list is used for calculating the numerator of the relative risk estimate, and the second level for each element in the list is used for calculating the denominator of the estimate. The default is a list containing the values "Poor" and "Good" for the first and second levels, respectively, of each element in the <code>stressor.var</code> argument.
<code>popcorrect</code>	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation, where TRUE = use the correction factors and FALSE = do not use the correction factors. The default is FALSE.
<code>pcfsize</code>	size of the resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample this argument must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
<code>N.cluster</code>	the number of stage one sampling units in the resource, which is required for calculation of finite and continuous population correction factors for a two-stage sample. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
<code>stage1size</code>	size of the stage one sampling units of a two-stage sample, which is required for calculation of finite and continuous population correction factors for a two-stage sample and must have the names attribute set to identify the stage one sampling unit codes. For a stratified sample, the names attribute must be set to identify both stratum codes and stage one sampling unit codes using a convention where the two codes are separated by the & symbol, e.g., "Stratum 1&Cluster 1". The default is NULL.

sizeweight	a logical value that indicates whether size-weights should be used in the analysis, where TRUE = use the size-weights and FALSE = do not use the size-weights. The default is FALSE.
vartype	the choice of variance estimator, where "Local" = local mean estimator and "SRS" = SRS estimator. The default is "Local".
conf	the confidence level. The default is 95%.

### Value

Value is a data frame of relative risk estimates for all combinations of population Types, subpopulations within Types, and response variables. Standard error and confidence interval estimates also are provided.

### Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>

### References

Sarndal, C.E., B. Swensson, and J. Wretman. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

### See Also

[relrisk.est](#)

### Examples

```
mysiteID <- paste("Site", 1:100, sep="")
mysites <- data.frame(siteID=mysiteID, Active=rep(TRUE, 100))
mysubpop <- data.frame(siteID=mysiteID, All.Sites=rep("All Sites", 100),
  Resource.Class=rep(c("Agr", "Forest"), c(55,45)))
mydesign <- data.frame(siteID=mysiteID, wgt=runif(100, 10, 100),
  xcoord=runif(100), ycoord=runif(100), stratum=rep(c("Stratum1",
  "Stratum2"), 50))
mydata.rr <- data.frame(siteID=mysiteID, RespVar1=sample(c("Poor", "Good"),
  100, replace=TRUE), RespVar2=sample(c("Poor", "Good"), 100, replace=TRUE),
  StressVar=sample(c("Poor", "Good"), 100, replace=TRUE), wgt=runif(100, 10,
  100))
relrisk.analysis(sites=mysites, subpop=mysubpop, design=mydesign,
  data.rr=mydata.rr, response.var=c("RespVar1", "RespVar2"),
  stressor.var=rep("StressVar", 2))
```

relrisk.est

*Compute the Relative Risk Estimate***Description**

This function calculates the relative risk estimate for a 2x2 table of cell counts defined by a categorical response variable and a categorical explanatory (stressor) variable for an unequal probability design. Relative risk is the ratio of two probabilities: the numerator is the probability that the first level of the response variable is observed given occurrence of the first level of the stressor variable, and the denominator is the probability that the first level of the response variable is observed given occurrence of the second level of the stressor variable. The standard error of the base e log of the relative risk estimate and confidence limits for the estimate also are calculated.

**Usage**

```
relrisk.est(response, stressor, response.levels=c("Poor", "Good"),
  stressor.levels=c("Poor", "Good"), wgt, xcoord=NULL, ycoord=NULL,
  stratum=NULL, cluster=NULL, wgt1=NULL, xcoord1=NULL, ycoord1=NULL,
  popcorrect=FALSE, pcfsz=pcfsz, N.cluster=NULL, stage1size=NULL,
  support=NULL, sizeweight=FALSE, swgt=NULL, swgt1=NULL, vartype="Local",
  conf=95, check.ind=TRUE, warn.ind=NULL, warn.df=NULL, warn.vec=NULL)
```

**Arguments**

response	the categorical response variable values.
stressor	the categorical explanatory (stressor) variable values.
response.levels	category values (levels) for the categorical response variable, where the first level is used for calculating the numerator and the denominator of the relative risk estimate. If response.levels is not supplied, then values "Poor" and "Good" are used for the first level and second level of the response variable, respectively. The default is c("Poor", "Good").
stressor.levels	category values (levels) for the categorical stressor variable, where the first level is used for calculating the numerator of the relative risk estimate and the second level is used for calculating the denominator of the estimate. If stressor.levels is not supplied, then values "Poor" and "Good" are used for the first level and second level of the stressor variable, respectively. The default is c("Poor", "Good").
wgt	the final adjusted weight (inverse of the sample inclusion probability) for each site, which is either the weight for a single-stage sample or the stage two weight for a two-stage sample.
xcoord	x-coordinate for location for each site, which is either the x-coordinate for a single-stage sample or the stage two x-coordinate for a two-stage sample. The default is NULL.

ycoord	y-coordinate for location for each site, which is either the y-coordinate for a single-stage sample or the stage two y-coordinate for a two-stage sample. The default is NULL.
stratum	the stratum for each site. The default is NULL.
cluster	the stage one sampling unit (primary sampling unit or cluster) code for each site. The default is NULL.
wgt1	the final adjusted stage one weight for each site. The default is NULL.
xcoord1	the stage one x-coordinate for location for each site. The default is NULL.
ycoord1	the stage one y-coordinate for location for each site. The default is NULL.
popcorrect	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation, where TRUE = use the correction factors and FALSE = do not use the correction factors. The default is FALSE.
pcfsize	size of the resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample this argument must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
N.cluster	the number of stage one sampling units in the resource, which is required for calculation of finite and continuous population correction factors for a two-stage sample. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
stage1size	size of the stage one sampling units of a two-stage sample, which is required for calculation of finite and continuous population correction factors for a two-stage sample and must have the names attribute set to identify the stage one sampling unit codes. For a stratified sample, the names attribute must be set to identify both stratum codes and stage one sampling unit codes using a convention where the two codes are separated by the & symbol, e.g., "Stratum 1&Cluster 1". The default is NULL.
support	the support value for each site - the value one (1) for a site from a finite resource or the measure of the sampling unit associated with a site from an extensive resource, which is required for calculation of finite and continuous population correction factors. The default is NULL.
sizeweight	a logical value that indicates whether size-weights should be used in the analysis, where TRUE = use the size-weights and FALSE = do not use the size-weights. The default is FALSE.
swgt	the size-weight for each site, which is the stage two size-weight for a two-stage sample. The default is NULL.
swgt1	the stage one size-weight for each site. The default is NULL.
vartype	the choice of variance estimator, where "Local" = local mean estimator and "SRS" = SRS estimator. The default is "Local".
conf	the confidence level. The default is 95%.
check.ind	a logical value that indicates whether compatibility checking of the input values is conducted, where TRUE = conduct compatibility checking and FALSE = do not conduct compatibility checking. The default is TRUE.



warn.ind	a logical value that indicates whether warning messages were generated, where TRUE = warning messages were generated and FALSE = warning messages were not generated. The default is NULL.
warn.df	a data frame for storing warning messages. The default is NULL.
warn.vec	a vector that contains names of the population type, the subpopulation, and an indicator. The default is NULL.

### Details

The relative risk estimate is computed using the ratio of a numerator probability to a denominator probability, which are estimated using cell and marginal totals from a 2x2 table of cell counts defined by a categorical response variable and a categorical stressor variable (Van Sickle and Paulsen, 2008). An estimate of the numerator probability is provided by the ratio of the cell total defined by the first level of the response variable and the first level of the stressor variable to the marginal total for the first level of the stressor variable. An estimate of the denominator probability is provided by the ratio of the cell total defined by the first level of response variable and the second level of the stressor variable to the marginal total for the second level of the stressor variable. Cell and marginal totals are estimated using the Horvitz-Thompson estimator. The standard error of the base e log of the relative risk estimate is calculated using a first-order Taylor series linearization (Sarndal et al., 1992).

### Value

If the function was called by the `relrisk.analysis` function, then `value` is a list containing the following components:

- `Results` - a list containing estimates, confidence bounds, and associated values
- `warn.ind` - a logical value indicating whether warning messages were generated
- `warn.df` - a data frame containing warning messages

If the function was called directly, then `value` is the `Results` list, which contains the following components:

- `RelRisk` - the relative risk estimate
- `RRnum` - numerator ("elevated" risk) of the relative risk estimate
- `RRdenom` - denominator ("baseline" risk) of the relative risk estimate
- `RRlog.se` - standard error for the log of the relative risk estimate
- `ConfLimits` - confidence limits for the relative risk estimate
- `WeightTotal` - sum of the final adjusted weights
- `CellCounts` - cell and margin counts for the 2x2 table
- `CellProportions` - estimated cell proportions for the 2x2 table

### Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>  
Tony Olsen <Olsen.Tony@epa.gov>  
John Vansickle <Vansickle.John@epa.gov>

## References

Van Sickle, J. and S. G. Paulsen. (2008). Assessing the attributable risks, relative risks, and regional extent of aquatic stressors. *Journal of the North American Benthological Society* **27**, 920-931.

Sarndal, C.E., B. Swensson, and J. Wretman. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

## Examples

```
response <- sample(c("Poor", "Good"), 100, replace=TRUE)
stressor <- sample(c("Poor", "Good"), 100, replace=TRUE)
wgt <- runif(100, 10, 100)
relrisk.est(response, stressor, wgt=wgt, vartype="SRS")

xcoord <- runif(100)
ycoord <- runif(100)
relrisk.est(response, stressor, wgt=wgt, xcoord=xcoord, ycoord=ycoord)
```

---

SC\_estuaries

*Estuaries in South Carolina*

---

## Description

A data frame containing attributes for estuaries in South Carolina.

## Usage

```
data(SC_estuaries)
```

## Format

A data frame with 135 rows and the following 10 columns:

**siteID** site ID value.

**xcoord** Albers projection x-coordinate.

**ycoord** Albers projection y-coordinate.

**wgt** survey design weight.

**Stratum** stratum code.

**Status** site evaluation status code.

**IBI\_score** IBI (index of biotic integrity) score.

**IBI\_status** status category of the IBI score.

**WQ\_score** WQ (water quality) score.

**WQ\_status** status category of the WQ score.

## Examples

```
data(SC_estuaries)
head(SC_estuaries)
```

---

`sp2shape`*Convert an sp Package Object to an ESRI Shapefile*

---

### Description

This function creates an ESRI shapefile from an sp package object. The type of shapefile, i.e., point, polyline, or polygon, is determined by the class of the sp object, which must be either "SpatialPointsDataFrame", "SpatialLinesDataFrame", or "SpatialPolygonsDataFrame".

### Usage

```
sp2shape(sp.obj, shpfilename="tempfile", prjfilename=NULL)
```

### Arguments

<code>sp.obj</code>	the sp package object.
<code>shpfilename</code>	name (without any extension) of the output shapefile. The default is "tempfile".
<code>prjfilename</code>	name (without any extension) of the projection file for the output shapefile. The default is NULL.

### Value

An ESRI shapefile of type point, polyline, or polygon.

### Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>

### References

ESRI Shapefile Technical Description: <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>

### Examples

```
## Not run:  
sp2shape(my.sp.object, "my.shapefile")  
  
## End(Not run)
```

---

SpatialDesign      *Create Objects of Class SpatialDesign*

---

### Description

Create objects of class [SpatialDesign](#) from survey design lists and objects of class [SpatialPointsDataFrame](#)-class.

### Usage

```
SpatialDesign(design, sp_obj)
```

### Arguments

design	object of class list containing specifications for the survey design.
sp_obj	object of class <a href="#">SpatialPointsDataFrame</a> containing spatial attributes that have spatial point locations.

### Value

Returns an object of class [SpatialDesign](#).

### See Also

[SpatialDesign-class](#)

---

SpatialDesign-class      *Class "SpatialDesign"*

---

### Description

Class for spatial attributes that have spatial point locations and survey design information

### Usage

```
## S4 method for signature 'SpatialDesign'
summary(object, ..., auxvar = NULL, spframe = NULL,
         tess_ind = TRUE, sbc_ind = FALSE, nrows = 5, dx dy = TRUE)
## S3 method for class 'SpatialDesign'
summary(object, ..., auxvar = NULL, spframe = NULL,
         tess_ind = TRUE, sbc_ind = FALSE, nrows = 5, dx dy = TRUE)
## S4 method for signature 'SpatialDesign,missing'
plot(x, y, ..., spframe = NULL,
     stratum = NULL, mdcaty = NULL, auxvar = NULL, pdf file = NULL, width = 8,
```

```

    height = 10)
  ## S3 method for class 'SpatialDesign'
  plot(x, y, ..., spframe = NULL, stratum = NULL,
       mdcaty = NULL, auxvar = NULL, pdffile = NULL, width = 8, height = 10)

```

### Arguments

object, x	SpatialDesign object
y	missing - this argument is not used
...	arguments passed through
auxvar	a vector containing the names of columns in the data slot of the SpatialDesign object that identify auxiliary variables to be used to summarize the survey design or create plots of the survey design. The default is NULL.
spframe	an sp package object of class SpatialPointsDataFrame, SpatialLinesDataFrame, or SpatialPolygonsDataFrame that contains the survey design frame. The default is NULL.
tess_ind	a logical variable indicating whether spatial balance metrics are calculated using proportions obtained from the intersection of Dirichlet tessellation polygons for the sample points with the frame object. TRUE means calculate the metrics. FALSE means do not calculate the metrics. The default is TRUE
sbc_ind	a logical variable indicating whether spatial balance metrics are calculated using proportions obtained from a rectangular grid superimposed on the sample points and the frame. TRUE means calculate the metrics. FALSE means do not calculate the metrics. The default is FALSE.
nrows	number of rows (and columns) for the grid of cells. The default is 5.
dxdy	indicator for equal x-coordinate and y-coordinate grid cell increments, where TRUE means the increments are equal and FALSE means the increments are not equal. The default is TRUE.
stratum	name of the column from the data slot of the spframe object that identifies stratum membership for each element in the frame. If stratum equals NULL, the design is unstratified, and a column named "stratum" (with all its elements equal to the stratum name specified in design) is added to the data slot of the spframe object. The default is NULL.
mdcaty	name of the column from the data slot of the spframe object that identifies the unequal probability category for each element in the frame. The default is NULL.
pdffile	a character variable containing the name of the pdf file to which output is written. If a value is not provided, output is written to the graphics window. The default is NULL.
width	width of the graphic region in inches. The default is 8.
height	height of the graphic region in inches. The default is 10.

### Slots

**design:** Object of class list containing specifications for the survey design.

**data:** Object of class data.frame containing the attribute data.

coords.nrs: Numeric object that records the column positions in data from which the coordinates were obtained.

coords: Object of class matrix containing the coordinates matrix, where points are rows in the matrix.

bbox: Object of class matrix containing the bounding box.

proj4string: Object of class CRS containing the projection string.

### Extends

Class "SpatialPointsDataFrame", directly. Class "SpatialPoints", by class "SpatialPointsDataFrame". Class "Spatial", by class "SpatialPoints".

### See Also

[SpatialDesign](#)

---

spbalance

*Calculate Spatial Balance Metrics for a Survey Design*

---

### Description

This function calculates spatial balance metrics for a survey design. Two options for calculation of spatial balance metrics are available: (1) use proportions obtained from the intersection of Dirichlet tessellation polygons for the sample points with the frame object and (2) use proportions obtained from a rectangular grid superimposed on the sample points and the frame object. In both cases the proportions are used to calculate the spatial balance metrics. Two metrics are calculated: (1) the Pielou evenness measure and (2) the chi-square statistic.

### Usage

```
spbalance(spsample, spframe = NULL, tess_ind = TRUE, sbc_ind = FALSE, nrows = 5,
          dxdy = TRUE)
```

### Arguments

spsample	an object of class SpatialDesign produced by either the grts or irs functions that contains survey design information and additional attribute (auxiliary) variables.
spframe	an sp package object of class SpatialPointsDataFrame, SpatialLinesDataFrame, or SpatialPolygonsDataFrame that contains the survey design frame. The default is NULL.
tess_ind	a logical variable indicating whether spatial balance metrics are calculated using proportions obtained from the intersection of Dirichlet tessellation polygons for the sample points with the frame object. TRUE means calculate the metrics. FALSE means do not calculate the metrics. The default is TRUE

sbc_ind	a logical variable indicating whether spatial balance metrics are calculated using proportions obtained from a rectangular grid superimposed on the sample points and the frame. TRUE means calculate the metrics. FALSE means do not calculate the metrics. The default is FALSE.
nrows	number of rows (and columns) for the grid of cells. The default is 5.
dxdy	indicator for equal x-coordinate and y-coordinate grid cell increments, where TRUE means the increments are equal and FALSE means the increments are not equal. The default is TRUE.

### Value

A list containing the following components:

tess	results for spatial balance metrics using tessellation polygons.
sbc	results for spatial balance metrics using a rectangular grid.

If either the tess\_ind or sbc\_ind arguments are set to FALSE, the corresponding component in the list is set to NULL. Otherwise, each component of the list is a lists that contains the following components:

J_subp	Pielou evenness measure.
chi_sq	chi-square statistic.
extent	frame extent for each Dirichlet tessellation polygon or rectangular grid cell.
prop	frame proportion for each Dirichlet tessellation polygon or rectangular grid cell.

### Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>

### References

Olsen, A. R., T. M. Kincaid, and Q. Payton (2012). Spatially balanced survey designs for natural resources. In R. A. Gitzen, J. J. Millsbaugh, A. B. Cooper, and D. S. Licht (Eds.), *Design and Analysis of Long-term Ecological Monitoring Studies*, pp. 126-150. Cambridge University Press.

### Examples

```
## Not run:
design <- list(Stratum1=list(panel=c(PanelOne=50), seltype="Equal", over=10),
  Stratum2=list(panel=c(PanelOne=50, PanelTwo=50), seltype="Unequal",
  caty.n=c(CatyOne=25, CatyTwo=25, CatyThree=25, CatyFour=25), over=75))
frame <- read.shp("shapefile")
samp <- grts(design=design, DesignID="Test.Site", type.frame="area",
  src.frame="shapefile", in.shape="shapefile", att.frame=frame@data,
  stratum="stratum", mdcaty="mdcaty", shapefile=TRUE,
  shapefilename="sample")
spbalance(samp, frame, sbc_ind = TRUE)

## End(Not run)
```

---

spsurvey.analysis      *Create an Object of Class spsurvey.analysis*

---

## Description

This function creates an object of class `spsurvey.analysis` that contains all of the information necessary to use the analysis functions in the `spsurvey` library.

## Usage

```
spsurvey.analysis(sites=NULL, subpop=NULL, design=NULL, data.cat=NULL,
  data.cont=NULL, siteID=NULL, wgt=NULL, sigma=NULL, var.sigma=NULL,
  xcoord=NULL, ycoord=NULL, stratum=NULL, cluster=NULL, wgt1=NULL, xcoord1=NULL,
  ycoord1=NULL, popsize=NULL, popcorrect=FALSE, pcfsite=NULL, N.cluster=NULL,
  stage1size=NULL, support=NULL, sizeweight=FALSE, swgt=NULL, swgt1=NULL,
  vartype="Local", conf=95, pctval=c(5,10,25,50,75,90,95))
```

## Arguments

- |        |  |
|--------|--|
| sites  | a data frame consisting of two variables: the first variable is site IDs and the second variable is a logical vector indicating which sites to use in the analysis. If this data frame is not provided, then the data frame will be created, where (1) site IDs are obtained either from the <code>design</code> argument, the <code>siteID</code> argument, or both (when <code>siteID</code> is a formula); and (2) a variable named <code>use.sites</code> that contains the value <code>TRUE</code> for all sites is created. The default is <code>NULL</code> .   |
| subpop | a data frame describing sets of populations and subpopulations for which estimates will be calculated. The first variable is site IDs and each subsequent variable identifies a Type of population, where the variable name is used to identify Type. A Type variable identifies each site with one of the subpopulations of that Type. If this data frame is not provided, then the data frame will be created, where (1) site IDs are obtained either from the <code>design</code> argument, the <code>siteID</code> argument, or both (when <code>siteID</code> is a formula); and (2) a single Type variable named <code>all.sites</code> that contains the value "All Sites" for all sites is created. The default is <code>NULL</code> . |
| design | a data frame consisting of design variables. If variable names are provided as formulas in the corresponding arguments, then the formulas are interpreted using this data frame. If this data frame is not provided, then the data frame will be created from inputs to the design variables in the argument list. The default is <code>NULL</code> . If variable names are not provided as formulas, then variables should be named as follows:<br><code>siteID</code> = site IDs<br><code>wgt</code> = final adjusted weights<br><code>xcoord</code> = x-coordinates for location<br><code>ycoord</code> = y-coordinates for location<br><code>stratum</code> = stratum codes<br><code>cluster</code> = stage one sampling unit codes        |



	<p>wgt1 = final adjusted stage one weights  xcoord1 = stage one x-coordinates for location  ycoord1 = stage one y-coordinates for location  support = support values  swgt = size-weights  swgt1 = stage one size-weights</p>
data.cat	a data frame of categorical response variables. The first variable is site IDs. Subsequent variables are response variables. Missing data (NA) is allowed. The default is NULL.
data.cont	a data frame of continuous response variables. The first variable is site IDs. Subsequent variables are response variables. Missing data (NA) is allowed. The default is NULL.
siteID	site IDs. This variable can be input directly or as a formula and must be supplied either as this argument or in the design data frame. The default is NULL.
wgt	the final adjusted weight (inverse of the sample inclusion probability) for each site, which is either the weight for a single-stage sample or the stage two weight for a two-stage sample. The default is NULL.
sigma	measurement error variance. This variable must be a vector containing a value for each response variable and must have the names attribute set to identify the response variable names. Missing data (NA) is allowed. The default is NULL.
var.sigma	variance of the measurement error variance. This variable must be a vector containing a value for each response variable and must have the names attribute set to identify the response variable names. Missing data (NA) is allowed. The default is NULL.
xcoord	x-coordinate for location for each site, which is either the x-coordinate for a single-stage sample or the stage two x-coordinate for a two-stage sample. The default is NULL.
ycoord	y-coordinate for location for each site, which is either the y-coordinate for a single-stage sample or the stage two y-coordinate for a two-stage sample. The default is NULL.
stratum	the stratum codes. This variable can be input directly or as a formula. The default is NULL.
cluster	the stage one sampling unit (primary sampling unit or cluster) codes. This variable can be input directly or as a formula. The default is NULL.
wgt1	the final adjusted stage one weights. This variable can be input directly or as a formula. The default is NULL.
xcoord1	the stage one x-coordinates for location. This variable can be input directly or as a formula. The default is NULL.
ycoord1	the stage one y-coordinates for location. This variable can be input directly or as a formula. The default is NULL.
popsize	known size of the resource, which is used to perform ratio adjustment to estimators expressed using measurement units for the resource. For a finite resource, this argument is either the total number of sampling units or the known sum of size-weights. For an extensive resource, this argument is the measure of the resource, i.e., either known total length for a linear resource or known total area

for an areal resource. The argument must be in the form of a list containing an element for each population Type in the subpop data frame, where NULL is a valid choice for a population Type. The list must be named using the column names for the population Types in subpop. If a population Type doesn't contain subpopulations, then each element of the list is either a single value for an unstratified sample or a vector containing a value for each stratum for a stratified sample, where elements of the vector are named using the stratum codes. If a population Type contains subpopulations, then each element of the list is a list containing an element for each subpopulation, where the list is named using the subpopulation names. The element for each subpopulation will be either a single value for an unstratified sample or a named vector of values for a stratified sample. The default is NULL.

Example popsize for a stratified sample:

```
popsize = list("Pop 1"=c("Stratum 1"=750,
"Stratum 2"=500,
"Stratum 3"=250),
"Pop 2"=list("SubPop 1"=c("Stratum 1"=350,
"Stratum 2"=250,
"Stratum 3"=150),
"SubPop 2"=c("Stratum 1"=250,
"Stratum 2"=150,
"Stratum 3"=100),
"SubPop 3"=c("Stratum 1"=150,
"Stratum 2"=150,
"Stratum 3"=75)),
"Pop 3"=NULL)
```

Example popsize for an unstratified sample:

```
popsize = list("Pop 1"=1500,
"Pop 2"=list("SubPop 1"=750,
"SubPop 2"=500,
"SubPop 3"=375),
"Pop 3"=NULL)
```

popcorrect	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation, where TRUE = use the correction factors and FALSE = do not use the correction factors. The default is FALSE.
pcfsize	size of the resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample this argument must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
N.cluster	the number of stage one sampling units in the resource, which is required for calculation of finite and continuous population correction factors for a two-stage sample. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.

stage1size	size of the stage one sampling units of a two-stage sample, which is required for calculation of finite and continuous population correction factors for a two-stage sample and must have the names attribute set to identify the stage one sampling unit codes. For a stratified sample, the names attribute must be set to identify both stratum codes and stage one sampling unit codes using a convention where the two codes are separated by the & symbol, e.g., "Stratum 1&Cluster 1". The default is NULL.
support	the support value for each site - the value one (1) for a site from a finite resource or the measure of the sampling unit associated with a site from an extensive resource, which is required for calculation of finite and continuous population correction factors. This variable can be input directly or as a formula. The default is NULL.
sizeweight	a logical value that indicates whether size-weights should be used in the analysis, where TRUE = use the size-weights and FALSE = do not use the size-weights. The default is FALSE.
swgt	the size-weight for each site, which is the stage two size-weight for a two-stage sample. This variable can be input directly or as a formula. The default is NULL.
swgt1	the stage one size-weight for each site. This variable can be input directly or as a formula. The default is NULL.
vartype	the choice of variance estimator, where "Local" = local mean estimator and "SRS" = SRS estimator. The default is "Local".
conf	the confidence level. The default is 95%.
pctval	the set of values at which percentiles are estimated. The default set is: {5, 10, 25, 50, 75, 90, 95}.

### Value

Value is a list of class `spsurvey.analysis`. Only those sites indicated by the logical variable in the sites data frame are retained in the output. The sites, subpop, and design data frames will always exist in the output. At least one of the `data.cat` and `data.cont` data frames will exist. Depending upon values of the input variables, other elements in the output may be NULL. The list is composed of the following components:

- `sites` - the sites data frame
- `subpop` - the subpop data frame
- `design` - the design data frame
- `data.cat` - the `data.cat` data frame
- `data.cont` - the `data.cont` data frame
- `sigma` - measurement error variance
- `var.sigma` - variance of the estimated measurement error variance
- `stratum.ind` - a logical value that indicates whether the sample is stratified, where TRUE = a stratified sample and FALSE = not a stratified sample
- `cluster.ind` - a logical value that indicates whether the sample is a two-stage sample, where TRUE = a two-stage sample and FALSE = not a two-stage sample

- `popsize` - the known size of the resource
- `pcfactor.ind` - a logical value that indicates whether the population correction factor is used during variance estimation, where TRUE = use the population correction factor and FALSE = do not use the factor
- `pcfsize` - size of the resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample
- `N.cluster` - the number of stage one sampling units in the resource
- `stage1size` - the known size of the stage one sampling units
- `swgt.ind` - a logical value that indicates whether the sample is a size-weighted sample, where TRUE = a size-weighted sample and FALSE = not a size-weighted sample
- `vartype` - the choice of variance estimator, where "Local" = local mean estimator and "SRS" = SRS estimator
- `conf` - the confidence level
- `pctval` - the set of values at which percentiles are estimated, where the default set is: 5, 25, 50, 75, 95

### Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>

### References

Diaz-Ramos, S., D.L. Stevens, Jr., and A.R. Olsen. (1996). *EMAP Statistical Methods Manual*. EPA/620/R-96/XXX. Corvallis, OR: U.S. Environmental Protection Agency, Office of Research and Development, National Health Effects and Environmental Research Laboratory, Western Ecology Division.

### See Also

[cat.analysis](#), [cont.analysis](#)

### Examples

```
# Categorical variable example:
mysiteID <- paste("Site", 1:100, sep="")
mysites <- data.frame(siteID=mysiteID, Active=rep(TRUE, 100))
mysubpop <- data.frame(siteID=mysiteID, All.Sites=rep("All Sites", 100),
  Resource.Class=rep(c("Good", "Poor"), c(55,45)))
mydesign <- data.frame(siteID=mysiteID, wgt=runif(100, 10,
  100), xcoord=runif(100), ycoord=runif(100), stratum= rep(c("Stratum1",
  "Stratum2"), 50))
mydata.cat <- data.frame(siteID=mysiteID, CatVar= rep(c("north", "south",
  "east", "west"), 25))
mypopsize <- list(All.Sites=c(Stratum1=3500, Stratum2=2000),
  Resource.Class=list(Good=c(Stratum1=2500, Stratum2=1500),
  Poor=c(Stratum1=1000, Stratum2=500)))
spsurvey.analysis(sites=mysites, subpop=mysubpop, design=mydesign,
  data.cat=mydata.cat, popsize=mypopsize)
```

```
# Continuous variable example - including deconvolution estimates:
mydesign <- data.frame(ID=mysiteID, wgt=runif(100, 10, 100),
  xcoord=runif(100), ycoord=runif(100), stratum=rep(c("Stratum1",
  "Stratum2"), 50))
ContVar <- rnorm(100, 10, 1)
mydata.cont <- data.frame(siteID=mysiteID, ContVar=ContVar,
  ContVar.1=ContVar + rnorm(100, 0, sqrt(0.25)),
  ContVar.2=ContVar + rnorm(100, 0, sqrt(0.50)))
mysigma <- c(ContVar=NA, ContVar.1=0.25, ContVar.2=0.50)
spsurvey.analysis(sites=mysites, subpop=mysubpop, design=mydesign,
  data.cont=mydata.cont, siteID=~ID, sigma=mysigma,
  popsize=mypopsize)
```

total.est

*Population Total, Mean, Variance, and Standard Deviation Estimates***Description**

This function calculates estimates of the population total, mean, variance, and standard deviation of a response variable, where the response variable may be defined for either a finite or an extensive resource. In addition the standard error of the population estimates and confidence bounds are calculated.

**Usage**

```
total.est(z, wgt, x=NULL, y=NULL, stratum=NULL, cluster=NULL, wgt1=NULL,
  x1=NULL, y1=NULL, popsize=NULL, popcorrect=FALSE, pcfsz=NULL,
  N.cluster=NULL, stage1size=NULL, support=NULL, sizeweight=FALSE, swgt=NULL,
  swgt1=NULL, vartype="Local", conf=95, check.ind=TRUE, warn.ind=NULL,
  warn.df=NULL, warn.vec=NULL)
```

**Arguments**

<code>z</code>	the response value for each site.
<code>wgt</code>	the final adjusted weight (inverse of the sample inclusion probability) for each site, which is either the weight for a single-stage sample or the stage two weight for a two-stage sample.
<code>x</code>	x-coordinate for location for each site, which is either the x-coordinate for a single-stage sample or the stage two x-coordinate for a two-stage sample. The default is NULL.
<code>y</code>	y-coordinate for location for each site, which is either the y-coordinate for a single-stage sample or the stage two y-coordinate for a two-stage sample. The default is NULL.
<code>stratum</code>	the stratum for each site. The default is NULL.
<code>cluster</code>	the stage one sampling unit (primary sampling unit or cluster) code for each site. The default is NULL.

wgt1	the final adjusted stage one weight for each site. The default is NULL.
x1	the stage one x-coordinate for location for each site. The default is NULL.
y1	the stage one y-coordinate for location for each site. The default is NULL.
popsiz	known size of the resource, which is used to calculate strata proportions for calculating estimates for a stratified sample. For a finite resource, this argument is either the total number of sampling units or the known sum of size-weights. For an extensive resource, this argument is the measure of the resource, i.e., either known total length for a linear resource or known total area for an areal resource. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
popcorrect	a logical value that indicates whether finite or continuous population correction factors should be employed during variance estimation, where TRUE = use the correction factors and FALSE = do not use the correction factors. The default is FALSE.
pcfsiz	size of the resource, which is required for calculation of finite and continuous population correction factors for a single-stage sample. For a stratified sample this argument must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
N.cluster	the number of stage one sampling units in the resource, which is required for calculation of finite and continuous population correction factors for a two-stage sample. For a stratified sample this variable must be a vector containing a value for each stratum and must have the names attribute set to identify the stratum codes. The default is NULL.
stage1siz	size of the stage one sampling units of a two-stage sample, which is required for calculation of finite and continuous population correction factors for a two-stage sample and must have the names attribute set to identify the stage one sampling unit codes. For a stratified sample, the names attribute must be set to identify both stratum codes and stage one sampling unit codes using a convention where the two codes are separated by the & symbol, e.g., "Stratum 1&Cluster 1". The default is NULL.
support	the support value for each site - the value one (1) for a site from a finite resource or the measure of the sampling unit associated with a site from an extensive resource, which is required for calculation of finite and continuous population correction factors. The default is NULL.
sizeweight	a logical value that indicates whether size-weights should be used in the analysis, where TRUE = use the size-weights and FALSE = do not use the size-weights. The default is FALSE.
swgt	the size-weight for each site, which is the stage two size-weight for a two-stage sample. The default is NULL.
swgt1	the stage one size-weight for each site. The default is NULL.
vartype	the choice of variance estimator, where "Local" = local mean estimator and "SRS" = SRS estimator. The default is "Local".
conf	the confidence level. The default is 95%.

check.ind	a logical value that indicates whether compatibility checking of the input values is conducted, where TRUE = conduct compatibility checking and FALSE = do not conduct compatibility checking. The default is TRUE.
warn.ind	a logical value that indicates whether warning messages were generated, where TRUE = warning messages were generated and FALSE = warning messages were not generated. The default is NULL.
warn.df	a data frame for storing warning messages. The default is NULL.
warn.vec	a vector that contains names of the population type, the subpopulation, and an indicator. The default is NULL.

### Details

The Horvitz-Thompson estimator is used to calculate the total, variance, and standard deviation estimates. The Horvitz-Thompson ratio estimator, i.e., the ratio of two Horvitz-Thompson estimators, is used to calculate the mean estimate. Variance estimates are calculated using either the local mean variance estimator or the simple random sampling (SRS) variance estimator. The choice of variance estimator is subject to user control. The local mean variance estimator requires the x-coordinate and the y-coordinate of each site. The SRS variance estimator uses the independent random sample approximation to calculate joint inclusion probabilities. Confidence bounds are calculated using a Normal distribution multiplier. The function can accommodate a stratified sample. For a stratified sample, separate estimates and standard errors are calculated for each stratum, which are used to produce estimates and standard errors for all strata combined. Strata that contain a single value are removed. For a stratified sample, when either the size of the resource or the sum of the size-weights of the resource is provided for each stratum, those values are used as stratum weights for calculating the estimates and standard errors for all strata combined. For a stratified sample when neither the size of the resource nor the sum of the size-weights of the resource is provided for each stratum, estimated values are used as stratum weights for calculating the estimates and standard errors for all strata combined. The function can accommodate single-stage and two-stage samples for both stratified and unstratified sampling designs. Finite population and continuous population correction factors can be utilized in variance estimation. The function checks for compatibility of input values and removes missing values.

### Value

If the function was called by the cont.analysis function, then value is a list containing the following components:

- Results - a data frame containing estimates and confidence bounds
- warn.ind - a logical value indicating whether warning messages were generated
- warn.df - a data frame containing warning messages

If the function was called directly, then value is a data frame containing estimates and confidence bounds.

### Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>

## References

Diaz-Ramos, S., D.L. Stevens, Jr., and A.R. Olsen. (1996). *EMAP Statistical Methods Manual*. EPA/620/R-96/XXX. Corvallis, OR: U.S. Environmental Protection Agency, Office of Research and Development, National Health Effects and Environmental Research Laboratory, Western Ecology Division.

## Examples

```
z <- rnorm(100, 10, 1)
wgt <- runif(100, 10, 100)
total.est(z, wgt, vartype="SRS")
```

```
x <- runif(100)
y <- runif(100)
total.est(z, wgt, x, y)
```

---

uniqueID

*Create Unique Site ID Values*

---

## Description

This function creates unique site IDs by appending a unique number to each occurrence of a site ID. It is intended for survey designs that have repeat visits to sites.

## Usage

```
uniqueID(siteID)
```

## Arguments

siteID            the vector of site IDs.

## Value

A vector of unique site ID values.

## Author(s)

Tom Kincaid <Kincaid.Tom@epa.gov>

## Examples

```
siteID <- paste("Site", c(1:10, 1, 3, 5), sep="")
uniqueID <- uniqueID(siteID)
```



---

UT_ecoregions	<i>Ecoregions in Utah</i>
---------------	---------------------------

---

**Description**

An object of class "SpatialPolygonsDataFrame" that was created from a shapefile of ecoregions in Utah.

**Usage**

```
data(UT_ecoregions)
```

**Format**

The format is: Formal class "SpatialPolygonsDataFrame" [package "sp"].  
The data slot contains 10 rows and the following 3 columns:

**Level3** Level 3 ecoregion code for the polygon.

**Level3\_Nam** Level 3 ecoregion name for the polygon.

**Area\_ha** area of the polygon in hectares.

**Examples**

```
data(UT_ecoregions)  
plot(UT_ecoregions)
```

---

write.dbf	<i>Write a Data Frame to the Attribute (dbf) File of an ESRI Shapefile</i>
-----------	--

---

**Description**

This function writes a data frame to the dbf file of an ESRI shapefile.

**Usage**

```
write.dbf(dframe, filename)
```

**Arguments**

dframe            a data frame to be written to the dbf file.

filename         name of the dbf file without any extension.

**Value**

The attribute (dbf) file of an ESRI shapefile.

**Author(s)**

Tom Kincaid <Kincaid.Tom@epa.gov>

**References**

ESRI Shapefile Technical Description: <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>

**See Also**

[read.dbf](#)

**Examples**

```
## Not run:
write.dbf(my.data.frame, "my.dbffile")

## End(Not run)
```

---

write.object

*Write an Object to a Plot*

---

**Description**

This function writes the contents of an object to a plot. The object may be either a data frame or a matrix. Values in the input data frame or matrix must be of class numeric, character, or factor.

**Usage**

```
write.object(obj, n.digits=2, r.names=TRUE, c.names=TRUE, r.cex=1, c.cex=1,
  miss="NA")
```

**Arguments**

obj	the object (either a data frame or a matrix).
n.digits	the number of digits after the decimal point for numeric values. The default is 2.
r.names	a logical value that indicates whether to print the row names, where TRUE = print the row names and FALSE = do not print the row names. The default is TRUE.
c.names	a logical value that indicates whether to print the column names, where TRUE = print the column names and FALSE = do not print the column names. The default is TRUE.
r.cex	character expansion parameter for the row labels. The default is 1.
c.cex	character expansion parameter for the column labels. The default is 1.
miss	the missing value code expressed as a character string. The default is "NA".

**Value**

The function returns NULL. Side effect of the function is to write contents of the input object to a plot.

**Author(s)**

Tom Kincaid <Kincaid.Tom@epa.gov>

**Examples**

```
z <- rnorm(100)
z.mean <- c(tapply(z, rep(1:4, rep(25,4)), mean), mean(z))
z.sd <- sqrt(c(tapply(z, rep(1:4, rep(25,4)), var), var(z)))
z.upper <- z.mean+1.96*z.sd
z.lower <- z.mean-1.96*z.sd
obj <- data.frame(rbind(z.mean, z.sd, z.upper, z.lower))
dimnames(obj) <- list(c("Mean Estimate", "Standard Deviation",
  "Lower 95% Conf. Bound", "Upper 95% Conf. Bound"), c(
  paste("Stratum", 1:4, sep=""), "AllStrata"))
write.object(obj, n.digits=3, r.cex=0.75)

obj <- data.frame(matrix(round(5 + runif(30), 1), nrow=6))
colnames(obj) <- c("United.States", "Russia", "Germany",
  "Japan", "France")
write.object(obj, n.digits=1, r.names=FALSE)
```

# Index

- \*Topic **IO**
  - read.sas, 74
  - read.shape, 75
  - sp2shape, 83
- \*Topic **aplot**
  - write.object, 98
- \*Topic **classes**
  - SpatialDesign-class, 84
- \*Topic **datasets**
  - decon\_data, 51
  - FL\_lakes, 53
  - IN\_streams, 64
  - Luck\_Ash\_streams, 70
  - NE\_lakes, 71
  - NLA\_2007, 72
  - SC\_estuaries, 82
  - UT\_ecoregions, 97
- \*Topic **distribution**
  - cdf.decon, 18
  - cdf.est, 22
  - cdf.plot, 26
  - cdf.test, 28
  - cont.analysis, 41
  - cont.cdfplot, 45
  - cont.cdfctest, 47
- \*Topic **manip**
  - SpatialDesign, 84
- \*Topic **misc**
  - adjwgt, 3
  - ash1.wgt, 5
  - marinus, 70
  - uniqueID, 96
- \*Topic **survey**
  - adjwgt, 3
  - albersgeod, 4
  - ash1.wgt, 5
  - attrisk.analysis, 6
  - attrisk.est, 8
  - cat.analysis, 12
  - category.est, 15
  - cdf.decon, 18
  - cdf.est, 22
  - cdf.plot, 26
  - cdf.test, 28
  - change.analysis, 33
  - change.est, 37
  - cont.analysis, 41
  - cont.cdfplot, 45
  - cont.cdfctest, 47
  - dsgnsum, 52
  - framesum, 54
  - geodalbers, 56
  - grts, 57
  - grtsarea, 60
  - grtslin, 61
  - grtspts, 62
  - irs, 64
  - irsarea, 67
  - irslin, 68
  - irspts, 69
  - read.dbf, 73
  - relrisk.analysis, 76
  - relrisk.est, 79
  - spbalance, 86
  - spsurvey.analysis, 88
  - total.est, 93
  - uniqueID, 96
  - write.dbf, 97
- \*Topic **survival**
  - attrisk.analysis, 6
  - attrisk.est, 8
  - relrisk.analysis, 76
  - relrisk.est, 79
- \*Topic **univar**
  - cat.analysis, 12
  - category.est, 15
  - change.analysis, 33
  - change.est, 37

- cont.analysis, 41
- total.est, 93
  
- adjwgt, 3
- albersgeod, 4, 59
- ash1.wgt, 5
- attrisk.analysis, 6
- attrisk.est, 8, 8
  
- cat.analysis, 12, 92
- category.est, 15, 15
- cdf.decon, 18
- cdf.est, 22, 45
- cdf.plot, 26, 47
- cdf.test, 28, 50
- change.analysis, 33
- change.est, 36, 37
- cont.analysis, 41, 92
- cont.cdfplot, 45
- cont.cdftest, 47
  
- decon\_data, 51
- dsgnsum, 52, 55, 59, 67
  
- FL\_lakes, 53
- framesum, 52, 54, 73
  
- geodalbers, 56
- grts, 52, 55, 57, 61–63
- grtsarea, 59, 60
- grtslin, 59, 61
- grtspts, 59, 62
  
- IN\_streams, 64
- interp.axis, 27
- interp.cdf, 27
- irs, 64, 68, 69
- irsarea, 67, 67
- irslin, 67, 68
- irspts, 67, 69
  
- Luck\_Ash\_streams, 70
  
- marinus, 70
  
- NE\_lakes, 71
- NLA\_2007, 72
  
- plot,SpatialDesign,missing-method  
(SpatialDesign-class), 84
  
- plot.SpatialDesign  
(SpatialDesign-class), 84
  
- read.dbf, 73, 75, 98
- read.sas, 74
- read.shape, 73, 75
- relrisk.analysis, 76
- relrisk.est, 78, 79
  
- SC\_estuaries, 82
- sp2shape, 83
- SpatialDesign, 84, 84, 86
- SpatialDesign-class, 84
- SpatialPointsDataFrame-class, 84
- spbalance, 86
- spsurvey.analysis, 88
- summary,SpatialDesign-method  
(SpatialDesign-class), 84
- summary.SpatialDesign  
(SpatialDesign-class), 84
  
- total.est, 45, 93
  
- uniqueID, 96
- UT\_ecoregions, 97
  
- write.dbf, 97
- write.object, 98