

# Package ‘svyPVpack’

February 20, 2015

**Type** Package

**Title** A package for complex surveys including plausible values

**Version** 0.1-1

**Date** 2014-03-06

**Author** Manuel Reif <Manuel.Reif@statistik.gv.at>, Jakob Peterbauer <Jakob.Peterbauer@statistik.gv.at>

**Maintainer** Manuel Reif <Manuel.Reif@statistik.gv.at>

**Description** This package deals with data which stem from survey designs including plausible values. This package has been created to handle data from Large Scale Assessments like PISA, PIAAC etc. which use complex survey designs to draw the sample and plausible values to report person related estimates. Various functions/statistics (mean, quantile, GLM etc.) are provided to handle this kind of data.

**License** GPL-3

**Depends** survey

**Suggests** testthat

**URL** <https://github.com/manuelreif/svyPVpack>

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2014-03-07 11:58:51

## R topics documented:

svyPVpack-package . . . . .	2
svy.exrep . . . . .	3
svyPVbenchmark . . . . .	3
svyPVcor . . . . .	5
svyPVeta . . . . .	6
svyPVglm . . . . .	7
svyPVlevel . . . . .	9
svyPVpm . . . . .	11
svyPVprob . . . . .	12
svyPVquantile . . . . .	14
svyPVttest . . . . .	15

---

svyPVpack-package      *Use survey designs including plausible values*

---

## Description

This package was built to properly estimate various statistics (mean, variance, glm ...) for survey designs which include plausible values. Plausible values are random draws from a posterior density and are typically used to make comparisons on the group level (this is typical for large scale assessment data as PISA, PIAAC etc.).

## Details

Package: svyPVpack  
Type: Package  
Version: 0.1-1  
Date: 2014-03-06  
License: GPL (>=2)

## Author(s)

Manuel Reif <Manuel.Reif@statistik.gv.at>, Jakob Peterbauer <Jakob.Peterbauer@statistik.gv.at>  
Maintainer: Manuel Reif <Manuel.Reif@statistik.gv.at>

## References

Saerndal, C.-E. & Swensson, B. & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.

See also:

OECD (2013). *Technical Report of the Survey of Adult Skills (PIAAC)*. Retrieved from: <http://www.oecd.org/site/piaac/All%20>

## See Also

[svyPVglm](#) [svyPVpm](#)

## Examples

```
data(svy_example1)

erg_ben <- svyPVbenchmark(by = ~ sex, svydat=svy.exrep,
pvs=c("plaus1", "plaus2", "plaus3"), BENCH=320)
```

erg\_ben

---

svy.exrep*Simulated survey design*

---

**Description**

Simulated survey design to show how the package works.

**Usage**

```
svy.exrep
```

**Format**

Survey Design

**Source**

Simulation

---

svyPVbenchmark*Estimate the proportion below and above a bechmark*

---

**Description**

This function works in a similar fashion like the `svyPVlevel` function. It discretizes the plausible values to a dichotomous variable and estimates the proportion of population totals above and below the benchmark within the comitted groups (by statement).

**Usage**

```
svyPVbenchmark(by, svydat, pvs, BENCH=NA, colN=FALSE)
```

**Arguments**

<code>by</code>	A formula statement is expected which splits the data into several subsets.
<code>svydat</code>	A survey design ( <code>svydesign</code> as well as <code>svrepdesign</code> ) which was generated by the survey package. To figure out how to create a survey design object, please read the help files for the survey package.
<code>pvs</code>	A character vector which includes the colnames of the plausible values. These variables must be part of the survey design comitted as <code>svydat</code> .

BENCH	Submit a benchmark (numeric vector of length = 1). A plausible value will be assigned to "< benchmark" if it is below the benchmark and assigned to ">= benchmark" if it is on or above the benchmark.
colN	If TRUE the colnames will equal the grouping variable names from the by statement. If FALSE, which is the default, the names will be Group1 up to Group k.

**Value**

The function returns a data.frame with the following columns

Group1..k	The first k-1 columns show the different levels of the k-1 subsetting groups, provided with by. The kth group column contains the benchmark variable.
Number.of.cases	Shows the unweighted number of cases (NA's excluded) within each group.
Sum.of.weights	Shows the sum of weights (NA's excluded) within each group.
Proportion	Contains the estimate of the conditional proportion of persons below and on/above the benchmark given the categories of the first k-1 groups.
Proportion.SE	Contains the SE of the proportion estimate.

**Author(s)**

Manuel Reif

**References**

- Lumley, T. (2010). *Complex Surveys*. Hoboken, NJ: Wiley.
- Saerndal, C.-E. & Swensson, B. & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.
- Chaudhuri, A. & Stenger, H. (2005). *Survey Sampling. Theory and Methods*. Boca Raton, FL: Chapman & Hall/CRC.

**See Also**

[svyPVlevel](#)

**Examples**

```
data(svy_example1)

erg_ben <- svyPVbenchmark(by = ~ sex, svydat=svy.exrep,
pvs=c("plaus1", "plaus2", "plaus3"), BENCH=320)

erg_ben
```

---

svyPVcor	<i>Survey-weighted Correlation Estimation with usage of plausible values.</i>
----------	---

---

### Description

svyPVcor estimates the Pearson product-moment correlation coefficient and its standard error for data from a complex survey design with plausible values.

### Usage

```
svyPVcor(formula, design, placeholder = 1:10)
```

### Arguments

formula	Formula, $x \sim y$ for the correlation between $x$ and $y$ (both variables have to be part of a survey design object created by the survey package). For a notation description for the plausible values see in 'details'.
design	A survey design which was generated by the survey package .
placeholder	A vector of symbols, which were used for numbering of the plausible values. For a detailed description see in 'details'.

### Details

All variables mentioned in the formula object must be part of the survey design object. Instead of the symbols, which were used for numbering the plausible values use '.' as notation (e.g. placeholder = 1:5 and PVLIT.. stands for PVLIT1, PVLIT2, PVLIT3, PVLIT4, PVLIT5).Missing values are deleted listwise.

### Value

The function returns a data.frame with the following columns

COR	Shows the Pearson product-moment correlation coefficient between $x$ and $y$ .
SE	Shows the SE for the Pearson product-moment correlation between $x$ and $y$ .
Number.of.cases	Shows the unweighted number of cases (NA's excluded) within each group.
Sum.of.weights	Shows the sum of weights (NA's excluded) within each group.

### Author(s)

Jakob Peterbauer

## References

- Lumley, T. (2010). *Complex Surveys*. Hoboken, NJ: Wiley.
- Saerndal, C.-E. & Swensson, B. & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.
- Chaudhuri, A. & Stenger, H. (2005). *Survey Sampling. Theory and Methods*. Boca Raton, FL: Chapman & Hall/CRC.

## See Also

[cov.wt](#), [svyPVeta](#)

## Examples

```
# data(svy_example1)
```

---

svyPVeta	<i>Survey-weighted estimation of the eta coefficient with usage of plausible values.</i>
----------	--

---

## Description

Computes the eta coefficient for data from a complex survey design with usage of plausible values.

## Usage

```
svyPVeta(formula, design, placeholder = 1:10)
```

## Arguments

formula	Formula, x~y for the eta coefficient between x and y (both variables have to be part of a survey design object created by the survey package). For a notation description for the plausible values see in 'details'.
design	A survey design which was generated by the survey package .
placeholder	A vector of symbols, which were used for numbering of the plausible values. For a detailed description see in 'details'.

## Details

All variables mentioned in the formula object must be part of the survey design object. Instead of the symbols, which were used for numbering the plausible values use '.' as notation (e.g. placeholder = 1:5 and PVLIT.. stands for PVLIT1, PVLIT2, PVLIT3, PVLIT4, PVLIT5).Missing values are deleted listwise.

**Value**

The function returns a data.frame with the following columns

ETA	Shows the eta coefficient between x and y.
SE	Shows the SE for the eta coefficient.
Number.of.cases	Shows the unweighted number of cases (NA's excluded) within each group.
Sum.of.weights	Shows the sum of weights (NA's excluded) within each group.

**Author(s)**

Jakob Peterbauer

**References**

Lumley, T. (2010). *Complex Surveys*. Hoboken, NJ: Wiley.

Saerndal, C.-E. & Swensson, B. & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.

Chaudhuri, A. & Stenger, H. (2005). *Survey Sampling. Theory and Methods*. Boca Raton, FL: Chapman & Hall/CRC.

**See Also**

[svyPVcor](#)

**Examples**

```
data(svy_example1)

res_eta <- svyPVeta(plaus.. ~ var3, svy.exrep, placeholder = 1:3)

res_eta
```

---

svyPVglm	<i>Survey-weighted generalised linear models with uasge of palusible values.</i>
----------	--

---

**Description**

Fit a generalised linear model to data from a complex survey design with usage of plausible values.

**Usage**

```
svyPVglm(formula, design, placeholder = 1:10, family = gaussian())
## S3 method for class 'svyPVglm'
summary(object, ...)
```

**Arguments**

formula	a model formula (all variables have to be part of a survey design object created by the survey package). For a notation description for the plausible values see in 'details'.
design	a survey design which was generated by the survey package
placeholder	a vector of symbols, which were used for numbering the plausible values. For a detailed description see in 'details'.
family	a description of the error distribution and link function to be used in the model. This has to be either a character string naming a family function, a family function or the result of a call to a family function.
object	A object of class svyPVglm.
...	

**Details**

All variables mentioned in the formula object must be part of the survey design object. Instead of the symbols, which were used for numbering the plausible values use '.' as notation (e.g. placeholder = 1:5 and PVLIT.. stands for PVLIT1, PVLIT2, PVLIT3, PVLIT4, PVLIT5). Missing values are deleted listwise.

**Value**

The function returns a list, which includes the following dataframes.

coef	Shows the regression coefficients and test statistics.
mod.fit	Shows the model test statistic. For more details see <a href="#">svyglm</a> .

**Author(s)**

Jakob Peterbauer

**References**

- Lumley, T. (2010). *Complex Surveys*. Hoboken, NJ: Wiley.
- Saerndal, C.-E. & Swensson, B. & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.
- Chaudhuri, A. & Stenger, H. (2005). *Survey Sampling. Theory and Methods*. Boca Raton, FL: Chapman & Hall/CRC.

**See Also**

[svyglm](#), [svyPVcor](#), [svyPVeta](#)



**Examples**

```
data(svy_example1)

res_glm <- svyPVglm(plaus.. ~ otherPlaus.. + age, svy.exrep, placeholder = 1:3)

res_glm
```

svyPVlevel

*Proportion in levels estimation***Description**

This function categorizes the plausible values into specific user defined levels, to estimate the proportion of population totals within this levels.

**Usage**

```
svyPVlevel(by, svydat, pvs, CATDEF,levlab=NA,right=TRUE, colN=FALSE)
```

**Arguments**

by	A formula statement is expected which splits the data into several subsets.
svydat	A survey design (svydesign as well as svrepdesign) which was generated by the survey package. To figure out how to create a survey design object, please read the help files for the survey package.
pvs	A character vector which includes the colnames of the plausible values. These variables must be part of the survey design comitted as svydat.
CATDEF	A numeric vector which contains the cut points with which the plausible values are categorized. The input follows the same notation as the break argument in the cut function.
levlab	A character vector which contains the levels labels. Default is NA – which means that the labels will be paste0("level", NUMBEROFLEVELS)
right	Corresponds to the right argument in the cut function. If TRUE, which is the default, the intervals built by this function are closed on the right.
colN	If TRUE the colnames will equal the grouping variable names from the by statement. If FALSE, which is the default, the names will be Group1 up to Group k.

**Details**

All variables, including the subsetting ones defined with by, must be part of the survey design object.

Missing values are deleted listwise.

Note that '.' is not allowed as part of the level string of a by variable. For example c("1.thing", "2.thing") is not allowed and will cause a error message.

**Value**

The function returns a data.frame with the following columns

Group1..k	The first k-1 columns show the different levels of the k-1 subsetting groups, provided with by. The kth group column contains the different levels defined by the CATDEF vector.
Number.of.cases	Shows the unweighted number of cases (NA's excluded) within each group.
Sum.of.weights	Shows the sum of weights (NA's excluded) within each group.
Proportion	Contains the estimate of the conditional proportion of persons on each level given the categories of the first k-1 groups.
Proportion.SE	Contains the SE of the proportion estimate.

**Author(s)**

Manuel Reif

**References**

- Lumley, T. (2010). *Complex Surveys*. Hoboken, NJ: Wiley.
- Saerndal, C.-E. & Swensson, B. & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.
- Chaudhuri, A. & Stenger, H. (2005). *Survey Sampling. Theory and Methods*. Boca Raton, FL: Chapman & Hall/CRC.

**See Also**

[svyPVbenchmark](#)

**Examples**

```
data(svy_example1)

erg_1 <- svyPVlevel(by = ~ sex, svydat=svy.exrep,
pvs=c("plaus1", "plaus2", "plaus3"), CATDEF=c(0,310,322,400))

erg_1
```

svyPVpm

*Mean and proportion estimation***Description**

This function estimates mean, standard deviation and proportion of subsets based on a survey design and plausible values.

**Usage**

```
svyPVpm(by, svydat, pvs, colN=FALSE)
```

**Arguments**

<code>by</code>	A formula statement is expected which splits the data into several subsets. Means and proportions will be estimated within these subsets.
<code>svydat</code>	A survey design which was generated by the survey package.
<code>pvs</code>	A character vector which includes the colnames of the plausible values. These variables must be part of the survey design committed as <code>svydat</code> .
<code>colN</code>	If TRUE the colnames will equal the grouping variable names from the <code>by</code> statement. If FALSE, which is the default, the names will be Group1 up to Group k.

**Details**

All variables, including the subsetting ones defined with `by`, must be part of the survey design object.

Missing values are deleted listwise.

Note that `'.'` is not allowed as part of the level string of a `by` variable. For example `c("1.thing", "2.thing")` is not allowed and will cause a error message.

**Value**

The function returns a data.frame with the following columns

<code>Group1..k</code>	The first k columns show the different levels of the k subsetting groups.
<code>Number.of.cases</code>	Shows the unweighted number of cases (NA's excluded) within each group.
<code>Sum.of.weights</code>	Shows the sum of weights (NA's excluded) within each group.
<code>Proportion</code>	Shows the (weighted) estimated proportion of persons within the categories.
<code>Proportion.SE</code>	Shows the Standard Errors of the proportion estimate.
<code>pvs_mean</code>	Shows the mean estimate of plausible values within each group.
<code>pvs_mean.SE</code>	Denotes the Standard error of the mean estimate.
<code>pvs_stddev</code>	Shows the standard deviation (sd) estimate of plausible values within each group.
<code>pvs_stddev.SE</code>	Denotes the Standard error of the sd estimate.

**Author(s)**

Manuel Reif

**References**

- Lumley, T. (2010). *Complex Surveys*. Hoboken, NJ: Wiley.
- Saerndal, C.-E. & Swensson, B. & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.
- Chaudhuri, A. & Stenger, H. (2005). *Survey Sampling. Theory and Methods*. Boca Raton, FL: Chapman & Hall/CRC.

**See Also**

[svyPVglm](#) [svyPVprob](#)

**Examples**

```
data(svy_example1)

erg_pm <- svyPVpm(by = ~ sex, svydat=svy.exrep, pvs=c("plaus1","plaus2","plaus3"))

erg_pm
```

---

svyPVprob

*Proportion estimation*

---

**Description**

This function was created to estimate the proportions of weighted observations within each group.

**Usage**

```
svyPVprob(by, svydat, pvs = NULL, colN=FALSE)
```

**Arguments**

by	A formula statement is expected which splits the data into several subsets.
svydat	A survey design (svydesign as well as svrepdesign) which was generated by the survey package.
pvs	Either a character vector which contains variablenames (these variables must exist within the survey) or a vector of length 1 which contains NULL. If plausible values (or any other variablenames) are provided by the argument pvs the weighted ratio of group members is computed after listwise deletion of those who contain any NA.

`colN` If TRUE the colnames will equal the grouping variable names from the by statement. If FALSE, which is the default, the names will be Group1 up to Group k.

### Value

The function returns a data.frame with the following columns

<code>Group1 . . k</code>	The first k columns show the different levels of the k subsetting groups.
<code>Number . of . cases</code>	Shows the unweighted number of cases (NA's excluded) within each group.
<code>Sum . of . weights</code>	Shows the sum of weights (NA's excluded) within each group.
<code>Proportion</code>	Shows the (weighted) estimated proportion of persons within the categories.
<code>Proportion . SE</code>	Shows the Standard Errors of the proportion estimate.

### Author(s)

Manuel Reif

### References

Lumley, T. (2010). *Complex Surveys*. Hoboken, NJ: Wiley.

Saerndal, C.-E. & Swensson, B. & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.

Chaudhuri, A. & Stenger, H. (2005). *Survey Sampling. Theory and Methods*. Boca Raton, FL: Chapman & Hall/CRC.

### See Also

[svyPVprob](#)

### Examples

```
data(svy_example1)

erg_p <- svyPVprob(by = ~ sex, svydat=svy.exrep, pvs=c("plaus1", "plaus2", "plaus3"))

erg_p
```

---

svyPVquantile	<i>Quantile estimation</i>
---------------	----------------------------

---

**Description**

This function estimates quantiles of plausible values within groups.

**Usage**

```
svyPVquantile(by, svydat, pvs, quantile, interval.type = "quantile", colN=FALSE, ...)
```

**Arguments**

by	A formula statement is expected which splits the data into several subsets. Within these subsets the quantiles are estimated.
svydat	A survey design (svydesign as well as svrepdesign) which was generated by the survey package. To figure out how to create a survey design object, please read the help files for the survey package.
pvs	A character vector which includes the colnames of the plausible values. These variables must be part of the survey design committed as svydat.
quantile	A numeric vector of length > 1 which contains the quantiles of interest.
interval.type	A character vector of length = 1 which denotes the kind of quantile estimation method. Valid inputs are: "quantile" and "probability" if a svrepdesign was submitted and in case of a ordinary survey design the valid inputs are: "Wald", "score" and "betaWald". More information about the different estimation techniques can be gained from the "Details" section of the svyquantile function in the survey package.
...	More arguments which are forwarded to svyquantile.
colN	If TRUE the colnames will equal the grouping variable names from the by statement. If FALSE, which is the default, the names will be Group1 up to Group k.

**Value**

The function returns a data.frame with the following columns

Group1..k	The first k columns show the different levels of the k subsetting groups.
Number.of.cases	Shows the unweighted number of cases (NA's excluded) within each group.
Sum.of.weights	Shows the sum of weights (NA's excluded) within each group.
q0...	These columns contain the quantile values which were estimated by means of the submitted plausible values.

**Author(s)**

Manuel Reif

## References

- Lumley, T. (2010). *Complex Surveys*. Hoboken, NJ: Wiley.
- Saerndal, C.-E. & Swensson, B. & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.
- Chaudhuri, A. & Stenger, H. (2005). *Survey Sampling. Theory and Methods*. Boca Raton, FL: Chapman & Hall/CRC.

## See Also

[svyPVpm](#)

## Examples

```
data(svy_example1)

erg_q <- svyPVquantile(by = ~ sex, svydat=svy.exrep,
pvs=c("plaus1", "plaus2", "plaus3"), quantile=c(0.2,0.7))

erg_q
```

---

svyPVttest	<i>Computation of a t-test for data from a complex survey design with usage of plausible values</i>
------------	---

---

## Description

Computes a one-sample or two-sample t-test for data from a complex survey design with usage of plausible values.

## Usage

```
svyPVttest(formula, design, placeholder = 1:10)
```

## Arguments

formula	Formula, outcome~group for two-sample, outcome~0 or outcome~1 for one-sample (all variables have to be part of a survey design object created by the survey package). For a notation description for the plausible values see in 'details'.
design	a survey design which was generated by the survey package.
placeholder	a vector of symbols, which were used for numbering the plausible values. For a detailed description see in 'details'.

**Details**

All variables mentioned in the formula object must be part of the survey design object. Instead of the symbols, which were used for numbering the plausible values use '..' as notation (e.g. placeholder = 1:5 and PVLIT.. stands for PVLIT1, PVLIT2, PVLIT3, PVLIT4, PVLIT5). Missing values are deleted listwise.

**Value**

The function returns a list, which includes the following three dataframes.

DESC	Shows mean and SE for each group.
TEST	Shows the test statistic (t.value, degf and Pr.t).
N	Shows the unweighted number of cases and the sum of weights (NA's excluded) within each group.

**Author(s)**

Jakob Peterbauer

**References**

- Lumley, T. (2010). *Complex Surveys*. Hoboken, NJ: Wiley.
- Saerndal, C.-E. & Swensson, B. & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.
- Chaudhuri, A. & Stenger, H. (2005). *Survey Sampling. Theory and Methods*. Boca Raton, FL: Chapman & Hall/CRC. .

**See Also**

[svytttest](#), [svyPVpm](#), [svyPVglm](#)

**Examples**

```
data(svy_example1)
res <- svyPVttest(plaus.. ~ sex, svy.exrep, placeholder = 1:3)

res
```



# Index

- \*Topic **Mean**
  - svyPVpm, [11](#)
- \*Topic **benchmark**
  - svyPVbenchmark, [3](#)
- \*Topic **correlation**
  - svyPVcor, [5](#)
- \*Topic **datasets**
  - svy.exrep, [3](#)
- \*Topic **glm**
  - svyPVeta, [6](#)
  - svyPVglm, [7](#)
- \*Topic **level**
  - svyPVlevel, [9](#)
- \*Topic **package**
  - svyPVpack-package, [2](#)
- \*Topic **proportion**
  - svyPVpm, [11](#)
  - svyPVprob, [12](#)
- \*Topic **quantile**
  - svyPVquantile, [14](#)
- \*Topic **ttest**
  - svyPVttest, [15](#)

cov.wt, [6](#)

summary.svyPVglm (svyPVglm), [7](#)

svy.exrep, [3](#)

svyglm, [8](#)

svyPVbenchmark, [3](#), [10](#)

svyPVcor, [5](#), [7](#), [8](#)

svyPVeta, [6](#), [6](#), [8](#)

svyPVglm, [2](#), [7](#), [12](#), [16](#)

svyPVlevel, [4](#), [9](#)

svyPVpack (svyPVpack-package), [2](#)

svyPVpack-package, [2](#)

svyPVpm, [2](#), [11](#), [15](#), [16](#)

svyPVprob, [12](#), [12](#), [13](#)

svyPVquantile, [14](#)

svyPVttest, [15](#)

svyttest, [16](#)