Package ‘svyVGAM’

March 31, 2023

**Title**  Design-Based Inference in Vector Generalised Linear Models

**Version**  1.2

**Author**  Thomas Lumley

**Description**  Provides inference based on the survey package for the wide range of parametric models in the ‘VGAM’ package.

**Imports**  stats, methods

**Depends**  VGAM, survey, R (>= 3.5.0)

**Suggests**  pscl, knitr, markdown, rmarkdown

**VignetteBuilder**  knitr

**Encoding**  UTF-8

**License**  GPL-3

**Maintainer**  Thomas Lumley <t.lumley@auckland.ac.nz>

**NeedsCompilation**  no

**Repository**  CRAN

**Date/Publication**  2023-03-30 22:00:02 UTC

**R topics documented:**

- nhanes_sxq . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1
- svy_vglm . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2

**Index**  5

---

**nhanes_sxq**

*Data from NHANES: number of sex partners*

---

**Description**

These data are from the NHANES 2003-2004 survey in the US. They provide an example of overdispersed count data that motivates a two-component zero-inflation model
svy_vglm

Usage

data("nhanes_sxq")

Format

A data frame with 2992 observations on the following 7 variables.

SDMVPSU  Primary Sampling Unit
SDMVSTRA  stratum
WTINT2YR  weights
malepartners  lifetime number of male sexual partners
RIDAGEYR  age in years
DMDEDUC  level of education: 1=less than high school, 2=high school, 3-more than high school, 7=refused
RIDRETH1  Race/ethnicity: 1=Mexican American, 2=Other Hispanic, 4=non-Hispanic White, 5=non-Hispanic Black, 5=Other

Source

NHANES files demo_c.xpt and sxq_c.xpt

See Also

Construction of the data set is described by https://notstatschat.rbind.io/2015/05/26/zero-inflated-poisson-from-complex-samples/

Examples

data(nhanes_sxq)
hdes = svydesign(id=~SDMVPSU,strat=~SDMVSTRA,weights=~WTINT2YR, nest=TRUE, data=nhanes_sxq)
svy_vglm(malepartners~RIDAGEYR+factor(RIDRETH1)+DMDEDUC, zipoisson(), design=hdes, crit = "coef")

svy_vglm  Design-based inference for vector generalised linear models

Description

This function provides design-based (survey) inference for Thomas Yee’s vector generalised linear models. It works by calling vglm with sampling weights, and then either using resampling (replicate weights) or extracting the influence functions and using a Horvitz-Thompson-type sandwich estimator.

Usage

svy_vglm(formula, family, design, ...)
svy_vglm

Arguments

formula        Model formula, as for vglm
family         Model family, as for vglm
design         Survey design object
...            Other arguments to pass to vglm

Value

An S3 object of class svy_glm with print, coef and vcov methods, containing the design in the
design component and a fitted vglm object in the fit component.

See Also

nhanes_sxq
vglm
svydesign svrepdesign

Examples

data(api)
dclus2<-svydesign(id=~dnum+snum, fpc=~fpc1+fpc2, data=apiclus2)

## Ordinary Gaussian regression
m1<-svyglm(api00~api99+mobility+ell, design=dclus2,family=gaussian)
## same model, but with the variance as a second parameter
m2<-svy_vglm(api00~api99+mobility+ell, design=dclus2,family=uninormal())
m1
m2
SE(m1)
SE(m2)

summary(m1)
summary(m2)

## Proportional odds model
dclus2<-update(dclus2, mealcat=as.ordered(cut(meals,c(0,25,50,75,100))))
a<-svyolr(mealcat~avg.ed+mobility+stype, design=dclus2)
b<-svy_vglm(mealcat~avg.ed+mobility+stype, design=dclus2, family=propodds())
a
b
SE(a)
SE(b) #not identical, because svyolr() uses approximate Hessian

## Zero-inflated Poisson
data(nhanes_sxq)

nhdes = svydesign(id=~SDMVPSU,strat=~SDMVSTRA,weights=~WTINT2YR,
est=TRUE, data=nhanes_sxq)

sv1<-svy_vglm(malepartners~RIDAGEYR+factor(RIDRETH1)+DMDEDUC,
svy_vglm(zipoisson(), design=nhdes, crit = "coef")

summary(sv1)

## Multinomial
## Reference group (non-Hispanic White) average older and more educated
## so coefficients are negative
mult_eth<- svy_vglm(RIDRETH1 ~ RIDAGEYR + DMDEDUC,
                      family=multinomial(reflevel=3), design=nhdes)

## separate logistic regressions are close but not identical
two_eth<- svyglm(I(RIDRETH1==1) ~ RIDAGEYR + DMDEDUC, family=quasibinomial,
                  design=subset(nhdes, RIDRETH1 %in% c(1,3)))

summary(mult_eth)
summary(two_eth)
Index

* datasets
  nhanes_sxq, 1
* survey
  svy_vglm, 2

nhanes_sxq, 1, 3
svrepdesign, 3
svy_vglm, 2
svydesign, 3
vglm, 3