

ASDA3 Analysis Examples Replication Chapter 6

R Code

```
# setwd ("P:\\ASDA3\\Data Sets for Analysis Examples and Stata R Code")

library(survey)
library(haven)

# Read NHANES data file in R format
load("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/nhanes1112.rdata")

# Create survey design object
nhanes.des <- svydesign(id = ~sdmvpstu, strata = ~sdmvstra, weights = ~wtmec2yr, nest = T, data = nhanes1112)

# Example 6.1
(ex61 <- svymean(~factor(irregular), subnhanes, se=T, na.rm=T, deff=T, ci=T, keep.vars=T))
confint(ex61)
ex61p <- svyciprop(~factor(irregular), subnhanes, se=T, na.rm=T, deff=T, ci=T, keep.vars=T)
ex61p

# Example 6.1, Franco et al. intervals

# Load key function from Franco et al. (2019) supplemental materials
source("http://umich.edu/~bwest/CIarrFcn.R")

# Define a function for computing the intervals in Franco et al. (2019)
# Author: Sofi Sinozich (sofi.sinozich@gmail.com)
surveyci <- function(formula, design, deff) {
  if(length(all.vars(formula))>1) {
    message("Can only be used for single variables.")
    return()
  }
  # compute design-adjusted estimates of the proportions and the DEFF
  p <- svymean(formula, design, na.rm=TRUE, deff=deff)
  # compute the effective sample size needed for the adjusted intervals
  n_eff <- dim(design$variables[all.vars(formula)])[1]/deff(p)
  # Compute and return the Franco et al. (2019) adjusted intervals
  CIarrFcn(p*n_eff,n_eff,0.05)
}

# Compute adjusted Franco et al. (2019) intervals
surveyci(~irregular, subset(nhanes.des, age >= 18), deff="replace")

# Comparison with more standard intervals
irr_p <- svymean(~irregular, subset(nhanes.des, age >= 18), na.rm=TRUE, deff="replace")
# Wald
confint(irr_p,df=17)
# Logit
confint(svyciprop(~irregular, subset(nhanes.des, age >= 18), method="logit", df=17))

##
# Bayesian Approach
# read in the full NHANES data set for the third edition, note the working directory is set already
load("nhanes1112.rdata")

# create complete data set with variables of interest
nhanes.red <- nhanes1112[nhanes1112$age >= 18 & !is.na(nhanes1112$age), c("irregular", "sdmvpstu", "sdmvstra",
"wtmec2yr")]
nhanes.red <- nhanes.red[complete.cases(nhanes.red),]
```

```

# need to normalize NHANES weights to match what is done for Stan modeling
nhanes.red$wtsc <- nhanes.red$wtmec2yr / mean(nhanes.red$wtmec2yr)

# survey design object
nhanes.des <- svydesign(id = ~sdmvpsu, strata = ~sdmvstra, weights = ~wtsc, nest = T, data = nhanes.red)

# Bayesian approach, flat prior
set.seed(41280)
model_formula <- formula("irregular|weights(wtsc) ~ 1")
mod.brms <- cs_sampling_brms(svydes = nhanes.des, brmsmod = brmsformula(model_formula), data = nhanes.red,
family = bernoulli())
mod.brms

# Example 6.2 NHANES ADULT DATA, Design Based Approach
ex62 <- svymean(~factor(racec), design=subnhanes, se=T, na.rm=T, deff=T, ci=T, keep.vars=T)
ex62
confint(ex62)

# Example 6.2 Bayesian Approach

# read in the full NHANES data set for the third edition
load("P:\\ASDA3\\Data Sets for Analysis Examples and Stata R Code\\nhanes1112.rdata")

# create complete data set with variables of interest
nhanes.red <- nhanes1112[nhanes1112$age >= 18 & !is.na(nhanes1112$age), c("ridreth1", "sdmvpsu", "sdmvstra",
"wtmec2yr")]
nhanes.red <- nhanes.red[complete.cases(nhanes.red),]

# need to normalize NHANES weights to match what is done for Stan modeling
nhanes.red$wtsc <- nhanes.red$wtmec2yr / mean(nhanes.red$wtmec2yr)

# survey design object
nhanes.des <- svydesign(id = ~sdmvpsu, strata = ~sdmvstra, weights = ~wtsc, nest = T, data = nhanes.red)

# Create jackknife replicate weights
nhanes.rep <- as.svrepdesign(nhanes.des)

# Define the Dirichlet multinomial model for Stan
mod_dm <- stan_model(model_code = load_wt_multi_model())

# Prepare the data
y <- as.factor(nhanes.rep$variables$ridreth1)
yM <- model.matrix(~y -1)
n <- dim(yM)[1]
K <- dim(yM)[2]
alpha <- rep(1,K)
weights <- nhanes.rep$pweights
data_stan <- list("y"=yM,"alpha"=alpha,"K"=K,"n"=n,"weights"=weights)
ctrl_stan <- list("chains"=1,"iter"=2000,"warmup"=1000,"thin"=1)

# Run the Stan model with cs_sampling
mod1 <- cs_sampling(svydes = nhanes.rep, mod_stan = mod_dm, data_stan = data_stan, ctrl_stan = ctrl_stan,
rep_design = TRUE)

# View results
> mod1$stan_fit

```

```

# Example 6.3 NHANES ADULT DATA
ex63 <- svymean(~factor(bp_catc), subnhanes, se=T, na.rm=T, deff=T, ci=T, keep.vars=T)
ex63
confint(ex63)

# EXAMPLE 6.4 ESS6 Russian Federation Data, Proportions of Russian 15+ by Marital Status
# GOF TOOL WITH PRE-SET PROPORTIONS (NOT AVAILABLE IN R)
# read SAS data set for this example (need Haven)

rfdata <- read_sas("P:\\ASDA3\\Data Sets for Analysis Examples and Stata R Code\\ess6_russia.sas7bdat")
summary(rfdata)

#create factor variables
rfdata$marcatc <- factor(rfdata$marcat, levels = 1:3, labels =c("Married", "Previous", "Never"))

rfsvy <- svydesign(strata=~stratify, id=~psu, weights=~pspwght, data=rfdata, nest=T)

ex6_4 <- svymean(~factor(marcatc), design=rfsvy, na.rm=T, se=T, deff=T, ci=T, keep.vars=T)
print(ex6_4)

# Analysis Example 6.5 PIE AND BAR CHARTS

# Pie chart of Marital Status Russian Federation Data
png("P:/ASDA3/Replication R/Chapter 5/fig65_pie.png")
ex6_5 <- svymean(~factor(marcatc), rfsvy, se=T, na.rm=T, deff=T, ci=T, keep.vars=T)
pie(ex6_5, col=c("black", "grey60", "blue", "red"), c("Married", "Previously Married", "Never Married"))
dev.off()

# Bar chart of marital status
png("P:/ASDA3/Replication R/Chapter 5/fig65_bar.png")
barplot(ex6_5, legend=c("Married", "Previously Married", "Never Married"), col=c("black","blue", "red"))
dev.off()

# Analysis Example 6.6, NCS-R DATA (see C5 for data prep), already complete for this analysis
(ex6_6 <- svymean(~interaction (sex, mde), ncsrsvyp1, se=T, na.rm=T, ci=T, keep.vars=T))
# obtain confidence intervals
confint(ex6_6)

# svyby analysis gives mean of mde by sex
(ex6_6a <- svyby(~mde, ~sex, ncsrsvyp1, svymean, se=T, na.rm=T, ci=T, keep.vars=T))

#CODES FOR SEX 1=MALE 2=FEMALE
#svychisq provides a 2 by 2 chisq test (F)
svychisq(~sex+mde, ncsrsvyp1, statistic="F")

# Analysis Example 6.7 MEAN OF MDE OVER SEX AND LINEAR COMPARISON TEST
(ex6_7 <- svyby(~mde, ~sex, ncsrsvyp1, svymean, se=T, na.rm=T, ci=T, keep.vars=T))
svycontrast(ex6_7,list(avg=c(.5,.5), diff=c(1,-1)))

```

```

# Example 6.7 Bayesian Approach

# read in the full NCS-R data set for the third edition, Rdata format
load("P:\\ASDA3\\Data Sets for Analysis Examples and Stata R Code\\ncsr.rdata")

# create complete data set with variables of interest
ncsr.red <- ncsr[, c("mde", "sexm", "seclustr", "sestrat", "ncsrwtsh")]
ncsr.red <- ncsr.red[complete.cases(ncsr.red),]

# need to normalize NHANES weights to match what is done for Stan modeling
ncsr.red$wtsc <- ncsr.red$ncsrwtsh / mean(ncsr.red$ncsrwtsh)

# survey design object
ncsr.des <- svydesign(id = ~seclustr, strata = ~sestrat, weights = ~wtsc, nest = T, data = ncsr.red)

# Bayesian approach, flat prior
set.seed(41279)
model_formula <- formula("mde|weights(wtsc) ~ sexm")
mod.brms <- cs_sampling_brms(svydes = ncsr.des, brmsmod = brmsformula(model_formula), data = ncsr.red, family =
bernoulli())
mod.brms$stan_fit

## Numbers in C6 of text: (0.151 / 0.849) / (0.229 / 0.771)

# Analysis Example 6.8 Independence of MDE and Gender
(ex6_8 <- svyby(~mde, ~sex, ncsrsvyp1, svymean, se=T, na.rm=T, ci=T, keep.vars=T))

#Codes for sex: 1=MALE 2=FEMALE
#svychisq provides a 2 by 2 chisq test (F)
svychisq(~sex+mde, ncsrsvyp1, statistic="F")

# Analysis Example 6.9 Independence of Education and Alcohol Dependence
ex6_9 <- svyby(~ald,~ed4catc, subset(ncsrsvyp2, age < 29 & !is.na(ED4CAT) & !is.na(ald)), svymean, na.rm=T,
ci=T)
# Codes for ed4cat : 1=0-11 2=12 3=13-15 4=16+ Years of Education
print(ex6_9)
svychisq(~ald+ ed4catc, subset(ncsrsvyp2, age < 29 & !is.na(ED4CAT) & !is.na(ald)), na.rm=T, statistic = "F")

# Analysis Example 6.10 Logistic Regression MDE regressed on Gender
(ex6_10 <- svyglm (mde~sexm, design=ncsrsvyp1, family=quasibinomial))
summary(ex6_10)
#note use exponent function with beta to obtain OR
or <- exp(-.516)
or

# Figure 6.8 Bar Chart of Marital Status in Russian Federation data
fig6_8 <- svyby(~factor(marcatc), ~gndr, rfsvy, svymean, na.rm=T)

# obtain ChiSQ test
svychisq(~mde+sexm, ncsrsvyp1, statistic="Chisq")

```

```
# Bar chart of marital status by gender
print(fig6_8)
```

```
png("P:/ASDA3/Replication R/Chapter 5/fig6_8.png")
barplot(fig6_8, legend=c("Married", "Previously Married", "Never Married"), col=c("black", "blue", "red"),
xlab=c("Male", "Female") )
dev.off()
```

```
# Analysis Example 6.11 Independence of Gender and MDE while controlling for Age, not available in R Survey
Package
```

```
# Analysis Example 6.12 Loglinear Model Examining Relationship Between MDE and Male
#null model run first
null <-svyloglin(~mde+sexm,ncsrsvyp1)
summary(null)
```

```
# saturated model, update null model with interaction of mde and sexm
saturated <- update(null, ~.+mde:sexm)
summary(saturated)
```

```
# obtain F test
svychisq(~mde+sexm, ncsrsvyp1)
# obtain ChiSQ test
svychisq(~mde+sexm, ncsrsvyp1, statistic="Chisq")
```

R Results

```
> # Read NHANES data file
> load("P:/ASDA3/Data Sets for Analysis Examples and Stata R Code/nhanes1112.rdata")

> # Create survey design object
> nhanes.des <- svydesign(id = ~sdmvpstu, strata = ~sdmvstrata, weights = ~wtmec2yr, nest = T, data = nhanes1112)

# Example 6.1
> (ex61 <- svymean(~factor(irregular), subnhanes, se=T, na.rm=T, deff=T, ci=T, keep.vars=T))
              mean      SE  DEff
factor(irregular)0 0.9835810 0.0016779 0.9367
factor(irregular)1 0.0164190 0.0016779 0.9367
> confint(ex61)
              2.5 %      97.5 %
factor(irregular)0 0.98029247 0.98686955
factor(irregular)1 0.01313045 0.01970753
> ex61p <- svyciprop(~factor(irregular), subnhanes, se=T, na.rm=T, deff=T, ci=T, keep.vars=T)
> ex61p
              2.5%  97.5%
factor(irregular) 0.0164 0.0132 0.0204

> # Load key function from Franco et al. (2019) supplemental materials
> source("http://umich.edu/~bwest/CIarrFcn.R")
>
> # Define a function for computing the intervals in Franco et al. (2019)
> # Author: Sofi Sinozich (sofi.sinozich@gmail.com)
> surveyci <- function(formula, design, deff) {
+   if(length(all.vars(formula))>1) {
+     message("Can only be used for single variables.")
+     return()
+   }
+   # compute design-adjusted estimates of the proportions and the DEFF
+   p <- svymean(formula, design, na.rm=TRUE, deff=deff)
+   # compute the effective sample size needed for the adjusted intervals
+   n_eff <- dim(design$variables[[all.vars(formula)]])[1]/deff(p)
+   # Compute and return the Franco et al. (2019) adjusted intervals
+   CIarrFcn(p*n_eff,n_eff,0.05)
+ }

> # Compute adjusted Franco et al. (2019) intervals
> surveyci(~irregular, subset(nhanes.des, age >= 18), deff="replace")
, , Wald

              Lowr      Upr      Width Flag LoInd
1 0.01327113 0.01956684 0.00629571    0    0

, , JeffPr

              Lowr      Upr      Width Flag LoInd
1 0.01349042 0.01979436 0.006303947    0    0

, , UnifPr

              Lowr      Upr      Width Flag LoInd
1 0.01356057 0.01987853 0.006317955    0    0

, , ClPe
```

	Lowr	Upr	Width	Flag	LoInd
1	0.01341811	0.01988169	0.006463582	0	0

, , Wils

	Lowr	Upr	Width	Flag	LoInd
1	0.01355469	0.01987635	0.00632166	0	0

, , AgCo

	Lowr	Upr	Width	Flag	LoInd
1	0.01354082	0.01989022	0.006349402	0	0

, , Assqr

	Lowr	Upr	Width	Flag	LoInd
1	0.01348974	0.01979935	0.006309606	0	0

, , Logit

	Lowr	Upr	Width	Flag	LoInd
1	0.01355058	0.01988234	0.006331755	0	0

```
> # Comparison with more standard intervals
> irr_p <- svymean(~irregular, subset(nhanes.des, age >= 18), na.rm=TRUE, deff="replace")
> # Wald
> confint(irr_p,df=17)
                2.5 %      97.5 %
irregular 0.01287902 0.01995896

> # Logit
> confint(svyciprop(~irregular, subset(nhanes.des, age >= 18), method="logit", df=17))
                2.5%      97.5%
irregular 0.01322983 0.02036105
```

Example 6.2 NHANES ADULT DATA, Design Based Approach

```
> ex62 <- svymean(~factor(racec), design=subnhanes, se=T, na.rm=T, deff=T, ci=T, keep.vars=T)
> ex62
```

	mean	SE	DEff
factor(racec)Mexican	0.079168	0.017251	22.9171
factor(racec)Other Hispanic	0.066224	0.015193	20.9551
factor(racec)White	0.659386	0.038892	37.8086
factor(racec)Black	0.117185	0.023370	29.6387
factor(racec)Other	0.078037	0.010917	9.3001

>

```
> confint(ex62)
                2.5 %      97.5 %
factor(racec)Mexican      0.04535738 0.11297837
factor(racec)Other Hispanic 0.03644730 0.09600144
factor(racec)White        0.58316044 0.73561243
factor(racec)Black        0.07138031 0.16298887
factor(racec)Other        0.05663959 0.09943387
```

```

> # Example 6.2 Bayesian Approach
> # Read in the full NHANES data set for the third edition, if already loaded can omit, we repeat for clarity

> load("P:\\ASDA3\\Data Sets for Analysis Examples and Stata R Code\\nhanes1112.rdata")
>
> # create complete data set with variables of interest
> nhanes.red <- nhanes1112[nhanes1112$age >= 18 & !is.na(nhanes1112$age), c("ridreth1", "sdmvpsu", "sdmvstra",
"wtmec2yr")]
> nhanes.red <- nhanes.red[complete.cases(nhanes.red),]
>
> # need to normalize NHANES weights to match what is done for Stan modeling
> nhanes.red$wtsc <- nhanes.red$wtmec2yr / mean(nhanes.red$wtmec2yr)
>
> # survey design object
> nhanes.des <- svydesign(id = ~sdmvpsu, strata = ~sdmvstra, weights = ~wtsc, nest = T, data = nhanes.red)
>
> # Create jackknife replicate weights
> nhanes.rep <- as.svrepdesign(nhanes.des)
>
> # Define the Dirichlet multinomial model for Stan
> mod_dm <- stan_model(model_code = load_wt_multi_model())
>
> # Prepare the data
> y <- as.factor(nhanes.rep$variables$ridreth1)
> yM <- model.matrix(~y -1)
> n <- dim(yM)[1]
> K <- dim(yM)[2]
> alpha <- rep(1,K)
> weights <- nhanes.rep$pweights
> data_stan <- list("y"=yM,"alpha"=alpha,"K"=K,"n"=n,"weights"=weights)
> ctrl_stan <- list("chains"=1,"iter"=2000,"warmup"=1000,"thin"=1)
>
> # Run the Stan model with cs_sampling
> mod1 <- cs_sampling(svydes = nhanes.rep, mod_stan = mod_dm, data_stan = data_stan, ctrl_stan = ctrl_stan,
rep_design = TRUE)
[1] "stan fitting"

```

SAMPLING FOR MODEL 'anon_model' NOW (CHAIN 1).

Chain 1:

Chain 1: Gradient evaluation took 0.006363 seconds

Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 63.63 seconds.

Chain 1: Adjust your expectations accordingly!

Chain 1:

Chain 1:

Chain 1: Iteration: 1 / 2000 [0%] (Warmup)

Chain 1: Iteration: 200 / 2000 [10%] (Warmup)

Chain 1: Iteration: 400 / 2000 [20%] (Warmup)

Chain 1: Iteration: 600 / 2000 [30%] (Warmup)

Chain 1: Iteration: 800 / 2000 [40%] (Warmup)

Chain 1: Iteration: 1000 / 2000 [50%] (Warmup)

Chain 1: Iteration: 1001 / 2000 [50%] (Sampling)

Chain 1: Iteration: 1200 / 2000 [60%] (Sampling)

Chain 1: Iteration: 1400 / 2000 [70%] (Sampling)

Chain 1: Iteration: 1600 / 2000 [80%] (Sampling)

Chain 1: Iteration: 1800 / 2000 [90%] (Sampling)

Chain 1: Iteration: 2000 / 2000 [100%] (Sampling)

Chain 1:

Chain 1: Elapsed Time: 126.532 seconds (Warm-up)

Chain 1: 125.485 seconds (Sampling)

Chain 1: 252.017 seconds (Total)

Chain 1:

[1] "gradient evaluation"

Warning messages:

1: Bulk Effective Samples Size (ESS) is too low, indicating posterior means and medians may be unreliable. Running the chains for more iterations may help. See

<https://mc-stan.org/misc/warnings.html#bulk-ess>

2: Tail Effective Samples Size (ESS) is too low, indicating posterior variances and tail quantiles may be unreliable.

Running the chains for more iterations may help. See

<https://mc-stan.org/misc/warnings.html#tail-ess>

>

> # View results

> mod1\$stan_fit

Inference for Stan model: anon_model.

1 chains, each with iter=2000; warmup=1000; thin=1;

post-warmup draws per chain=1000, total post-warmup draws=1000.

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
lambda[1]	0.40	0.02	0.19	0.13	0.27	0.38	0.50	0.86	69	1.02
lambda[2]	0.34	0.02	0.15	0.11	0.23	0.32	0.42	0.68	68	1.03
lambda[3]	3.36	0.18	1.53	1.13	2.21	3.14	4.18	6.91	69	1.03
lambda[4]	0.60	0.03	0.27	0.20	0.40	0.56	0.75	1.22	68	1.02
lambda[5]	0.40	0.02	0.18	0.13	0.27	0.37	0.49	0.83	68	1.03
theta[1]	0.08	0.00	0.00	0.07	0.08	0.08	0.08	0.09	962	1.00
theta[2]	0.07	0.00	0.00	0.06	0.06	0.07	0.07	0.07	1102	1.00
theta[3]	0.66	0.00	0.01	0.65	0.66	0.66	0.66	0.67	1167	1.00
theta[4]	0.12	0.00	0.00	0.11	0.11	0.12	0.12	0.13	1150	1.00
theta[5]	0.08	0.00	0.00	0.07	0.08	0.08	0.08	0.09	1052	1.00
loglam[1]	-1.01	0.06	0.48	-2.01	-1.32	-0.98	-0.69	-0.15	54	1.02
loglam[2]	-1.19	0.07	0.48	-2.19	-1.49	-1.15	-0.87	-0.38	53	1.02
loglam[3]	1.10	0.07	0.47	0.12	0.79	1.14	1.43	1.93	53	1.02
loglam[4]	-0.62	0.07	0.48	-1.62	-0.92	-0.58	-0.29	0.20	53	1.02
loglam[5]	-1.03	0.07	0.48	-2.04	-1.32	-0.99	-0.71	-0.19	53	1.03
lp__	-6492.18	0.12	1.64	-6496.09	-6493.09	-6491.83	-6490.93	-6490.02	179	1.01

Samples were drawn using NUTS(diag_e) at Tue Mar 11 12:47:54 2025.

For each parameter, n_eff is a crude measure of effective sample size, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat=1).

Example 6.3 NHANES ADULT DATA

> ex63 <- svymean(~factor(bp_catc), subnhanes, se=T, na.rm=T, deff=T, ci=T, keep.vars=T)

> ex63

	mean	SE	DEff
factor(bp_catc)Normal	0.4722233	0.0155209	5.1762
factor(bp_catc)Pre-HBP	0.4279854	0.0120353	3.1685
factor(bp_catc)Stage 1 HBP	0.0797780	0.0058154	2.4669
factor(bp_catc)Stage 2 HBP	0.0200133	0.0043847	5.2494

	2.5 %	97.5 %
factor(bp_catc)Normal	0.44180279	0.50264372
factor(bp_catc)Pre-HBP	0.40439669	0.45157418
factor(bp_catc)Stage 1 HBP	0.06837999	0.09117605
factor(bp_catc)Stage 2 HBP	0.01141946	0.02860712

```

# EXAMPLE 6.4 ESS6 Russian Federation Data, Proportions of Russian 15+ by Marital Status
> # GOF TOOL WITH PRE-SET PROPORTIONS (NOT AVAILABLE IN R)

> rfsvy <- svydesign(strata=~stratify, id=~psu, weights=~pspwght, data=rfddata, nest=T)
>
>
> ex6_4 <- svymean(~factor(marcatc), design=rfsvy, na.rm=T, se=T, deff=T, ci=T, keep.vars=T)
> print(ex6_4)

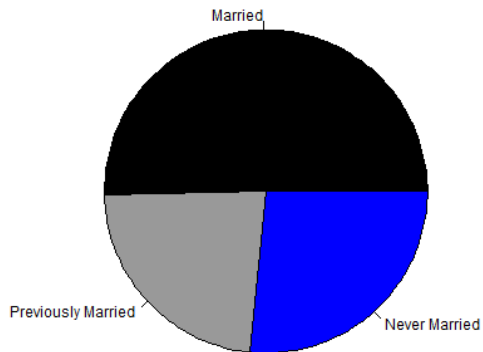
```

	mean	SE	DEff
factor(marcatc)Married	0.503860	0.012878	442.85
factor(marcatc)Previous	0.230066	0.011536	501.57
factor(marcatc)Never	0.266074	0.013401	613.97

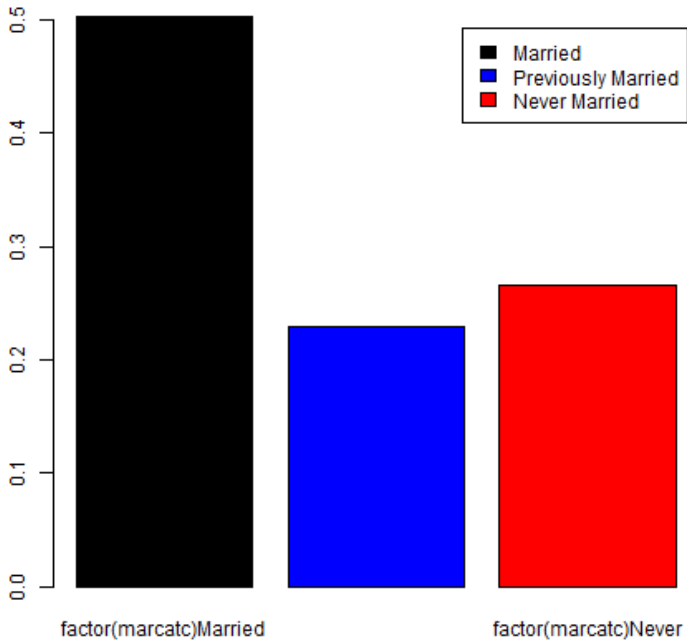
```

> # Pie of Marital Status Russian Federation Data
> ex6_5 <- svymean(~factor(marcatc), rfsvy, se=T, na.rm=T, deff=T, ci=T, keep.vars=T)
> pie(ex6_5, col=c("black", "grey60", "blue", "red"), c("Married", "Previously Married", "Never Married"))

```



```
> # Bar chart of marital status
> barplot(ex6_5, legend=c("Married", "Previously Married", "Never Married"), col=c("black", "blue", "red"))
>
```



```
> # Analysis Example 6.6, NCS-R DATA

> # svyby analysis gives mean of mde by sex
> (ex6_6a <- svyby(~mde, ~sex, ncsrsvyp1, svymean, se=T, na.rm=T, ci=T, keep.vars=T))
  sex    mde      se
1   1 0.1507933 0.007747811
2   2 0.2293083 0.005647255

> (ex6_6 <- svymean(~interaction (sex, mde), ncsrsvyp1, se=T, na.rm=T, ci=T, keep.vars=T))
              mean      SE
interaction(sex, mde)1.0 0.406644 0.0070
interaction(sex, mde)2.0 0.401644 0.0054
interaction(sex, mde)1.1 0.072208 0.0034
interaction(sex, mde)2.1 0.119504 0.0030
> # obtain confidence intervals
> confint(ex6_6)
              2.5 %    97.5 %
interaction(sex, mde)1.0 0.39296383 0.42032513
interaction(sex, mde)2.0 0.39113771 0.41215085
interaction(sex, mde)1.1 0.06546993 0.07894551
interaction(sex, mde)2.1 0.11356911 0.12543793

> # svyby analysis gives mean of mde by sex
> (ex6_6a <- svyby(~mde, ~sex, ncsrsvyp1, svymean, se=T, na.rm=T, ci=T, keep.vars=T))
> #sex=1=M sex=2=F
  sex    mde      se
1   1 0.1507933 0.007747811
2   2 0.2293083 0.005647255

> #svychisq provides a 2 by 2 chisq test (F)
```

```
> svychisq(~sex+mde, ncsrsvyp1, statistic="F")
```

Pearson's X²: Rao & Scott adjustment

```
data: svychisq(~sex + mde, ncsrsvyp1, statistic = "F")
F = 57.978, ndf = 1, ddf = 42, p-value = 1.947e-09
```

```
> # Analysis Example 6.7 MEAN OF MDE OVER SEX AND LINEAR COMPARISON TEST
> (ex6_7 <- svyby(~mde, ~sex, ncsrsvyp1, svymean, se=T, na.rm=T, ci=T, keep.vars=T))
      sexc      mde      se
Male   Male 0.1507933 0.007747811
Female Female 0.2293083 0.005647255
```

```
> svycontrast(ex6_7, list(avg=c(.5, .5), diff=c(1, -1)))
      contrast      SE
avg    0.190051 0.0048
diff -0.078515 0.0096
Warning message:
In vcov.svyby(stat) : Only diagonal elements of vcov() available
```

```
> confint(ex6_7)
      2.5 %      97.5 %
Male   0.1356079 0.1659788
Female 0.2182399 0.2403767
```

```
> # Bayesian Approach Example 6.7 Results on log scale
> mod.brms$stan_fit
Inference for Stan model: anon_model.
1 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=1000.
```

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
b[1]	-0.52	0.00	0.05	-0.62	-0.55	-0.52	-0.48	-0.42	523	1
Intercept	-1.44	0.00	0.03	-1.49	-1.46	-1.44	-1.42	-1.39	776	1
lprior	-2.13	0.00	0.01	-2.14	-2.13	-2.13	-2.12	-2.11	776	1
b_Intercept	-1.21	0.00	0.04	-1.28	-1.24	-1.21	-1.19	-1.14	678	1
lp__	-4492.61	0.04	0.97	-4495.23	-4493.00	-4492.34	-4491.92	-4491.67	508	1

Samples were drawn using NUTS(diag_e) at Tue Mar 11 13:08:36 2025.
For each parameter, `n_eff` is a crude measure of effective sample size,
and `Rhat` is the potential scale reduction factor on split chains (at
convergence, `Rhat=1`).

```
> # Analysis Example 6.8 Independence of MDE and Gender
> (ex6_8 <- svyby(~mde, ~sex, ncsrsvyp1, svymean, se=T, na.rm=T, ci=T, keep.vars=T))
      sex      mde      se
1    1 0.1507933 0.007747811
2    2 0.2293083 0.005647255
> #Codes for sex: 1=MALE 2=FEMALE
> #svychisq provides a 2 by 2 chisq test (F)
> svychisq(~sex+mde, ncsrsvyp1, statistic="F")
```

Pearson's X²: Rao & Scott adjustment

```
data: svychisq(~sex + mde, ncsrsvyp1, statistic = "F")
F = 57.978, ndf = 1, ddf = 42, p-value = 1.947e-09
```

```
> # Analysis Example 6.9 Independence of Education and Alcohol Dependence
```

```

> ex6_9 <- svyby (~ald,~ed4catc, subset(ncsrsvyp2, age < 29 & !is.na(ED4CAT) & !is.na(ald)), svymean, na.rm=T,
ci=T)
>
>
> # CODES FOR ED4CAT 1=0-11 2=12 3=13-15 4=16+ YEARS OF EDUCATION
> print(ex6_9)
      ed4catc      ald      se
0-11    0-11 0.09128575 0.02937999
12      12 0.04855850 0.01345971
13-15   13-15 0.04895775 0.01004206
16+     16+ 0.06903765 0.01364029

> svychisq(~ald+ ed4catc, subset(ncsrsvyp2, age < 29 & !is.na(ED4CAT) & !is.na(ald)), na.rm=T, statistic = "F")

      Pearson's X^2: Rao & Scott adjustment

data: svychisq(~ald + ed4catc, subset(ncsrsvyp2, age < 29 & !is.na(ED4CAT) & !is.na(ald)), na.rm = T,
statistic = "F")
F = 1.6498, ndf = 2.7506, ddf = 112.7754, p-value = 0.1858

> # Analysis Example 6.10 Logistic Regression MDE regressed on Gender
> (ex6_10 <- svyglm(mde~sexm, design=ncsrsvyp1, family=quasibinomial))
Stratified 1 - level Cluster Sampling design (with replacement)
With (84) clusters.
svydesign(strata = ~sestrat, id = ~seclustr, weights = ~ncsrwtsh,
data = ncsr, nest = T)

Call: svyglm(formula = mde ~ sexm, design = ncsrsvyp1, family = quasibinomial)

Coefficients:
(Intercept)      sexm
-1.2122      -0.5162
Degrees of Freedom: 9281 Total (i.e. Null); 41 Residual
Null Deviance: 9072
Residual Deviance: 8979      AIC: NA
> summary(ex6_10)

Call:
svyglm(formula = mde ~ sexm, design = ncsrsvyp1, family = quasibinomial)

Survey design:
svydesign(strata = ~sestrat, id = ~seclustr, weights = ~ncsrwtsh,
data = ncsr, nest = T)

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.21222    0.03195 -37.935 < 2e-16 ***
sexm        -0.51617    0.06820  -7.568 2.63e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for quasibinomial family taken to be 1.000108)

Number of Fisher Scoring iterations: 4
# Odds Ratio using exp
> or <- exp(-.516)
> or
[1] 0.5969034
> # Figure 6.8 Bar Chart of Marital Status in Russian Federation data
> fig6_8 <- svyby(~factor(marcatc), ~gndr, rfsvy, svymean, na.rm=T)

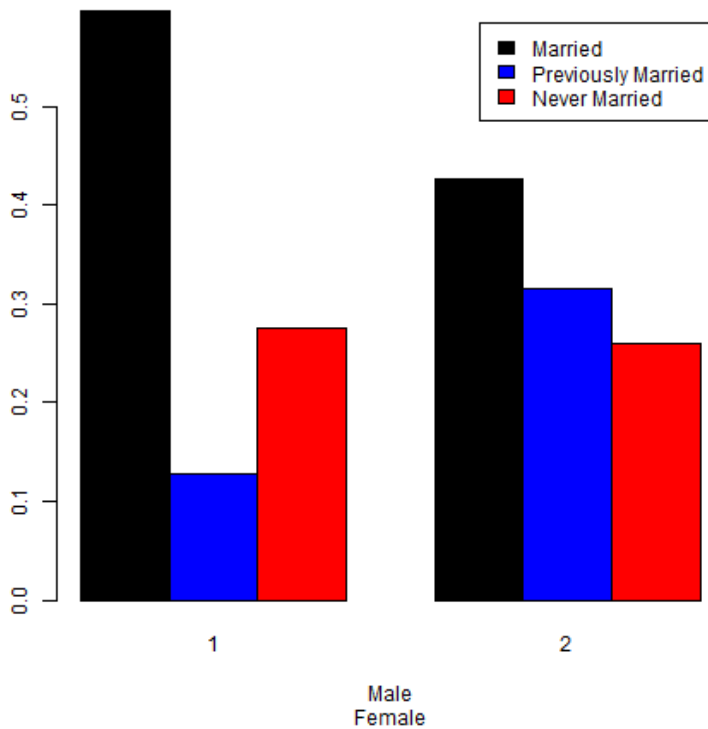
```

```

> # Bar chart of marital status by gender
> print(fig6_8)
  gndr factor(marcatc)Married factor(marcatc)Previous factor(marcatc)Never
1    1             0.5972364             0.1281528             0.2746108
2    2             0.4261892             0.3148384             0.2589724
  se.factor(marcatc)Married se.factor(marcatc)Previous se.factor(marcatc)Never
1                      0.01707498             0.01197434             0.01631450
2                      0.01632922             0.01610291             0.01722477

barplot(fig6_8, legend=c("Married", "Previously Married", "Never Married"), col=c("black", "blue", "red"),
xlab=c("Male", "Female") )

```



#NOTE: Example 6.11 not available in R (see text for example using Sudaan software)

```

> # Analysis Example 6.12 Loglinear Model Examining Relationship Between MDE and Male
> # null model run first
> null <-svyloglin(~mde+sexm,ncsrsvyp1)
> summary(null)
Loglinear model: svyloglin(~mde + sexm, ncsrsvyp1)
              coef          se          p
mde1  0.71946455 0.01573591 0.000000e+00
sexm1 0.04232084 0.01065001 7.073982e-05

> # saturated model, update null model with interaction of mde and sexm
> saturated <- update(null, ~.+mde:sexm)
> summary(saturated)
Loglinear model: update(null, ~. + mde:sexm)
              coef          se          p
mde1      0.7351533 0.01716099 0.000000e+00
sexm1      0.1228566 0.01151009 1.349733e-26
mde1:sexm1 -0.1290428 0.01705080 3.786148e-14

> # obtain F test
> svychisq(~mde+sexm, ncsrsvyp1)

          Pearson's X^2: Rao & Scott adjustment

data:  svychisq(~mde + sexm, ncsrsvyp1)
F = 57.978, ndf = 1, ddf = 42, p-value = 1.947e-09

> # obtain ChiSQ test
> svychisq(~mde+sexm, ncsrsvyp1, statistic="Chisq")

          Pearson's X^2: Rao & Scott adjustment

data:  svychisq(~mde + sexm, ncsrsvyp1, statistic = "Chisq")
X-squared = 92.15, df = 1, p-value = 2.65e-14

```