`Survey` package: Domain and poststratification estimation

Week 5

Stat 260, St. Clair

Domain estimation

- You have a `design_srs` object

```
> svyby(~y, ~domainvar, ~design_srs, svymean)  # domain means/props
> svyby(~y, ~domainvar, ~design_srs, svytotal)  # domain totals
```

response variable

domain variable

design object

function

---

svyby → stratified design object

svyby(~y, ~stratavar, ~strat, fun)

↳ SRS est. for each strat

Lohr Examples 4.7

Estimate regional means using the `agsrs` SRS data

```
> agsrs <- read.csv("http://math.carleton.edu/kstclair/data/agsrs.csv
> agsrs$n <- nrow(agsrs)
> agsrs$N <- 3078
> agsrs$wts <- agsrs$N/agsrs$n
> design_srs <- svydesign(id= ~1,
+                         fpc= ~N,
+                         weights= ~wts,
+                         data= agsrs)
```

SRS

Lohr Examples 4.7

Estimate regional means using the `agsrs` SRS data

y domain

```
> region_tot <-svyby(~acres92, ~region, design_srs, svytotal)
> region_tot   # domain estimates of total
   region   acres92       se
NC     NC 384557574 41022160
NE     NE  17722098  4490614
S       S 275091387 35287421
W       W 239556051 46090457
> confint(region_tot, df=degf(design_srs))
       2.5 %    97.5 %
NC 303828848 465286299
NE   8884885  26559311
S  205648224 344534549
W  148853274 330258829
```

Poststratification estimation

- You have a `design` object which is turned into a poststratified design object

```
pop_Nh <- data.frame(strat = c(..fill in..),
                     Nh = c(..fill..in))
design_post <- postStratify(design, ~poststrat, pop_Nh)
```

N_h known

Strat	N_h
level 1	N_1
level 2	N_2
⋮	⋮
level H	N_h

Poststratification estimation

Poststratify the `agsrs` SRS data.

```
> levels(agsrs$region)  # names of poststrata levels
NULL
> # need data frame: poststrata variable, poststrata pop. sizes
> pop_Nh <- data.frame(region=c("NC","NE","S","W"),
+                      Nh=c(1054, 220, 1382, 422))
> pop_Nh
  region   Nh
1     NC 1054
2     NE  220
3      S 1382
4      W  422
```

Poststratification estimation

Poststratify the `agsrs` SRS data.

```
> design_post<- postStratify(design_srs, strata= ~region, population=
> summary(design_post)
Independent Sampling design
postStratify(design_srs, strata = ~region, population = pop_Nh)
Probabilities:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.09242 0.09407 0.09407 0.09771 0.10152 0.10909
Population size (PSUs): 3078
Data variables:
 [1] "county"   "state"    "region"   "acres92"  "acres87"  "acres82"
 [7] "farms92"  "farms87"  "farms82"  "largef92" "largef87" "largef82
[13] "smallf92" "smallf87" "smallf82" "n"        "N"        "wts"
```

Poststratification estimation

Poststratified estimate of mean 1992 farming acres per county

```
> svymean(~acres92, design_post)  # poststratified estimate
           mean    SE
acres92 299778 17513
```