

Other statistics than the mean

Peter Lugtig

30 september, 2021

Introduction

So far, I have always asked you to compute the mean as a statistic. The mean is a very useful statistic, as many concepts in statistics can be expressed as means. For example, a regression model including variables X and Y, can be expressed as a set of link-functions that describe the relation between values of x and the conditional mean (or predicted mean) in y.

Still, you may in practice want to compute other things than the mean. The survey package includes many different functions, which I will not show all, but I will show you some functions that may be useful. The idea of this exercise is that you simply run the code first, and then change some variables to test and see what every option does.

Let's first reload our example dataset.

```
boys <- readRDS("boys.RDS")
boys <-
  boys %>%
    mutate(agecat = cut(age,
                        breaks = c(0, 1, 5, 10, 22),
                        labels = c("Younger than 1",
                                   "1 - 5",
                                   "5 - 10",
                                   "Older than 10"))))
```

#Creating a sample, and survey design object

Below I will draw a simple SRS of size 200, like you did in week 38. You can of course also use a cluster or stratified sample

```
# SRS of 200
set.seed(11)
boys$srs <- srswor(200, nrow(boys))

srs <- subset(boys, srs==1)
srs$fpc <- nrow(boys)

# now with fpc - this is correct!
srsdesign <- svydesign(id=~1, fpc=~fpc, data=srs)
svymean(~wgt, na.rm=T, design=srsdesign, deff=T) # you can use a different y-variable
```

#describing the dependent variable Now, lets ask for some statistics for the 'wgt' income.

Question 1.

- Inspect the output: do you understand what is going on?
- Can you ask for the 50th quantile (the median), or the 95th?

```
# can we also compute other things? Sure! For example, the total weight.  
# In the context of this dataset it makes no sense, but if you do a business survey,  
# the total may show the total revenue, or total production for example.  
sum(boys$wgt, na.rm=T) #  
svytotal(~wgt, na.rm=T, design=srsdesign, deff=T)  
# get estimates of weighted percentiles (here the 25th )  
svyquantile(~wgt, na.rm=T, srsdesign, quantiles = 0.25, ci = TRUE)
```

#Saving the results as an object

You can also ask for multiple means, and save these as an object. This has other benefits, because you can do calculations on this object, e.g. get a Confidence Interval

Question 2.

Run the code below, and inspect how the output looks like

```
mysvymean <- svymean(~wgt+hgt, na.rm=T, srsdesign, deff = TRUE)  
coef(mysvymean)  
SE(mysvymean)  
deff(mysvymean)  
# and get a confidence interval  
confint(mysvymean)
```

finally, bi- and multivariate relations Perhaps you are more interested in computing cross-tables or regression-coefficients. This can also be done using the functions below.

```
#we can estimate crosstables, and t-tests  
svytable(~town+agecat, design=srsdesign) # population tables  
svyby(~wgt, by=~agecat, svymean, design=srsdesign)  
  
# we can extend to do a chi.sq test  
svychisq(~town+agecat, srsdesign, statistic="adjWald")  
  
# or regression  
svyglm(wgt~hgt+age+town, design=srsdesign)  
# and we can access parts of the output  
coef(svyglm(wgt~hgt+age+town, design=srsdesign))
```

Question 3.

Can you rerun this code using a stratified or cluster sample (you may choose a design from week 40 or specify a new design), and inspect how results change?

#EOF