MT4608 Sampling Theory

Computer Practical 1

David Borchers

Septemeber 2020

In this practical we are going to use the `R` package `survey` to reproduce some of the results we have calculated manually in class and in tutorials, and then to . . .

1 Getting started with the pacakge `sampling`

The `sampling` package should already be installed in RStudio Cloud, but if you are working on another machine, you need to install it from CRAN.

We start by loading the `survey` package and the caribou dataset of Table 2.1 of the notes. The dataset is in the package `mt4608`, written specifically for this module[1].

```
library(survey) # load survey package
library(mt4608) # load this module package
```

The caribou survey data are contained in the package `mt4608`, and can be accessed as follows:

```
data(caribou) # get the dataset
caribou # look at it
help(caribou) # get a description of it
```

Calculate the total area and total number of strips in the survey region:

```
A = unique(caribou$area[caribou$stratum==1]) +
  unique(caribou$area[caribou$stratum==2])
N = unique(caribou$N[caribou$stratum==1]) +
  unique(caribou$N[caribou$stratum==2])
```

Estimate the mean number of animals per survey strip, together with a 95% confidence interval. To do this with the `survey` package, you first need to associate a survey design with the survey data, using the command `svydesign`. Look at the help for `svydesign` to understand more about the arguments; here `ids=~1` tells it that there are no clusters (you have to specify `ids`), and `fpc` if the "finite population correction" factor that we call f in the notes.

---

[1]This package is not as robust as packages on CRAN, so please be a bit patient/tolerant - things may go a little wrong in situations that I did not anticipate when writing it.

```
n = dim(caribou)[1] # sample size
srs <- svydesign(ids=~1,data=caribou,fpc=rep(n/N,n))
```

Now using the survey data and the design, we can estimate the population mean and 95% confidence interval very easily using the commands `svymean` and `confint`. To do this you have to tell `svymean` which column of the data frame is the y-variable (the response). In our case it is the `count`:

```
ybar <- svymean(~count,srs)
ybar
confint(ybar)
```

1.1   Exercise 1

1. By doing appropriate calculations with the data, decide whether the function `confint` assumed that the sample mean has a normal distribution or a t-distribution.

2. Estimate the total number of caribou in the survey region, together with a 95% confidence interval