Stats with Sparrows - 4

Julia Schroeder

September 2016

Standard errors

Standard errors (se) are a good way to display uncertainty. You can caluclate them using this equation:

$$se = \sqrt{\frac{s^2}{n}}$$

Let's do this for Tarsus in our dataset. Some basic housekeeping to start with:

```
rm(list=ls())

setwd("~/Box Sync/Teaching/IntroStats")

d<-read.table("SparrowSize.txt", header=TRUE)
d1<-subset(d, d$Tarsus!="NA")
seTarsus<-sqrt(var(d1$Tarsus)/length(d1$Tarsus))
seTarsus

## [1] 0.02096211
```

and now only for 2001:

```
d12001<-subset(d1, d1$Year==2001)
seTarsus2001<-sqrt(var(d12001$Tarsus)/length(d12001$Tarsus))
seTarsus2001

## [1] 0.1030623
```

The SE of 2001 is about four times the one for the total population. That's shocking! So, what determines how larger or small our measurement of precision, the SE is? Let's have a closer look at it:

$$se = \sqrt{\frac{s^2}{n}}$$

with a bit of Algebra we find out this is the same as

$$se = \frac{s}{\sqrt{n}}$$

The important bit in this part is to have a good look at the denominator. N represents sample size. s^2 is, as we know, the variance. Thus, is se is an indicator of the uncertainty, then what do we have to do to make this number the smallest possible? We have to increase the sample size! We can do some basic math to find out how much larger our sample size needs to be to half our standard error:

$$\frac{se}{2} = \frac{s}{2\sqrt{n}} = \frac{s}{\sqrt{4n}}$$

This is the square root law of sample size: to improve your precisision by doubling it, you need to increase your sample size by it's squared term - 4!

Another reason why I showed you these equations for the standart error is that it is not only related to variance, but also to our 95% confidence interval. Remember that? Here is the equation to calculate the 95% confidence interval. It is also a measure of precision. Can you spot the standard error?

$$CI_{95\%} = \pm 1.96 \frac{s}{\sqrt{n}}$$

or

$$CI_{95\%} = \pm 1.96 se$$

Note that this only works for large sample sizes (approx. N>50).

Excercises:

Calculate the standard error of Tarsus, Mass, Wing and Bill length of the complete population sample (as opposed to all sparrows in this world)

Note N of each.

Then, subset the dataset to only 2001 data d1<-subset(d, d$Year==2001)

Calculate SE for Tarsus, Tarsus, Mass, Wing and Bill length for the 2001 sample

Present in a table on whiteboard!

Extra points: can you calculate the 95% CI of each variable?