How to calculate confidence intervals

Consider our snow vole case study… We will calculate the confidence intervals for the mean of the weight of the individuals, for the foot length, and for different groups.

library(Rmisc)

## Warning: il pacchetto 'Rmisc' è stato creato con R versione 4.3.3

## Caricamento del pacchetto richiesto: lattice

## Caricamento del pacchetto richiesto: plyr

library(Hmisc)

##   
## Caricamento pacchetto: 'Hmisc'

## I seguenti oggetti sono mascherati da 'package:plyr':  
##   
## is.discrete, summarize

## I seguenti oggetti sono mascherati da 'package:base':  
##   
## format.pval, units

library(ggplot2)  
library(boot)

##   
## Caricamento pacchetto: 'boot'

## Il seguente oggetto è mascherato da 'package:lattice':  
##   
## melanoma

dati <- read.csv("data/captures.csv", sep=";")

# Calculating and plotting CIs with the traditional method

The function CI in the Rmisc package produces confidence intervals for the mean using the traditional method:

CI(dati$weight\_g, ci=0.95)

## upper mean lower   
## NA NA NA

So what?? How can we fix that??

CI(na.omit(dati$weight\_g), ci=0.95)

## upper mean lower   
## 31.37342 29.83193 28.29044

What about the CIs for different group of animals?

dati2 <- na.omit(dati)  
group.CI(weight\_g ~ sex,  
 data= dati2,  
 ci=0.95)

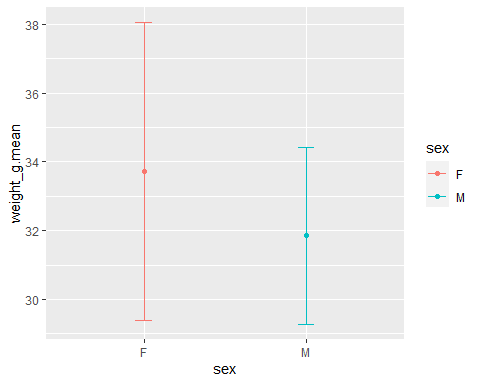
## sex weight\_g.upper weight\_g.mean weight\_g.lower  
## 1 F 38.06126 33.72727 29.39329  
## 2 M 34.41998 31.84615 29.27233

intervalli <- group.CI(weight\_g ~ sex,  
 data= dati2,  
 ci=0.95)  
intervalli

## sex weight\_g.upper weight\_g.mean weight\_g.lower  
## 1 F 38.06126 33.72727 29.39329  
## 2 M 34.41998 31.84615 29.27233

ggplot(intervalli, aes(x=sex, y=weight\_g.mean, colour=sex)) +  
 geom\_errorbar(aes(ymin=weight\_g.lower, ymax=weight\_g.upper), width=.1) +  
 geom\_line() +  
 geom\_point()

## `geom\_line()`: Each group consists of only one observation.  
## ℹ Do you need to adjust the group aesthetic?



Do it on your own:

* plot just one of the confidence intervals (for males or for females, of for the overall mean)
* plot the confidence intervals for the mean weights, grouped for age
* repeat for the footlength

as.data.frame(CI(na.omit(dati$weight\_g), ci=0.95)) -> CImean  
CImean

## CI(na.omit(dati$weight\_g), ci = 0.95)  
## upper 31.37342  
## mean 29.83193  
## lower 28.29044

# Calculating and plotting CIs with the bootstrapping method

We first need to apply the bootstrap to our data. For this purpose, we use the function boot in the boot package:

boot.data <- boot(dati2$weight\_g,  
 function(x,i) mean(x[i]),   
 R=10000)

Now we can use the bootstrapped data to calculate the CIs, using the boot.ci function:

boot.ci(boot.data,  
 conf = 0.95)

## Warning in boot.ci(boot.data, conf = 0.95): varianze bootstrap necessarie per  
## intervalli studentizzati

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS  
## Based on 10000 bootstrap replicates  
##   
## CALL :   
## boot.ci(boot.out = boot.data, conf = 0.95)  
##   
## Intervals :   
## Level Normal Basic   
## 95% (30.61, 34.79 ) (30.58, 34.83 )   
##   
## Level Percentile BCa   
## 95% (30.58, 34.83 ) (30.46, 34.71 )   
## Calculations and Intervals on Original Scale

The function generates different types of non-parametric confidence intervals. For routine use, it is recommended to use the BCa or percentile methods. The formulae on which the calculations are based can be found in Chapter 5 of Davison and Hinkley (1997). Davison, A.C. and Hinkley, D.V. (1997) Bootstrap Methods and Their Application, Chapter 5. Cambridge University Press.

Do it on your own:

* Calculate the bootstrapped intervals for the sex groups
* Plot the bootstrapped intervals

## Optional readings:

* ‘Calculating the Confidence interval for a mean using a formula’ from Statistics Learning Center. (Dr. Nic). 2013. www.youtube.com/watch?v=s4SRdaTycaw
* Davison, A.C. and Hinkley, D.V. (1997) Bootstrap Methods and Their Application, Chapter 5. Cambridge University Press.