Statistical tests

Duncan Golicher

1/12/2019

Table of Contents

# How to use the crib sheet

The crib sheets contain R code for running analyses. There is no accompanying text to explain the output nor advice on why to use the method. You should consult the course material for that.

In order to use the cribsheets you **must** first become completely familiar with the process of loading data into R memory using either read.csv for comma separated variable files or read\_excel for importing data directly from an excel spreadsheet file. You need to know how to put together your own annotated markdown files, with embedded code chunks.

For each analysis an example data set is provided that is loaded from the /home/aqm/data folder on the server. The file is converted into a data table in the cribsheet that can be exported and used as the template for your own analysis.

To use the cribsheet, first look at the format of the example data. Download this file and modify it in Excel by changing it to include your own data. Then build your own markdown file using your own data as the input. Providing you paste in chunks from the crib sheet **in the right order** you can then build an analysis for your data that will reporoduce the results. Order is important as some code chnuks are precursors to others. If you understand the logic of the analysis this will not be a proablem.

## Chi squared contingency tables

### Data formats

#### Long format

The data will originally have been collected though classifying each observation. So, if the data consists of mud cores that have been classified into two categories of substrate, mud or sand, and two categories depending whether ragworm are present or absent you will produces a csv file with the format as shown.

d<-read.csv("/home/aqm/data/HedisteCat.csv")
dt(d)



#### Tablular format

You might already have tabulated the data in Excel. Providing that the table is in the top cells of the first sheet of an Excel spreadsheet, this code will load the data.

library(readxl)
ct <-read\_excel("contingency\_table.xlsx")
dt(ct)



### Table of counts

#### Table of counts using the ct format

ct<-as.data.frame(ct)
row.names(ct) <- ct[,1]
ct<-ct[,-1]
ct<-as.table(as.matrix(ct))
ct

## Absent Present
## Mud 23 27
## Sand 44 16

ct %>% kable() %>% kable\_styling(bootstrap\_options = "striped", full\_width = F, position = "left")

Absent

Present

Mud

23

27

Sand

44

16

#### Table of counts using the data frame format

ct<-table(d)
ct

## Cat
## Substrate Absent Present
## Mud 23 27
## Sand 44 16

ct %>% kable() %>% kable\_styling(bootstrap\_options = "striped", full\_width = F, position = "left")

Absent

Present

Mud

23

27

Sand

44

16

### Table of Proportions

#### Table of proportions

pt<-round(prop.table(ct) \*100,1)
pt

## Cat
## Substrate Absent Present
## Mud 20.9 24.5
## Sand 40.0 14.5

pt %>% kable() %>% kable\_styling(bootstrap\_options = "striped", full\_width = F, position = "left")

Absent

Present

Mud

20.9

24.5

Sand

40.0

14.5

#### Table of proportions for rows

ptr<-round(prop.table(ct,margin=1) \*100,1)
ptr

## Cat
## Substrate Absent Present
## Mud 46.0 54.0
## Sand 73.3 26.7

ptr %>% kable() %>% kable\_styling(bootstrap\_options = "striped", full\_width = F, position = "left")

Absent

Present

Mud

46.0

54.0

Sand

73.3

26.7

#### Table of proportions for columns

ptc<-round(prop.table(table(d),margin=2) \*100,1)
ptc

## Cat
## Substrate Absent Present
## Mud 34.3 62.8
## Sand 65.7 37.2

ptc %>% kable() %>% kable\_styling(bootstrap\_options = "striped", full\_width = F, position = "left")

Absent

Present

Mud

34.3

62.8

Sand

65.7

37.2