

Statistical methods in genetic relatedness and pedigree analysis

Oslo, January 2020

Magnus Dehli Vigeland and Thore Egeland

Exercise set II. Introduction to R and the *ped suite* packages

Before you start

The aim of these exercises is to introduce you to the *ped suite*, which is a collection of R packages for pedigree analysis. For now we will only need two of the packages, `pedtools` and `pedprobr`. The first of these is the core of the *ped suite*, containing all the basic utilities for handling and plotting pedigrees. The `pedprobr` package is more specialised, focused on computing pedigree likelihoods.

Some warm-up exercises in base R are included, which you can skip if you don't need them.

To get started, open Rstudio and load `pedtools` and `pedprobr` by writing

```
library(pedtools)
library(pedprobr)
```

If this results in an error like `there is no package called pedtools`, you must first install the packages:

```
install.packages("pedtools")
install.packages("pedprobr")
```

(Don't forget the quotes.) Ask for help if you run into trouble!

Finally, if you enjoy working with the `pedtools` package, I recommend glancing through the user manual (called a *vignette* in the R world). This gives a more detailed overview of the utilities offered by this package, many of which are not covered in this course. You can find the vignette on the CRAN page of `pedtools`, or access it in R with the command

```
vignette("pedtools")
```

Warm-up: Basic R

If you are an experienced R user, feel free to skip this section.

Exercise II-1

Guess the outcome of the following R commands. Execute the commands to see if you were correct.

a) `2 - 4 / 2 + 3 ^ 2`

b) `3:1 + 1 + c(0, 0, 10)`

c) `a = 1:4`
`a * (a + 1)`

d) `x = 1:10`
`x[6:10] = -9`
`x`

e) `a1 = 3.14`
`a2 = seq(10, 22, by = 5)`

```
c(a1, a2, 0)
```

```
f) y = list(my_numbers = 1:3, my_strings = "foo")
y$my_numbers = c(y$my_numbers, 4:6)
y$my_strings = c(y$my_strings, "bar")
y$my_bool = c(T, F)
y
```

Exercise II-2 (Basic R plots)

- a) Use the following code to plot the graph of $y = x^3 - x$ for x between -3 and 3

```
x = seq(-3, 3, length = 100)
y = x^3 - x
plot(x, y, type = "l", main = "My graph", xlab = "x", ylab = "y")
```

- b) Plot the graph of $y = 1/x + 10\sqrt{x}$ for x between 0 and 3 . How do you deal with the 0 ?

Pedigrees and likelihoods

Exercise II-3 (Creating pedigrees)

This exercise displays some common methods of creating pedigrees in `pedtools`. I recommend running `plot(x)` after each line of code, to get a feeling of what's going on.

- a) Use the code below to create the pedigree in Figure 1. Plot the pedigree after each line.

```
x = nuclearPed(nch = 2, sex = 2:1)
x = addSon(x, 3)
x = addDaughter(x, 4)
```

- b) Use the code below to create the pedigree in Figure 2. Plot the pedigree after each line.

```
x = nuclearPed(nch = 1)
x = addParents(x, id = 1)
x = addParents(x, id = 2)
x = relabel(x, old = c(4, 5, 6, 7, 1, 2, 3), new = 1:7)
```

- c) Use the code below to create the pedigree in Figure 3. Plot the pedigree after each line.

```
x = nuclearPed(father = 1, mother = 2, child = 4)
x = addChildren(x, father = 1, mother = 3, nch = 2, sex = 2:1)
x = addChildren(x, father = 4, mother = 5, nch = 1)
```

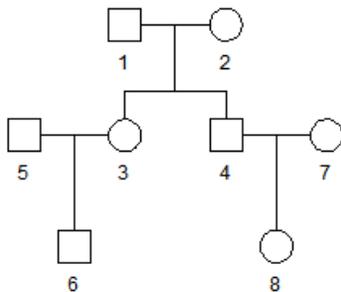


Figure 1

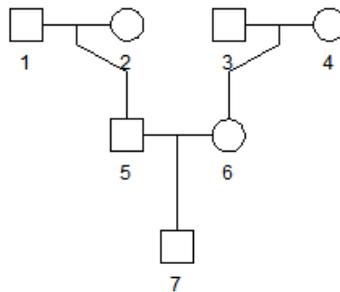


Figure 2

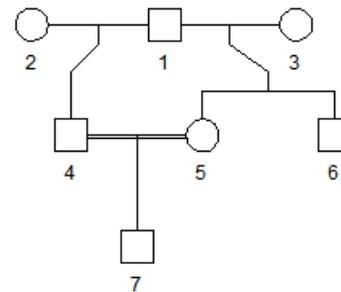


Figure 3

Exercise II-4 (Pedigree shortcuts)

Several shortcuts for common pedigree structures are available in `pedtools`. In many cases one of these will give you the pedigree you want in a single command, possibly save a few details to be sorted out at the end.

- a) Check that the following code produces the pedigree in Figure 1.

```
x = cousinPed(degree = 1)
x = swapSex(x, c(3, 8))
```

- b) Check that the following code produces the pedigree in Figure 2.

```
x = ancestralPed(g = 2)
```

- c) Check that the following code produces the pedigree in Figure 3.

```
x = halfSibPed(nch1 = 1, nch = 2, sex1 = 1, sex2 = 2:1)
x = addChildren(x, father = 4, mother = 5, nch = 1)
```

Exercise II-5 (Marker genotypes and other plotting options)

- a) Use the code below to create the pedigree in Figure 4.

```
x = nuclearPed(3)
m = marker(x, alleles = c("A", "B"))
genotype(m, 1) = c("A", "A")
genotype(m, 3:5) = c("A", "B")
plot(x, marker = m)
```

- b) Keeping `x` and `m` as above, execute the following plotting commands and study the results. Experiment with the different options on your own. The help page for pedigree plotting contains additional information: type `?plot.ped` to open it.

```
plot(x, id.labels = NULL)
plot(x, title = "MY PED")
plot(x, shaded = c(1,5), deceased = 3)
plot(x, col = list(red = 3:4, blue = 5))
plot(x, marker = m, skip.empty.genotypes = TRUE)
plot(x, marker = m, sep = " | ", missing = "0")
```

- c) Try to recreate exactly the pedigree plot in Figure 5 (use `x` and `m` as above).

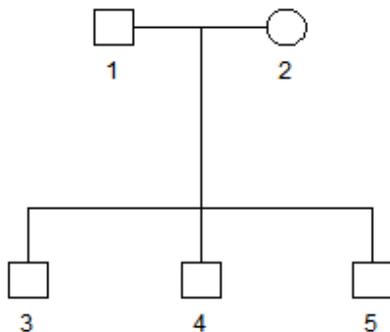


Figure 4

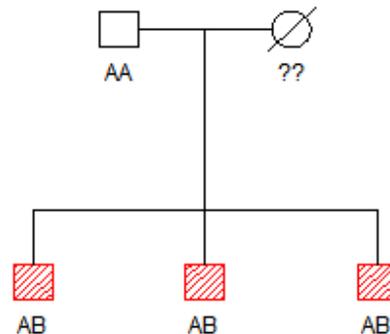


Figure 5

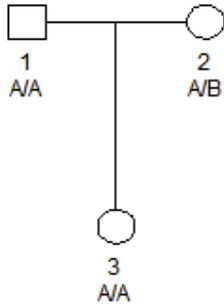


Figure 6

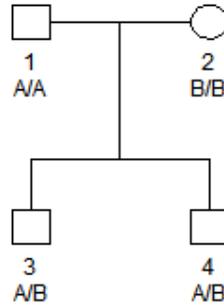


Figure 7

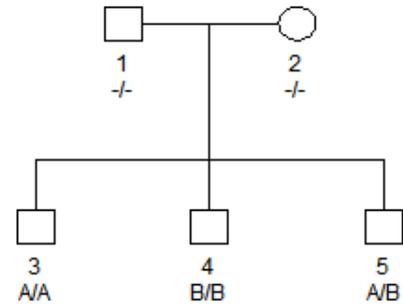


Figure 8

Exercise II-6

Figures 6, 7 and 8 show pedigrees with genotypes for an autosomal SNP in HWE, with allele frequencies $P(A) = 0.9$ and $P(B) = 0.1$.

- a) Use the following code to compute the likelihood of the pedigree in Figure 6.

```
x = nuclearPed(1, sex = 2)
m = marker(x, afreq = c(A = 0.9, B = 0.1))
genotype(m, c(1, 3)) = c("A", "A")
genotype(m, 2) = c("A", "B")
plot(x, m)
likelihood(x, m)
```

- b) Compute the likelihood of the pedigree in Figure 7.
 c) Compute the likelihood of the pedigree in Figure 8.

Exercise II-7 (Creating and loading a pedigree file)

- a) Create a pedigree file describing the family in Figure 3, and save it as “figure3.ped” somewhere on your computer. I recommend using a simple text editor (e.g. notepad) for this task. If you use Excel, make sure you save the file as plain tab-delimited text (xlsx will not work).
 b) Load the pedigree into R using the `readPed()` function. If you included headers in the ped file (e.g. ID, FID, MID, and SEX) the following should work without further options:

```
x = readPed("path/to/your/folder/figure3.ped")
```

Without headers you have to indicate the position of the essential columns, i.e.,

```
x = readPed("path/to/your/folder/figure3.ped", id = 1, fid = 2, mid = 3, sex = 4)
```

- c) Plot the pedigree and check that it looks like Figure 3.

Exercise II-8

The shaded male in Figure 9 has hemophilia A, an X-linked recessive disease. None of the other family members are affected, but the female cousin (8) is worried that she might be carrying the disease allele. We assume that the disease allele D has frequency 0.001 in the population, and that HWE holds. We also assume that no mutations at the disease locus arose within the pedigree.

- a) Write in all the genotypes you can deduce from the given information.
 b) What is the probability that individual 8 is a carrier of the disease allele? Were the assumptions about frequency and HWE needed to answer this?

- c) The information about the grandfather is limited, and it is possible that he may have had the disease. Explain how this complicates the computation of the carrier probability.
- d) The `oneMarkerDistribution` of `pedprobr` can be used to compute such probabilities. For example, the following code is a good starting point:

```
ped = cousinPed(1)
ped = swapSex(ped, c(3, 6, 8))
dis = marker(ped, "5" = "D", afreq = c(D = 0.001, N = 0.999), chrom = "X")
plot(ped, marker = dis, shaded = 5)

oneMarkerDistribution(ped, id = 8, partialmarker = dis)
```

What is the reported probability that she is a carrier? Discuss the output and possible weaknesses of this calculation.

