ISFG summer school – virtual edition 2023

Pedigree analysis: Basic

Magnus Dehli Vigeland and Thore Egeland

Exercise set II. Kinship testing

Before you start, run the following commands to download some of the datasets used in this course.

```
url = "https://magnusdv.github.io/pedinr/datasets/data.zip"
download.file(url, destfile = "data.zip")
unzip("data.zip")
```

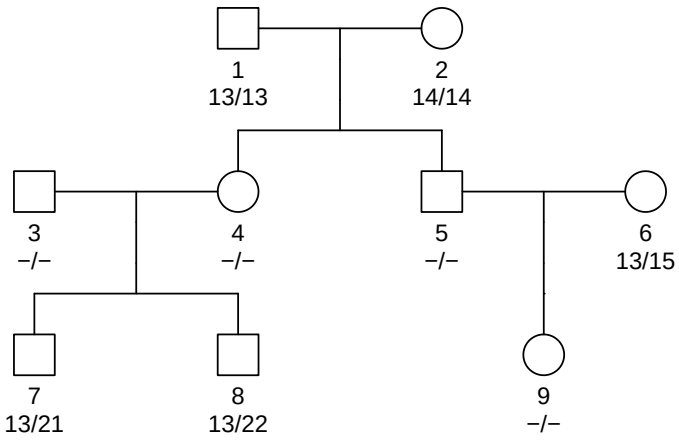
Note: The files are put in a subfolder `data` in the current working directory. You can identify this by with the command `getwd()`.

You should also load the **pedsuite** packages:

```
library(pedsuite)
```

Exercise II-1

Consider the following pedigree, in which some members have been typed with a single DNA marker.



a) What kind of marker is this: SNP or STR? Autosomal or X-linked? How do you know?
b) How many different alleles are observed in the family. What do the allele labels (e.g. `13`) mean?
c) What are the genotypes of individuals 4 and 5?
d) Can you determine the genotype of individual 3?
e) What are the possible genotypes for individual 9, and how likely is each of them?

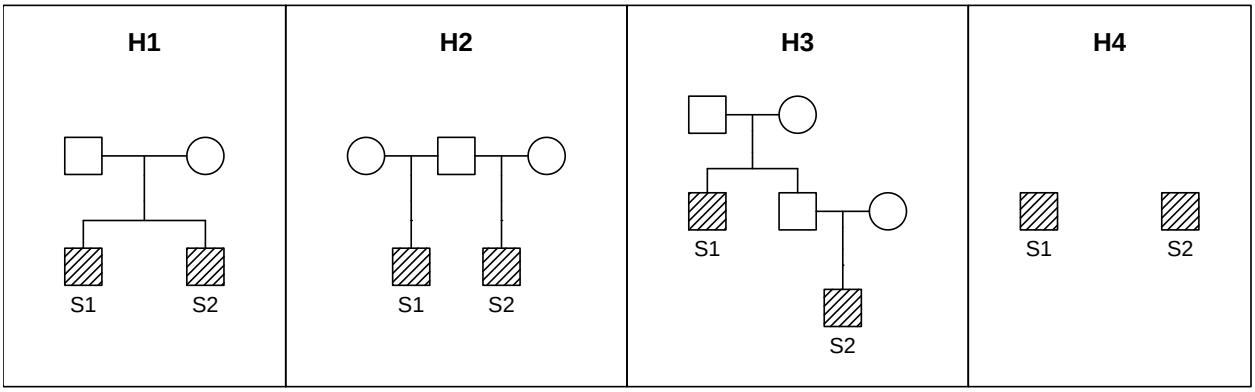
Exercise II-2

In a paternity case, the alleged father has genotype a/b for a certain marker, while the child has genotype a/c. The mother is not genotyped.

a) Formulate the natural hypotheses H1 and H2, and create the corresponding pedigree objects in R.
b) Use `plotPedList()` to plot H1 and H2.
c) Use `kinshipLR()` to compute the LR if the allele frequencies are $p_a = 0.01$, $p_b = 0.3$ and $p_c = 0.69$.
d) (Optional) Find a general formula for LR expressed by the allele frequency p_a.

Exercise II-3

In this exercise we will analyse the relationship between two males, S1 and S2, who are genotyped with 15 STR markers. The hypothesised pedigrees are shown below.



a) Define the first three pedigrees by running:

```
ids = c("S1", "S2")
H1 = nuclearPed(children = ids)
H2 = relabel(halfSibPed(), old = 4:5, new = ids)
H3 = relabel(avuncularPed(), old = c(3,6), new = ids)
```

b) Load the marker data and check that the result contains S1 and S2 as singletons.

```
H4 = readPed("data/kinship-riddle.ped")
H4 = setFreqDatabase(H4, database = NorwegianFrequencies)
summary(H4)
```

c) Find the LRs using H4 as the reference.

d) Notice that H2 and H3 give identical LRs. Do you think this is a coincidence? Explain!

e) Include a hypothesis specifying that S1 and S2 are first cousins, and find the LR against H4.
Hint:

```
H5 = relabel(cousinPed(1), old = 7:8, new = ids)
```

Exercise II-4

This exercise demonstrates how you can continue projects in Windows `Familias` (freely available from https://familias.no/) in R. Once the `Familias` data, i.e., the `.fam` file, has been converted to pedigree objects in R, we have access to all the functionality of the pedsuite.

A child (CH) and the alleged father (AF) are genotyped with 21 STR markers. The mother is not disputed. We consider the following two hypotheses:

- H1: The alleged father is the biological father.
- H2: The alleged father and the child are unrelated.

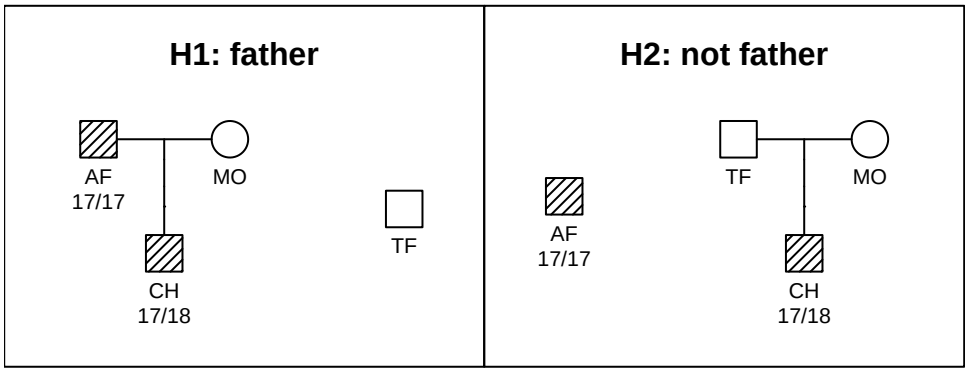
a) Use the `readFam()` function to download the Familias file and convert to R.

```
dat = readFam("http://familias.name/norbisRelatedness/paternityCase.fam")
```

b) The next step is to understand the output by `readFam()`. Run the following commands to explore the `dat` object.

```
names(dat)
summary(dat)
plotPedList(dat)
```

c) Produce the plot below as accurately as possible, including hatched symbols for the typed members, and genotypes for the first marker.



d) To simplify matters below, extract the two hypotheses as separate objects, `H1` and `H2`:

```
H1 = dat[[1]]
H2 = dat[[2]]
```

e) Show that LR comparing H1 to H2 is 0, by running:

```
res = kinshipLR(H1, H2)
res
```

f) Find the marker with LR = 0.

g) Calculate the LR once more, after first removing the marker giving LR = 0. (Note: This is a practice we advise strongly against!)

Rather than removing incompatible markers, we introduce a mutation model. The possible mutation models include `custom`, `equal`, `proportional`, `stepwise` and `onestep` and are described in the documentation of `pedmut::mutationModel()`. Different models can be used for females and males. Note that `custom` is completely general as you can define the mutation matrix. Below we use the `proportional` and `equal` models.

h) Apply a proportional model to the PENTA_E marker and recalculate LR:

```
H2 = setMutmod(H2, marker = "PENTA_E", model = "proportional", rate = 0.00001)
lr = kinshipLR(H1, H2, source = 2)
```

What is the total LR now? What is the new LR for PENTA_E?

i) What is the total LR when the `proportional` model with `rate = 0.00001` is used for *all* markers?

j) What is the total LR when the `equal` model with `rate = 0.00001` is used for *all* markers?

k) (Optional.) Run the below commands and explain what it does. Comment on the resulting plot.

```
mutrate = 10^(-c(1:6))

LR = sapply(mutrate, function(r) {
  H2 = setMutmod(H2, model = "equal", rate = r)
  kinshipLR(H1, H2, source = 2)$LRtotal[1]
})

plot(mutrate, LR, type = "b", log = "xy", main = "LR with mutation model")
```