# ISFG summer school - virtual edition 2021

# Pedigree analysis in R

Magnus Dehli Vigeland and Thore Egeland

## Exercise set III. Kinship testing

### Exercise III-1

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.
b) Use `forrel` to compute the LR for this marker, if the allele frequencies are $p_a = 0.01, p_b = 0.3$ and $p_c = 0.69$.
c) Plot H1 and H2.
d) (Optional.) Find a general formula for LR expressed by the allele frequency $p_a$.

### Exercise III-2

If you haven't done it already, run the below commands to download various datasets used in this course.

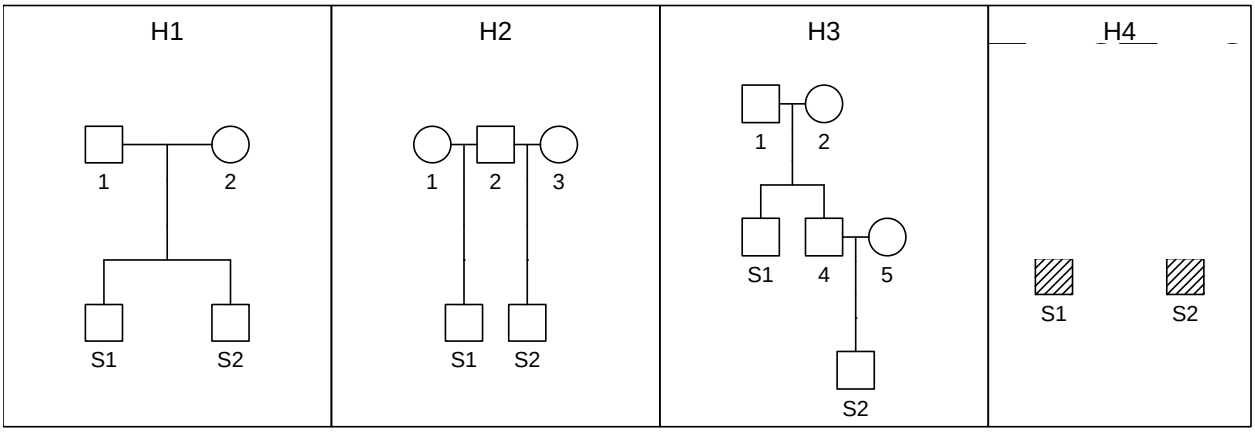
```
url = "https://magnusdv.github.io/pedinr/datasets/data.zip"
download.file(url, destfile = "data.zip")
unzip("data.zip")
```

*Comment*: The files are put in the subfolder `data` of the working directory of you current R session, which you can find by running:

```
getwd()
```

### Exercise III-3

In the first part of this exercise you are asked to reproduce the code of the "Relationship Riddle" example of Lecture 3. The 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(cousinPed(deg = 0, rem = 1),
             old = c(3,6), new = ids)
```

b) Read marker data, add pedigree for H4 and assign database by running:

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

c) Find the LRs when H4 is the reference.

d) Include a hypothesis specifying that S1 and S2 are first cousins and find the LRs. Use H4 as the reference.
*Hint*:

```
H5 = relabel(cousinPed(1), old = 7:8, new = c("S1", "S2"))
```

**Exercise III-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 `ped suite`. The function `forrel::readFam` used for conversion from `Familias` to R.

We consider the following two hypotheses:

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

The alleged father and the child are genotyped. The mother is not disputed. There are 21 markers.

a) Run the commands to read the data:

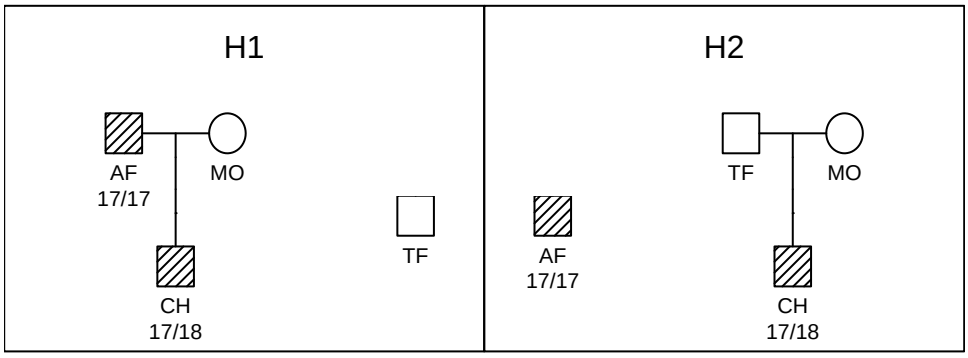
```
library(forrel)
dat = readFam("http://familias.name/norbisRelatedness/paternityCase.fam")
```

The next step is to understand the structure of the converted data returned by `readFam()`, i.e., `dat` object above. In this case the `.fam` file includes data for both hypotheses, which are converted into a list.

b) Extract data for H1 and H2 by running:

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

c) Produce the below plot, where the genotypes are for the first marker.



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

```
res = kinshipLR(H1, H2, ref = 2)
```

e) Find the marker with LR = 0.

f) Calculate the LR once more, but now with the marker giving 0 likelihood ratio removed (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.

g) Run the below commands:

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

What is the total LR and LR for the marker "PENTA_E"?

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

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

j) Run the below commands and explain what is done:

```
rvec = 10^{-c(1:6)}

lrs = sapply(rvec, function(r) {
  H2 = setMutationModel(H2, model = "equal", rate = r)
  kinshipLR(H1, H2, source = 2)$LRtotal[1]
})

plot(log10(rvec), log10(lrs))
```

k) Plot once more, but now with axes labels and title added.