

ISFG summer school - virtual edition 2021

Pedigree analysis in R

Thore Egeland and Magnus Dehli Vigeland

Exercise set VI. Introduction to pedigrees, genetics and R

Load `pedsuite` and `dvir`:

```
library(pedsuite)
library(dvir)
```

If you haven't downloaded the datasets already, do so by running

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

Exercise VI-1 (Missing person case)

In this exercise we will go through some of the analysis in the “Case Study” of Lecture 6.

a) Read the data by running

```
mpi = readPed("data/mpi-example.ped")
mpi = setFreqDatabase(mpi, "data/mpi-example.freq")
```

b) Inspect the data by running

```
summary(mpi)
```

c) Make a missing person plot by running

```
ref = mpi$Reference
missingPersonPlot(ref, missing = "MP")
```

d) We would like to test if POI1 is the missing person. Formulate the natural hypotheses and find the LR by running

```
poi1 = mpi$POI1
mpiTest1 = missingPersonLR(ref, missing = "MP", poi = poi1)
mpiTest1
```

Give a conclusion.

e) Produce a plot of the LR for each marker by running

```
lr1 = mpiTest1$LRperMarker
cols = ifelse(lr1 > 1, 8, 2)
barplot(lr1, col = cols, ylab = "LR", las = 2, cex.names = 0.8)
abline(h = 1, lty = 2)
```

Give the names of the markers with LR = 0.

f) Find the LR for POI2. Which marker gives the largest LR? Extract the allele frequencies of the marker with the largest LR, and explain why this marker gives a high LR for POI2, but not for POI1.

g) Find the exclusion power by running

```
ep = missingPersonEP(ref, missing = "MP")
ep
```

Interpret the output.

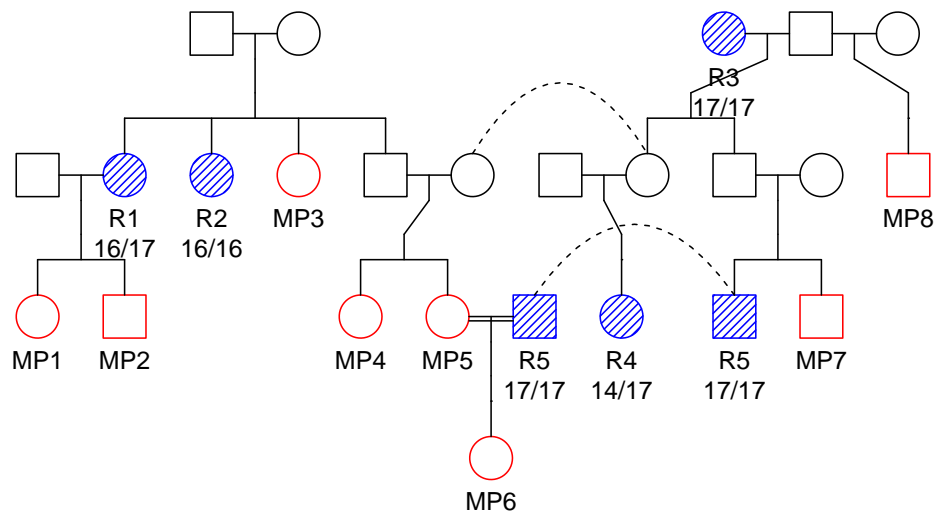
- h) Find the inclusion power by running

```
ip = missingPersonIP(ref, missing = "MP", nsim = 1000, threshold = 10000, seed = 17)
ip
```

Interpret the output.

Exercise VI-2 (DVI analysis)

We will analyse the DVI dataset `grave` included in the `dvir` package. Here is the reference pedigree:



- a) Show that the total number of a priori possible solutions is 52,564, as follows:

```
ncomb(5, 5, 3, 3)
```

Explain the input to the function.

- b) To save typing later on, extract the three components of the `grave` dataset:

```
pm = grave$pm           # The list of missing persons
am = grave$am           # The reference family pedigree
missing = grave$missing # The names of the missing persons
```

Now try to reproduce the plot of the reference family above. *Hint:* Here is a good start:

```
refs = typedMembers(am)
plot(am, labs = c(refs, missing), col = list(red = missing, blue = refs))
```

- c) Describe the relationship between the parents of MP6. Does the pedigree contain inbreeding? Find the inbreeding coefficient of MP6.
- d) Use `pairwiseLR()` to compute the LR matrix consisting of individual likelihood ratios $LR_{i,j}$ comparing the assignment $V_i = M_j$ to the null hypothesis of no identification. *Hint:* Check `?pairwiseLR`.
- e) Use `jointDVI()` to find the optimal (joint) solution, and inspect the top five alternatives. Comment on your findings. *Hint:* Check `?jointDVI`.