

ISFG summer school - virtual edition 2021

Pedigree analysis in R

Magnus Dehli Vigeland and Thore Egeland

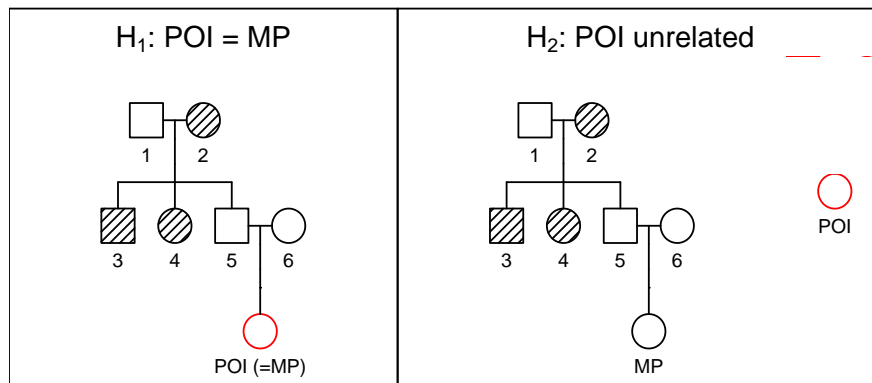
Solutions for exercise set VI

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

Exercise VI-1

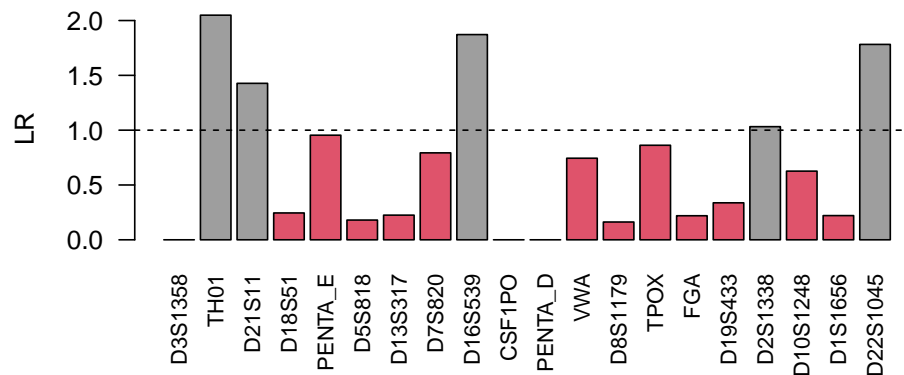
Some answers are in the exercise; missing answers are added below.

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



- d) The commands gives $\text{LR} = 0$, which excludes POI1 as the missing person (adding the usual caveats).
e) The markers with $\text{LR} = 0$ are D3S1358, CSF1PO and PENTA_D, as can be seen in the plot below.

```
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)
```



f) Using similar code as for POI1, we find:

```
poi2 = mpi$POI2
mpiTest2 = missingPersonLR(ref, missing = "MP", poi = poi2)
mpiTest2
```

```
##      H1:H2
## 2369.091
```

The LR_s for the individual markers are found in the LR_{perMarker} slot of the output. We sort them and round them to 2 decimal places:

```
round(sort(mpiTest2$LRperMarker), 2)
```

```
## D8S1179 D13S317 D7S820 D22S1045 TPOX D16S539 D5S818 D3S1358
##      0.16      0.65      0.79      0.90      1.00      1.06      1.24      1.32
## PENTA_D      VWA D21S11 D10S1248 FGA D1S1656 TH01 D19S433
##      1.34      1.35      1.43      1.51      1.64      1.86      2.05      2.22
## PENTA_E D2S1338 CSF1PO D18S51
##      2.33      2.91      3.25      15.22
```

The marker D18S51 gives the largest LR, with LR = 15.22. The allele frequencies of this marker are (rounded to 5 decimal places):

```
round(afreq(mpi, "D18S51"), 5)
```

```
##      9      10      11      12      12.2      13      14      15      16      16.2
## 0.00042 0.01242 0.00835 0.13489 0.00005 0.12680 0.18728 0.14246 0.11626 0.00010
##      17      18      18.1      18.3      19      20      21      22      23      24
## 0.10092 0.07697 0.00010 0.00005 0.05343 0.02066 0.00746 0.00595 0.00235 0.00141
##      25      26      27
## 0.00104 0.00031 0.00031
```

The allele 14 (frequency 0.19) in POI2 is observed several times in the reference pedigree, while the alleles of POI1 do not occur in the reference. This explains the difference in the LR_s for this particular marker:

```
c(LR1 = mpiTest1$LRperMarker["D18S51"], LR2 = mpiTest2$LRperMarker["D18S51"])
```

```
## LR1.D18S51 LR2.D18S51
## 0.2449387 15.2168009
```

g) The exclusion power calculation gives the following output:

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

```
## Potential mismatches: 8 (D3S1358, D7S820, CSF1PO, PENTA_D, VWA, TPOX, D19S433, D2S1338)
## Expected mismatches: 1.679
## P(at least 1 mismatch): 0.863
```

We conclude that the overall probability of exclusion is EP = 0.86, which is normally considered to be sufficient.

h) We estimate the inclusion power as suggested in the exercises:

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

```
## Mean LR: 46998.29
## Mean log10(LR): 3.223
```

```
## Estimated power:
## P(LR >= 10000) = 0.282
```

The output shows that LR exceeds 10000 with probability 0.282. *Comment:* The answer depends to some extent on the seed, but with as many as 1000 simulations the numbers will not change too much.

Exercise VI-2

a) The total *a priori* number of solutions is `ncomb(5,5,3,3)`, where the numbers are, respectively:

- the number of female victims
- the number of female missing persons
- the number of male victims
- the number of male missing persons

b) The plot was produced with the following command:

```
refs = typedMembers(am)
plot(am, marker = 1, hatched = refs, labs = c(refs, missing),
     col = list(red = missing, blue = refs))
```

c) The parents of MP6 are first cousins, hence he has inbreeding coefficient $f = 1/16$. As an aside, we mention that there is an experimental package, **verbalisr**, that can help to describe relationships in confusing pedigrees. To try it, first install it from GitHub:

```
install.packages("devtools")
devtools::install_github("magnusdv/verbalisr")
```

Then we can apply it to the relationship between MP5 and R5 (the parents of MP6) as follows:

```
library(verbalisr)
verbalise(am, ids = c("MP5", "R5"))
```

```
## first cousins (common ancestors: GF2, R3)
```

d) The pairwise LR matrix is found as follows:

```
prw = pairwiseLR(pm, am, missing)
prw$LRmatrix
```

```
##          MP1          MP2          MP3          MP4          MP5          MP6
## V1 479971259          0 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
## V2          0 67760107189 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
## V3          0          0 6.409841e+14 0.000000e+00 0.000000e+00 0.000000e+00
## V4          0          0 0.000000e+00 1.803600e+12 1.803600e+12 0.000000e+00
## V5          0          0 0.000000e+00 1.030067e+11 1.030067e+11 0.000000e+00
## V6          0          0 0.000000e+00 0.000000e+00 0.000000e+00 8.817392e+12
## V7          0          0 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
## V8          0          0 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##          MP7          MP8
## V1          0 0.0000000
## V2          0 0.5512209
## V3          0 0.0000000
## V4          0 0.0000000
## V5          0 0.0000000
## V6          0 0.0000000
## V7 16946051 295.8389523
## V8          0 0.2684890
```

e) The most likely joint solutions are found as follows:

```
res = jointDVI(pm, am, missing)
head(res)
```

```
##      V1 V2 V3 V4 V5 V6 V7 V8      loglik          LR      posterior
## 1 MP1 MP2 MP3 MP4 MP5 MP6 MP7 * -737.0038 1.374125e+90 6.904732e-01
## 2 MP1 MP2 MP3 MP4 MP5 MP6 MP7 MP8 -737.8061 6.159953e+89 3.095266e-01
## 3 MP1 MP2 MP3 MP4 MP5 MP6 * * -752.3418 2.997856e+83 1.506369e-07
## 4 MP1 MP2 MP3 MP4 MP5 MP6 * MP8 -753.3430 1.101543e+83 5.535057e-08
## 5 MP1 MP2 MP3 * MP5 MP6 MP7 MP8 -773.8441 1.375599e+74 6.912137e-17
## 6 MP1 MP2 MP3 * MP5 MP6 MP7 * -774.8047 5.263521e+73 2.644825e-17
```

Comment: The optimal solution is $V_i = M_i, i = 1, \dots, 7$. The LR comparing this solution to the one where $V_8 = MP_8$ is $13.74/6.160 = 2.2$. Hence, it is not possible to reliably distinguish these solutions. This is not unexpected as there is scarce evidence in the pedigree for identifying MP8. The remaining solutions, from number 3 and onwards, are much less likely. In fact, the posterior for the two top solutions, calculated using a flat prior for each of the 52564 possible assignments, i.e., $1/52674$, is close to 1.