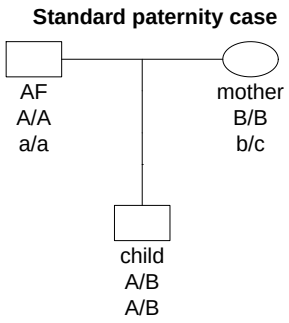

Figure 1: A standard paternity case with genotypes for two markers marker used in Exercise 1.

Familias 3 exercises. Independent markers

1. The purpose of this exercise is to consider the simplest possible paternity case. Figure 1 shows a mother (undisputed), an alleged father (AF) and a child.

   We consider

   H_1**:** The alleged father (AF) is the real father.

   H_2**:** The alleged father and the child are unrelated.

   The mother is undisputed.

   a) Consider first only one one autosomal locus, called S1, with alleles A, B and C, see Figure 1. The allele frequencies are $p_A = p_B = 0.05$ and $p_C = 0.9$. Explain why the likelihood ratio is $LR = 1/p_A$. How do you interpret the LR?

   b) Calculate the LR using Familias.

   c) There is a second autosomal locus, called S2, with alleles a, b, c and d with allele frequencies 0.1, 0.1, 0.1 and 0.7, respectively. See Figure 1. Calculate the LR for this marker by hand. the likelihood ratio for the two first markers using Familias.

Allele	Frequency
14	0.072
15	0.082
16	0.212
17	0.292
18	0.222
19	0.097
20	0.02
21	0.003

Table 1: Allele frequencies for 2

d) It can be shown that the likelihood ratio for two first markers is $(1/p_A)(1/p_B)$. Use this to verify the Familias answer.

e) Generate a report using the `Save` results with the `Both` option. The report includes all input and all output. Check that the report file is correct. In particular check that the LR for markers S1 and S2 as well as the combined likelihood ratio is correct.

f) Save the Familias file (we suggest the file extension fam). Exit Familias.

g) Start Familias and read the previously saved file. Calculate 1/RMP for AF.

h) We next consider theta (θ) correction. For simplicity we will only use the first marker, S1. The θ parameter is called kinship parameter in Familias and is set using the `Options` button in Familias. Set the kinship parameter to 0.02. Calculate the LR for the first marker S1. To get calculations for selected markers only, in this case S1, use the `Systems included` button. Check that your answer coincide with the following theoretical result

$$LR = \frac{1 + 3\theta}{2\theta + (1 - \theta)p_A}. \tag{1}$$

i) Discuss the assumptions underlying the calculations of this exercise.

2. We consider a motherless paternity case, see Figure 2, with one marker, VWA:

   The case data and hypotheses are:

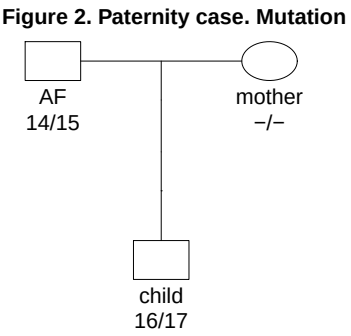

Figure 2: A paternity case with possible mutation.

- Alleged father: 14, 15
- Child: 16, 17
- H_1: Alleged father is the true father.
- H_2: Alleged father and the child are unrelated.

a) Explain why $LR = 0$. Confirm this answer using Familias.

b) Use the mutation model Equal probability with mutation rate $R = 0.007$ for both males females and calculate LR.

c) Let. It can be shown (see Exercise 8) that

$$LR = \frac{m(p_{16} + p_{17})}{2p_{16}p_{17}}. \tag{2}$$

Use this formula to confirm the Familias calculation.

3. The College of American Pathologists (CAP) has several proficiency testing programs targeted to laboratories that perform DNA typing of STR loci. The below is a test from 2011: Hikers come across human skeletal remains in a forest. Evidence around the site provides a clue as to the identity of the individual. You are asked to test a bone to determine if the individual (bone) is related to an alleged mother (AM) and the mother's other daughter, the alleged full sister, (AS) see Figure 3. Consider the following two hypotheses and problems

H_1 : The individual providing the bone is the daughter of AM and sister of AS.

H_2 : The individual providing the bone is not related to the tested individuals (AM,AS) of Figure 3.

a) Enter the data for the marker shown in Figure 3 manually into Familias and calculate the LR.

b) Read input from the file ExS3.fam. Calculate the LR based on all markers.

c) Use the output from Familias to find the LRs of the individual markers. Check that the answer for F13B corresponds to the one you found in problem a) above. Calculate LR for all markers.

d One of the markers, D7S820, gives a very large LR, namely 11189. What do you think is the reason for this large LR? What is the combined LR if marker D7S820 is removed?

Allele frequencies for F13B are: 14:0.122, 15:0.258, 17:0.197, rest:0.423.

4. Two individuals, GF and GS, are submitted to the laboratory for testing. We consider the hypotheses

   H_1 : GF is the grandfather of GS.

   H_2 : The individuals GF and GS are unrelated.

   Figure 4 shows the pedigree corresponding to H_1 for the first marker (D3S1358).

   a) Enter the data manually into Familias and calculate the LR for the first marker shown in Figure 4. The allele frequencies are 14: 0.122; 15:0.258;17:0.197

   b) Calculate the LR based on all markers. Read input from the file ExS4.fam.

   c) For the discussion: Formulate a conclusion. In the CAP exercise it was stated that GF and GS share the same Y-haplotype and that the frequency of this haplotype is 0.0025. Can this information be used?

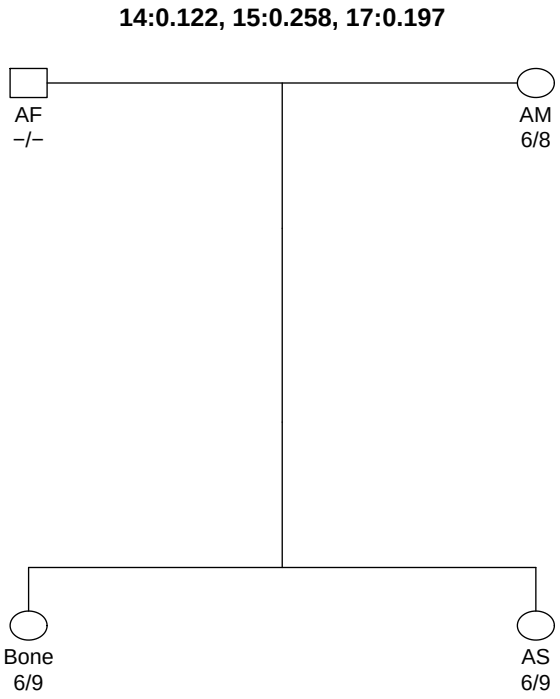

Figure 3: The case of the missing sister.

5. We revisit Exercise 1. Rather than calculating the LR we will now calculate the Essen-Møller index W defined as the probability of H_1 conditional on the genotypic data. Assume apriori that the hypotheses H_1 and H_2 are apriori equally likely. Then it can be shown that

$$W = P(H_1|\text{data}) = \frac{LR}{LR+1} \tag{3}$$

a) Recall that $LR = 20$ for the first marker. Calculate W.

b) Recall that $LR = 200$ for two markers. Calculate W.

c) Use Familias to calculate W for the two above cases.

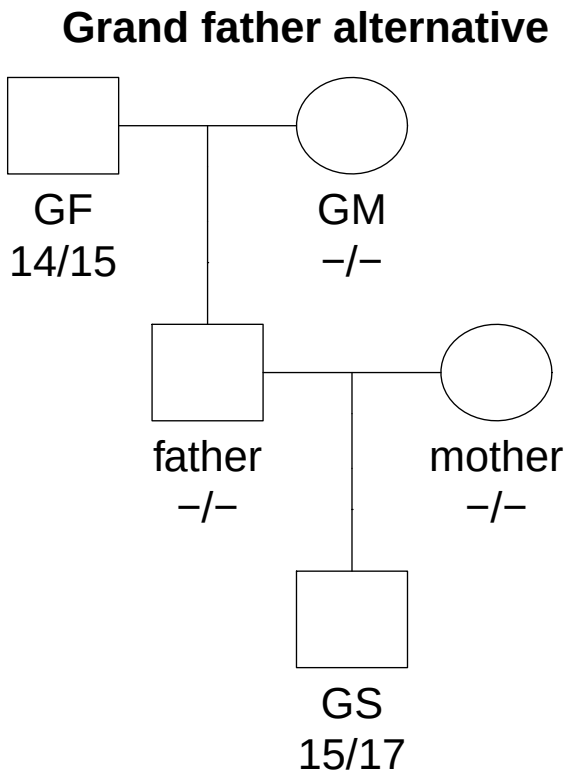

Figure 4: Grandfather-grandson.

d) For the discussion: Do you prefer LR or W?

6. Consider the following hypotheses

   H_1 : AF, the father of mother (undisputed), is the father also of her child.

   H_2 : An unrelated man is the father of the child.

   Figure 6 shows the pedigree corresponding to hypothesis H_1. The allele frequencies are $p_1 = p_2 = p_3 = 0.05$.

   a) Use Familias to calculate the LR. Check that your result coincide

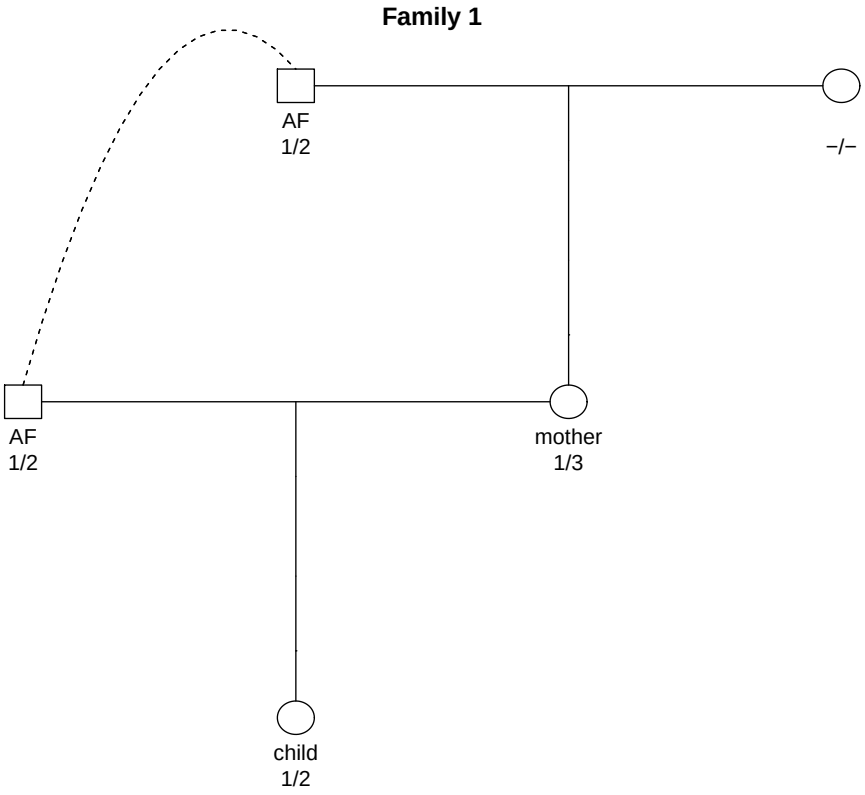

Figure 5: Incest by father.

with the theoretical Does the incest influence the result LR in this particular case?

b) The defense claims that one should rather consider the following three hypotheses

H_1 AF, the father of mother, is the father also of her child.

H_2 An unrelated man is the father of child.

H_3 The brother of mother is the father of the child. See Figure 6.

The LR can be calculated in several ways depending on the choice of the reference. Calculate Check that

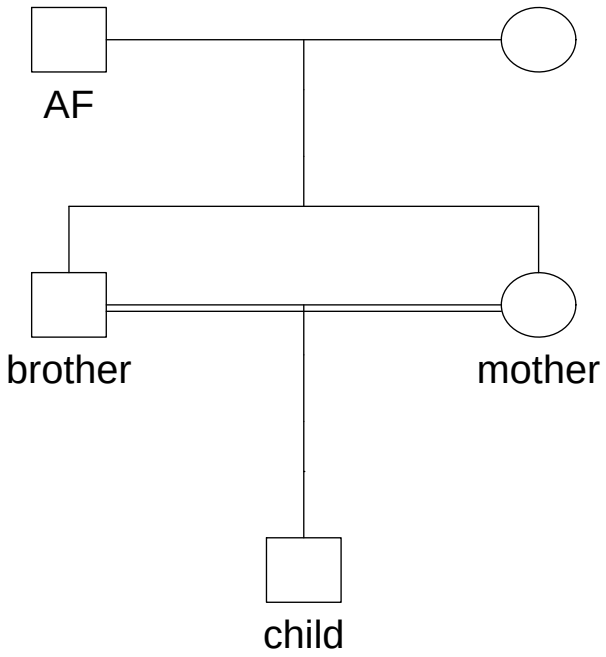

Figure 6: Incest by brother.

c) When there are more than two hypotheses, as above, some prefer to rather calculate posterior probabilities for the hypotheses. Assume that each of the three hypotheses are equally likely and calculate the posterior probabilities

7. In this exercise, which extends on Exercise 2, we will try the different mutation models. Throughout we consider data from the system VWA given in Exercise 2. The case data and hypotheses are:

   - Alleged father:14,15,
   - child:16,17,

- H_1 Alleged father is the true father.
- H_2 Alleged father and child unrelated.

There are five different mutation models in Familias. In this exercise, all will be tried. We will use overall mutation rate R=0.005 and the same model for females and males. The answer for this exercise will be obtained using Familias; Exercise 8, on the other hand, is based on theoretical calculations.

a) Load the data ExS7.fam into Familias.

b) Use model four `Equal probability`. Calculate the LR. (Answer; LR=2.9E-03=0.0029. Comment: For this model mutations to all other alleles are equally likely).

c) Use model three `Probability proportional to frequency`. Calculate the LR. (Answer LR=6.3E-03=0.0063. Comment: For this model it is more likely to mutate to a common allele compared to a rare).

d) Use model two `Stepwise (unstationary)`. For this model there are two parameters. The first is as before and should be set to R=0.005. Set the second parameter, Mutation range to r=0.5. Calculate the LR. (Answer: LR=4.7E-03=0.0047. Comment: For this model, mutation depends on the size of the mutation. With r=0.5, a two step mutation occurs with half the probability of a one-step mutation. A three step mutation occurs with half the probability of a two step mutation and so on.)

e) Use model one `Stepwise (stationary)` with parameters as above. (Answer: LR=6.4E-03=0.0064.)

f) Use model `Extended stepwise` with last parameter 0.1. In this case, with no microvariants, the results is the same as for Stepwise (unstationary) regardless of values of last parameter.

g) Comment: Model 1 (`Probability decreasing with range (stable)`) and model 3 (`Probability proportional to frequency`) and 4 are stationary whereas model 2,4 and 5 are non-staionary. If a model is stable, introducing a new untyped person, say the father of the alleged father, does not change the LR. This is a reasonable

property of a model as introducing irrelevant information should not change the result. For an unstable model, however, the LR will change slightly as allele frequencies may then differ from one generation to the next.

Verify that for model 1 and model 3, that the LR does not change if a father of the alleged father is introduced, whereas a slight change occurs for the two other models.

h) Comment: There is an other subtle point of all mutation models: In this case only five alleles (14, 15, 16, 17 and Rest allele) are needed rather than the 8 alleles defined. A five allele model will lead to slightly changing LRs.

Verify the above and comment.

8. In this exercise, which serves to confirm the answers obtained in Exercise 7, we will fill in some mathematical details related to the mutation models. The data and hypotheses are as in Exercise 7.

a) Show that

$$LR = \frac{p_{16}(m_{14,17} + m_{15,17}) + p_{17}(m_{14,16} + m_{15,16})}{4p_{16}p_{17}} \tag{4}$$

where p_{16} is the allele frequency for allele 16 and $m_{14,17}$ is the probability of a mutation from 14 to 17 and so on.

b) For the Equal probability mode $m_{ij} = m = R/(n-1)$ n is the number of alleles. Explain why the formula in Exercise 2 follows from (4). Show that when $n = 8$ and $R = 0.005$, $LR = 0.0029$, as in Exercise 7 b).

b) Consider next the proportional model. By definition,

$$m_{ij} = kp_j \text{ for } i \neq j, \tag{5}$$
$$m_{ii} = 1 - k(1 - p_i). \tag{6}$$

Show that

$$LR = k = \frac{R}{\sum_{i=1}^{n} p_i(1 - p_i)}.$$

Person	S1	S2	S3	S4	S5
Mother	A1/A2	A1/A2	A2/A3	A2/A4	A2/A3
Sister 1	A2/A3	A2/A3	A3/A4	A3/A4	A3/A4
Sister 2	A2/A3	A2/A3	A1/A3	A1/A3	A1/A3

Table 2: Marker data for Exercise 10

Comment: Exact calculations for the last two mutation models are more technical and are explained in Section A.1.4 of the manual for Familias 2.

9. Load the file ExS9.fam. Consider the hypotheses of Exercise 1.

   a) Show that the $LR = 0$.

   b) There is one marker where the child and the alleged father do not share an allele. Find this marker.

   c) Use model one `Stepwise (stationary)` for females and males with mutation rate 0.001 and mutation range 0.5 for the marker PENTA_E and calculate LR.

   d) Assume you are asked to consider the hypotheses H_3: Brother of alleged father is father. Calculate LR (H_1/H_3).

   e) For the discussion: Is there a best mutation model? Should a mutation model be used routinely for all markers?

10. We would like to determine whether two girls (called sister1 and sister2 in the left part of Figure 7 below) are sisters (corresponding to hypothesis H_1) or if they are half sisters (corresponding to hypothesis H_2 shown on the right hand side of Figure 7). The alleles are given in Table 1 below. The allele frequencies are 0.1 for systems S1 and S2, and 0.05 for systems S3-S5.

    a) What is the LR comparing the full sisters-alternative compared to the half-sisters-alternative?

    a) For the discussion: The LR in this case does not give rise to a clear conclusion. How many, which, further markers should be used?

11. See Figure 7. This is a paternity case where there is suspicion of a silent allele. Include a silent allele frequency of 0.05, and calculate the LR using Familias. The allele frequencies of A and B are both 0.1.

12. Verify theoretically the formula in Exercise 1 h.

13. This exercise expands on the previous Exercise **??** by introducing theta correction: Calculate the LR based on all markers. To save time you can read input from the file ExS4.fam and introduce a theta (kinship-parameter) of 0.02. (Hint: This is done in the pedigree window and you should press the `Options` button.)

14. A practical way to start work with Familias is to begin by reading a Familias file containing the relevant database. Sometimes it is, however, of interest to read and write databases and case data and this will be the topic below.

    a) Read input from the file ExS3.fam.

    a) Export the data base from the `General DNA data` window. Call the output file database.txt

    a) Import the case data from the `Case Related DNA Data`. Call the output file casedata.txt.

    a) Open a new, blank, project in Familias.

    a) Import database.txt

    a) Import casedata.txt

    a) Define the pedigrees, see Exercise S3, and calculate the LR once more.

15. Load the file ExS15.fam. Consider the hypotheses

    H_1 : The alleged father (AF) is the real father.

    H_2 : The alleged father and the child are unrelated.

    The mother is undisputed.

    a) Confirm that that the LR=0 and find the one marker where the child and the alleged father do not share an allele.

b) Set the dropout probability to 0.1 for this marker, choose consider dropout for the child and recalculate the LR.

c) This dropout could also be treated as a mutation. Compare the LR with the LR you get with the mutation model (for this marker) Stepwise (stationary) for females and males with mutation rate 0.001 and mutation range 0.5 for this marker. Remove drop-out.

d) Discuss: Should dropout be used for all homozygous markers?

16. See the paternity case in Figure 7. We will consider two scenarios, first that there is a silent allele inherited from the alleged father to the child, and second that there is a dropout in both individuals. Let the allele frequencies of A and B both be 0.2.

    a) Include a silent allele frequency of 0.05, and calculate the LR using Familias.

    b) Remove the silent allele frequency and instead include a dropout probability of 0.05 for both the alleged father and child. Calculate LR.

17. Load the file ExS17.fam. The hypotheses considered are as in Exercise 15.The file contains no genotype information for the alleged father and child. Use the simulation in Familias to simulate genotypes for both individuals. Use 1000 simulations and find

    a) The expected LR(H_1/H_2) when H_1 is true.

    b) The expected LR(H_1/H_2) when H_2 is true.

    c) The probability of observing a LR larger than 50 when H_1 is true.

Solutions for Familias 3 exercises. Independent markers

The solutions for the exercises in **??** follow. In addition, there are some videos[1], currently only for Familias 2, available as solutions for some exercises.

[1](http://arken.umb.no/~theg/alcala/Familiasvideos.pdf

1. **a)** We can write

$$LR = \frac{P(child|mother, father)}{P(child|mother}$$

Consider first the numerator. The only possible genotype for the child, given the mother and the AF, is A/B, and therefore the probability is 1. For the denominator, the father must have passed on the A allele, and therefore the probability is. Hence . The standard interpretation is *The data is 20 times more likely assuming AF to be the father compared to the alternative that some unknown man is the father.*

b) This is implemented in Familias as explained in Section **??**.

c) Repeat step 1 above to define the marker S2 and step 3 to define the genotypes for all individuals. Steps 2 and 4 need not be repeated. The LR is then obtained as before and the answer is: LR=200.

d) $(1/p_A) * (1/p_B) = (1/0.05) * (1/0.05) = 200$.

g) $1/RMP = 40000$. This is found by marking AF in the `Case-related DNA-data` window and pressing `compare DNA`.

h) $LR = (1 + 3 * 0.02)/(2 * 0.02 + (1 - 0.02) * 0.05) = 11.91$

i) Note that Hardy-Weinberg equilibrium is required for the LR derivation for each marker. This assumption is not needed when we use theta-correction. Furthermore, linkage disequilibrium is needed. We have also assumed that there are no mutations or silent alleles.

2. **a)** The numerator, and therefore LR=0 as confirmed by Familias.

b) $LR = 4.07E - 03 = 0.00407$

c) $LR = 0.01 * (0.212 + 0.292)/(2 * 0.212 * 0.292) = 0.00407$.

3. The pedigree corresponding to H_1, Figure 8, is specified as follows Most answers are given in the rightmost column of the below table from the report generated by Familias. D7S820, gives a very large LR, namely 11189 since the allele 11.1 is so rare. If this marker is omitted the new LR is 530437141.9/11189.4=47407.

Likelihood ratio versus H2: 530437140,978139

System:	Likelihood:	LR versus H2:
D3S1358	0,002107356328	24,3295665373587
D21S11	0,00106135425	4,42583732057416
D18S51	0,0008102440125	26,7722473604827
D7S820	5,037122916E-07	11189,4499975311
D16S539	0,000811482	0,899280575539568
CSF1PO	0,0130390037415	2,25229008488432
F13B	0,001423488512	5,88485536018151
LPL	0,0007365230775	1,37960642346607

4. **a)** LR(grand father/unrelated) = 0.98 for D3S1358.
   b) LR(grand father/unrelated) = 0.0085 for all markers.
   c) Whether autosomal and haplotype markers can be combined and in case how is a big question and there appears to be no consensus.

5. **a)** $W = 20/(1 + 20) = 0.952$.
   b) $W = 200/(1 + 200) = 0.995$
   c) Same answers as above.
   d) This is a again a big question. Most recommendations are in favor of LR. The problem with W is to decide on the prior. W may be easier to interpret.

6. **a)** LR=10. Comment: The formula provided follows from The same result is obtained in this particular case if inbreeding is ignored.

b) LR (H1/H2)= 10, LR(H1/H3)=1.905. We are asked to verify that

c)

7. Most answers are given in problem. Regarding question f. We can add the following: If a model is unstable, allele frequencies change with generations. This adds some intuition as to why introducing an extra person, say a grandparent, typically changes results slightly for an unstable model.

   Regarding question g. Consider the first model listed in Familias Stepwise (unstationary) which gives $LR = 6.4E-03 = 0.0064$. Remove alleles so that only 14, 15, 16,17 and a `Rest allele` remains. Then $LR = 7.8E-03 = 0.00784$, a change. Again this effect does not have to with Familias; there are two different models, one with 8 alleles and one with alleles (which constrains mutations within those five alleles) and different models typically give different results.

8. **a)** We find denominator by a direct argument and the numerator by conditioning on the paternal allele:

$$LR = \frac{P(data|H_1)}{P(data|H_2)} = \frac{p_{16}(m_{14,17}+m_{15,17})+p_{17}(m_{14,16}+m_{15,16})}{4p_{16}p_{17}}.$$

   b) The formula in Exercise 2 follows by setting $m_{14,17} = m_{15,17}) = m_{14,16} = m_{15,16})$ and the numerical result $LR = 0.0029$ follows.

   c) We find $LR = k$ by using the definition $m_{ij} = kp_j$. Furthermore,

$$1 - R = \sum_{i=1}^{n} p_i(1-p_i) \;\Rightarrow\; k = \frac{R}{\sum_{i=1}^{n} p_i(1-p_i)} = 0.0063$$

9. **a)** $LR(H_1 : \; father/H_2 : \text{not father}) = 0$ from Familias.

   b) $LR = 0$ for PENTA_E as can be seen from report from Familias.

   c) $LR = 4.5 \cdot 10^6$

   d) $LR = 1.34$.

10. **a)** $LR = 0.791/0.209 = 3.8$. Details on Familias implementation, including definition of pedigrees, can be seen from Exercise 10.

11. The LR (not-father/father) is 1.36 How to do it: Enter the allele system setting the silent allele frequency to 0.05. Enter the persons and their DNA data as usual. Construct the pedigrees manually and Calculate. A theoretical argument, not requested, is given below: Let pA = pB = 0.1 and pS = 0.05. According to http://dna-view.com/patform.htm,

$$\begin{aligned} LR &= \frac{p_s(p_A + p_s)}{(p_A + p_s)^2(p_B + 2p_s) + p_s p_A(p_B + 2p_s)} \\ &= \frac{0.05 \cdot 0.15}{0.15^2 \cdot 0.2 + 0.05 \cdot 0.1 \cdot 0.2} = 1.36 \end{aligned} \tag{7}$$

which coincides with Familias.

12. See pages 36-37 of the manual for Familias 2.

13. $LR = 0.0068$.

14. See video.

15. *Paternity case with dropout.*

    a) LR(H1: father/H2:not father)=0 from Familias and we see in the report that LR=0 for PENTA_E.

    b) `General DNA data`: Click on the marker penta_E. In the new window click on options and set Dropout to 0.1. Save. `Case-related DNA data`: choose the child and tick `consider dropout` in the new window. A message will appear saying that Familias will model dropout, click OK. Calculate $LR = 2679875170$.

    c) `Case-related DNA data`: Untick 4 `consider dropout` for the child. `General DNA data`: Choose mutation model for Penta_E, see previous exercises on how to do this. Calculate $LR = 1078633$.

16. *Silent allele or dropout?* Note that both alleged father and child must be given as homozygotes in Familias.

    a) General DNA data: when editing the allele data, choose options and include a silent allele frequency of 0.05. Note that allele frequencies and silent allele frequency should add to 1. We find $LR = 0.57$.

b) General DNA data: Remove the silent allele frequency and include a dropout probability of 0.05. Case-related DNA data: Tick `onsider dropout` for both the alleged father and child. LR=0.34.

17. Load file with allele frequencies. Add alleged father and child without any genotype information. Add pedigrees for both hypotheses. Simulation: In Pedigrees click Simulate. Move both AF and CH to `Will be genotyped`. The simulation will produce slightly different results each time it is run unless a seed is set. If you untick `random seed` and set seed to 12345, you should get the exact same results as below. Click simulate. It may take some time depending on the number of simulations you have chosen. In the new window choose display and tick verb@Mean@. The mean LR is shown for both H_1 true and H_2 true.

    a) The expected LR when H1 is true is 0.8988

    b) The expected LR when H2 is true is 40.7

    c) Click `LR limit`, choose LR threshold 50 and click update. The probability of observing a LR larger than 50 is 0.089.

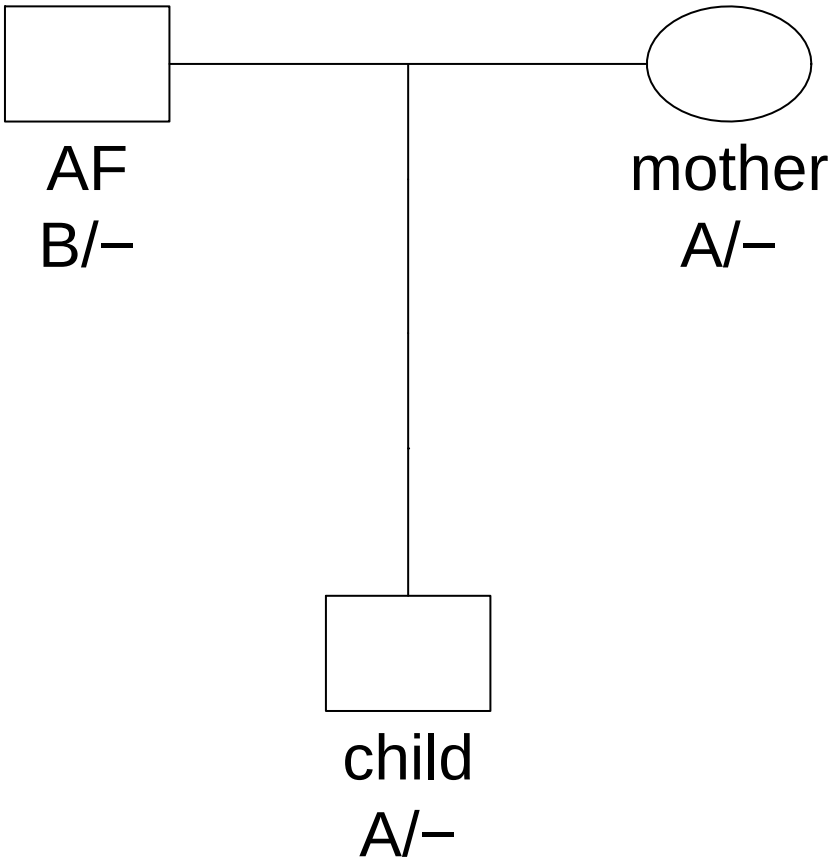

Figure 7: Pedigree for Exercise 11 and 16.

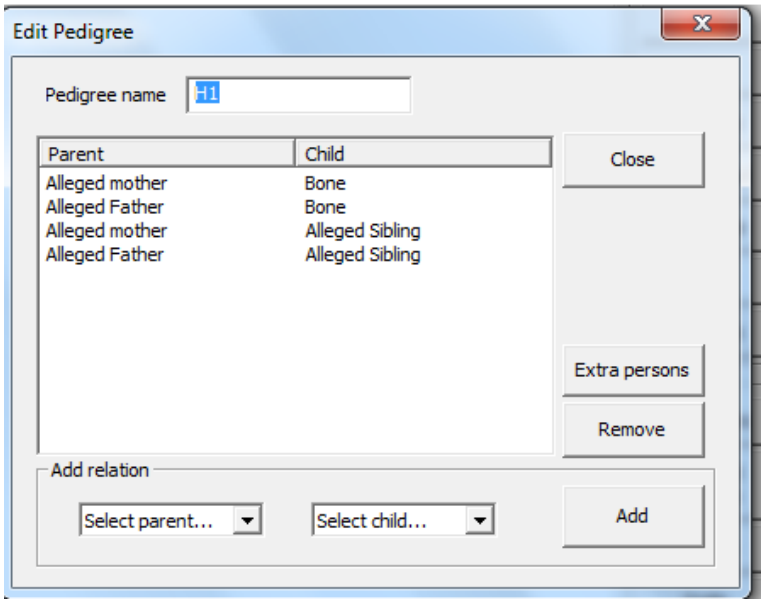

Figure 8: Defining the pedigree.