

Homework 4

Lorenzo Galizia, Vincenzo Genna

16/12/2017

2. Data set information:

Individuals	SNPs	Missing data(%)
139	28	39.54

3. $2^{n_{SNP}}$ (= 268435456) haplotypes can theoretically be found if all the genotypes are different.

4. 6 haplotypes were found using the EM algorithm:

- List:

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	SNP10
1	C	A	A	C	C	A	G	T	A	G
2	T	A	A	C	C	A	G	C	G	G
3	T	A	A	C	C	A	G	C	G	G
4	T	A	A	C	C	A	G	T	A	G
5	T	A	G	G	C	A	G	T	A	G
6	T	C	A	C	C	A	C	T	A	A

	SNP11	SNP12	SNP13	SNP14	SNP15	SNP16	SNP17	SNP18	SNP19
1	C	G	C	C	A	A	C	C	C
2	C	A	A	T	G	G	T	C	T
3	C	A	A	T	G	G	T	C	T
4	C	A	A	T	G	A	C	C	T
5	C	G	C	C	A	A	C	C	C
6	T	G	C	C	A	A	C	A	C

	SNP20	SNP21	SNP22	SNP23	SNP24	SNP25	SNP26	SNP27	SNP28
1	C	G	C	C	C	C	A	C	T
2	C	A	T	T	C	C	A	C	T
3	C	G	C	T	C	C	A	C	T
4	C	G	C	T	C	C	A	C	T
5	C	G	C	C	C	C	A	C	T
6	G	G	C	C	C	C	A	C	T

- Most common haplotype:

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	SNP10
6	T	C	A	C	C	A	C	T	A	A

	SNP11	SNP12	SNP13	SNP14	SNP15	SNP16	SNP17	SNP18	SNP19
6	T	G	C	C	A	A	C	A	C

	SNP20	SNP21	SNP22	SNP23	SNP24	SNP25	SNP26	SNP27	SNP28
6	G	G	C	C	C	C	A	C	T

5. The haplotypic constitution is ambiguous for the individuals in the listIDs:

```
## [1] 1 2 4 5 8 10 11 12 13 15 16 20 22 23 35 37 40
## [18] 42 43 44 46 49 50 51 52 53 54 56 57 61 64 74 75 76
## [35] 81 83 84 85 88 89 92 93 97 100 101 102 104 111 114 115 118
## [52] 119 121 126 127 129 130 131 133 134 135 137
```

6. Sembra che i risultati non cambiano però il bisognerebbe sapere il perchè...

- List:

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	SNP10
1	C	A	A	C	A	G	T	A	G	C
2	T	A	A	C	A	G	C	G	G	C
3	T	A	A	C	A	G	C	G	G	C
4	T	A	A	C	A	G	T	A	G	C
5	T	A	G	G	A	G	T	A	G	C
6	T	C	A	C	A	C	T	A	A	T

	SNP11	SNP12	SNP13	SNP14	SNP15	SNP16	SNP17	SNP18	SNP19
1	G	C	C	A	A	C	C	C	C
2	A	A	T	G	G	T	C	T	C
3	A	A	T	G	G	T	C	T	C
4	A	A	T	G	A	C	C	T	C
5	G	C	C	A	A	C	C	C	C
6	G	C	C	A	A	C	A	C	G

	SNP20	SNP21	SNP22	SNP23	SNP24	SNP25	SNP26	SNP27
1	G	C	C	C	C	A	C	T
2	A	T	T	C	C	A	C	T
3	G	C	T	C	C	A	C	T
4	G	C	T	C	C	A	C	T
5	G	C	C	C	C	A	C	T
6	G	C	C	C	C	A	C	T

- Haplotype frequencies:

	1	2	3	4	5	6
freq	0.02	0.06	0.02	0.00	0.18	0.73