

732A96 Advanced Machine Learning

LAB 1 Bioinformatics

Hugo Morvan
William Wiik

STIMA
Department of Computer and Information Science
Linköpings universitet

2024-11-12

Contents

1	Question 1	1
1.1	Question 1.1	1
1.2	Question 1.2	1
2	Question 2	3

1 Question 1

We consider a gene locus with two possible alleles (say A and a) and a diploid population with N individuals. Hence, there are $2N$ alleles in the population. Let p be the proportion of A s in the allele population and q the population of a s (of course $p + q = 1$). A population is said to be in Hardy-Weinberg equilibrium if the proportion of AA homozygotes is p^2 , aa homozygotes is q^2 and the proportion of heterozygotes (Aa) is $2pq$.

1.1 Question 1.1

Show that with random mating (i.e. both alleles of the offspring are just randomly, with proportions p and q , drawn from the parental allele population) Hardy-Weinberg equilibrium is attained in the first generation. What is the proportion of A and a alleles in the offspring population? Hence, with random mating, can a population in Hardy-Weinberg equilibrium ever deviate from it?

If the mating is random then all individuals inherit alleles independently with probabilities p and q from each parent. Then:

$$P(AA) = p \cdot p = p^2$$

$$P(aa) = q \cdot q = q^2$$

$$P(Aa) = (p \cdot q) + (p \cdot q) = 2pq$$

The population can not deviate from Hardy-Weinberg equilibrium unless factors such as mutation changes the alleles.

1.2 Question 1.2

We look at the MN blood group, it has two possible codominating (both contribute to heterozygotes) alleles L^M (denoted M) and L^N (denoted N). In a population of 1000 Americans of Caucasian descent the following genotype counts were observed, 357 individuals were MM , 485 were MN and 158 were NN . Use a chi-square goodness of fit test to test if the population is in Hardy-Weinberg equilibrium.

```
MM <- 357
MN <- 485
NN <- 158

population <- 1000

# p = the proportion of Ms in the allele population
p <- (2 * MM + MN) / (2 * population)
q <- 1 - p

# Hardy-Weinberg equilibrium
HW_MM <- p**2 * population
HW_NN <- q**2 * population
HW_MN <- q * p * population * 2
```

```
# chisq = sum[(Observed - Expected)^2 / (Expected)]  
##pchisq()
```

2 Question 2

In this exercise, you will use ...