Practical 5

Statistical Genetics: Genetic Association Analysis

Anna Putina Marine Mazeau

2023-12-10

```
library(stats)
library(HardyWeinberg)
```

A particular SNP is supposed to be involved in Alzheimer's disease. A case control study has been performed, obtaining the following results:

```
X <- matrix(c(112, 278, 150, 206, 348, 150), byrow=TRUE, ncol=3)
colnames(X) <- c("AA", "Aa", "aa")
rownames(X) <- c("Cases", "Controls")
print(X)
```

```
##           AA  Aa  aa
## Cases    112 278 150
## Controls 206 348 150
```

Question 1:

Perform the alleles test for this data set. Provide the p-value and the odds ratio and comment on the results.

Alleles table:

```
Y <- cbind(2*X[,1]+X[,2], 2*X[,3]+X[,2])
colnames(Y) <- c("A","a")
print(Y)
```

```
##            A   a
## Cases    502 578
## Controls 760 648
```

```
allele_test <- chisq.test(Y, correct=FALSE)
fisher_test <- fisher.test(Y)

p_value <- allele_test$p.value
odds_ratio <- fisher_test$estimate
conf_int <- fisher_test$conf.int

cat("P-value chisq test:", p_value, "\n")
```

```
## P-value chisq test: 0.0002100232
```

```
cat("Odds Ratio:", odds_ratio, "\nwith the confidence interval:", conf_int)
```

```
## Odds Ratio: 0.7406221
## with the confidence interval: 0.6295932 0.8709495
```

P-value of chi square test is lower than 0.05 so we can conclude that the association between alleles and case-control status is statistically significant.

Odds ratio of the fisher test is different from 1, i.e. we can reject the null hypothesis that the probability of A is the same in both cases and controls groups. As the odds ratio is lower than 1, it means that the odds of disease with allele a are higher than with allele A.

Thanks to these alleles tests, we can say that there seems to be a dependance between the allele's value and the presence of the disease. It seems that allele a leads more often to the presence of the disease than allele A. Disease is 0.741 more frequent with allele A.

Question 2:

Test for association using a codominant, a dominant and a recessive model. Provide the p-values for all the tests and comment on the results.

Codominant model

```
codom_chisq <- chisq.test(X)
codom_fisher <- fisher.test(X)

cat("P-value for chi-squared test:", codom_chisq$p.value, "\n")
```

```
## P-value for chi-squared test: 0.0008085403
```

```
cat("P-value for fisher test:", codom_fisher$p.value, "\n")
```

```
## P-value for fisher test: 0.0007782748
```

P-values are both lower than 0.05 : we reject the null hypothesis that the probability of disease with all the genotypes is the same.

Dominant model : A is dominant

```
X.dom <- cbind(X[,1]+X[,2], X[,3])
colnames(X.dom) <- c("AA or Aa", "aa")
print(X.dom)
```

```
##          AA or Aa  aa
## Cases         390 150
## Controls      554 150
```

```
dom_chisq <- chisq.test(X.dom)
dom_fisher <- fisher.test(X.dom)

cat("P-value for chi-squared test:", dom_chisq$p.value, "\n")
```

```
## P-value for chi-squared test: 0.009952705
```

```
cat("P-value for fisher test:", dom_fisher$p.value, "\n")
```

```
## P-value for fisher test: 0.009094037
```

P-values are both lower than 0.05 : we reject the null hypothesis that the probability of disease does not depend on A.

Recessive model : A is recessive

```
X.rec <- cbind(X[,1], X[,2]+X[,3])
colnames(X.rec) <- c("AA", "Aa or aa")
print(X.rec)
```

```
##           AA Aa or aa
## Cases    112      428
## Controls 206      498
```

```
rec_chisq <- chisq.test(X.rec)
rec_fisher <- fisher.test(X.rec)

cat("P-value for chi-squared test:", rec_chisq$p.value, "\n")
```

```
## P-value for chi-squared test: 0.0008108124
```

```
cat("P-value for fisher test:", rec_fisher$p.value, "\n")
```

```
## P-value for fisher test: 0.0006464142
```

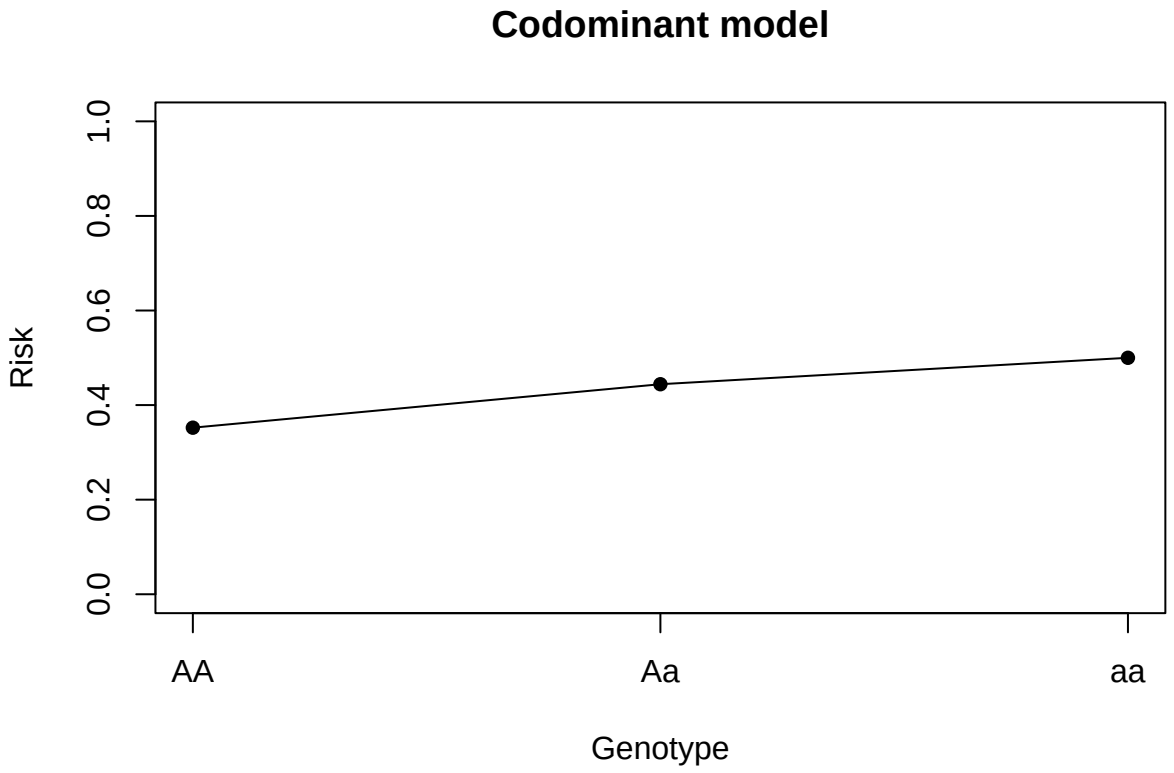
P-values are both lower than 0.05 : we reject the null hypothesis that the probability of disease does not depend on a.

As a conclusion, the probability of the disease seems to depend on both alleles a and A, and the probability of the disease among all genotypes is different.

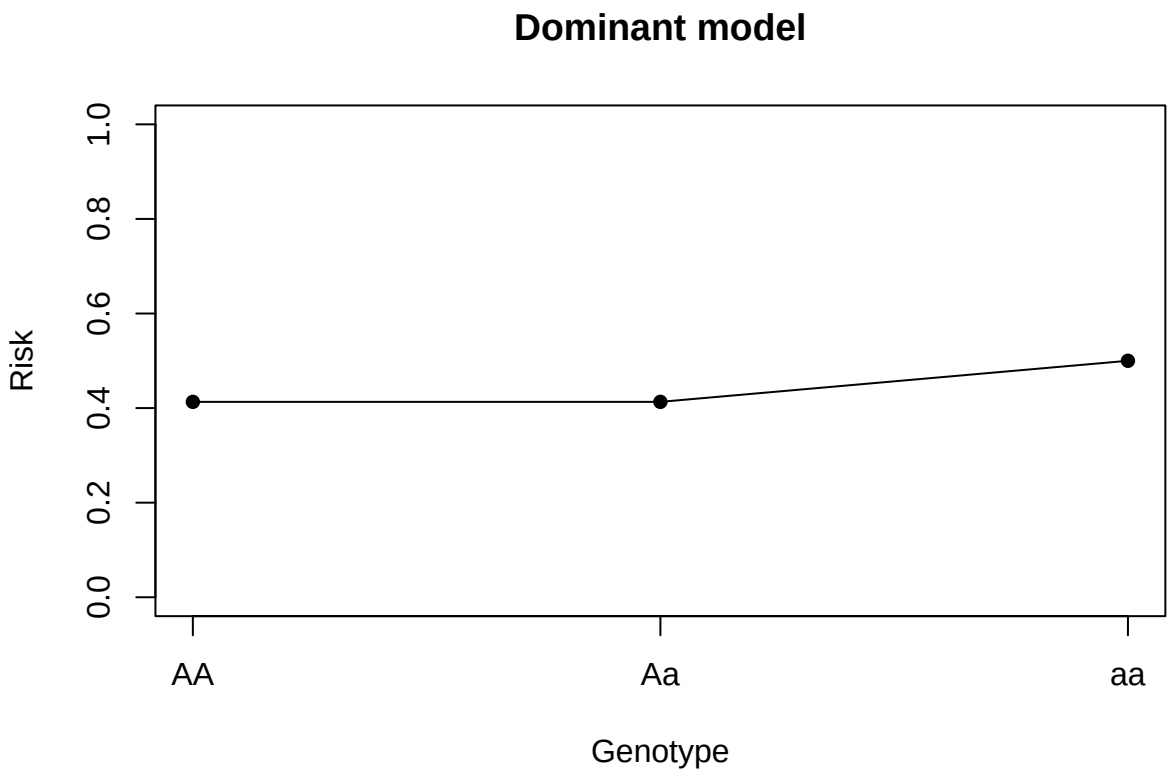
Question 3:

Plot the risk of disease as a function of the number of m alleles. Comment on the results. Which model seems most appropriate?

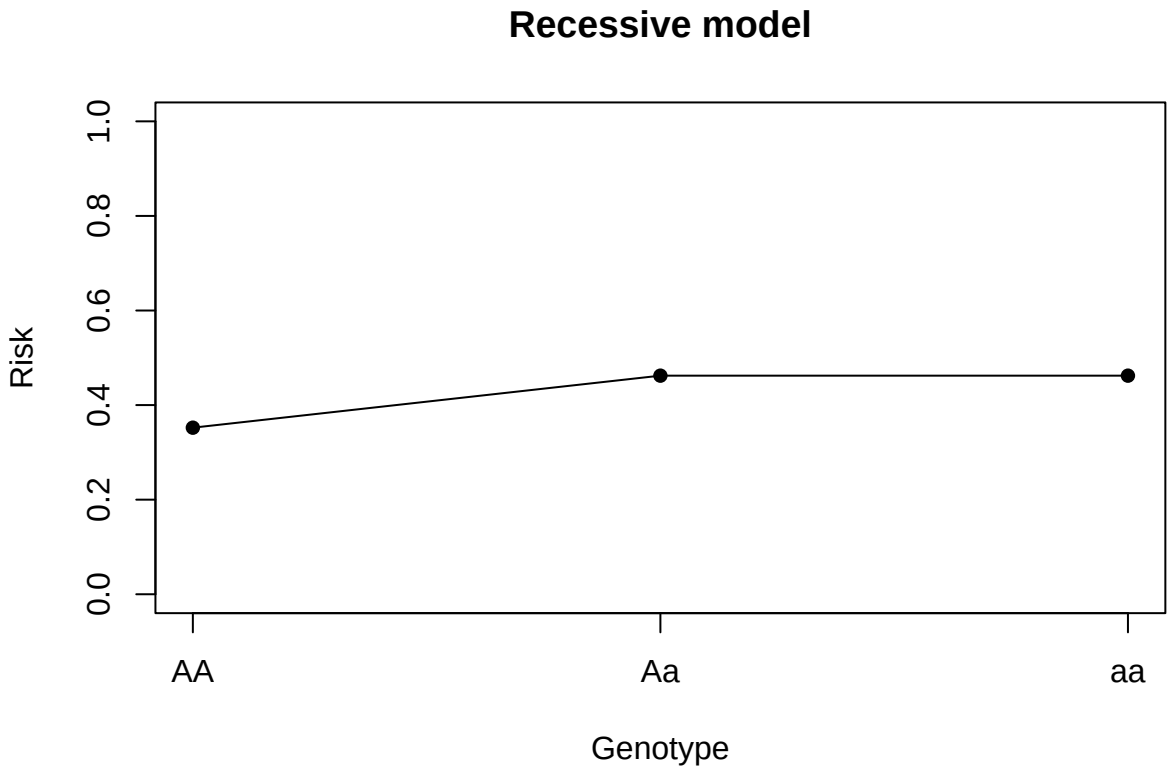
```
proportions_codom_cases <- X["Cases", ] / colSums(X)

plot(c(0,1,2), proportions_codom_cases, pch = 16,
     main = "Codominant model",
     xlab = "Genotype", ylab = "Risk", ylim = c(0, 1), xaxt = "n")
lines(c(0,1,2), proportions_codom_cases, type = "l")
axis(1, at = 0:2, labels = c("AA", "Aa", "aa"))
```



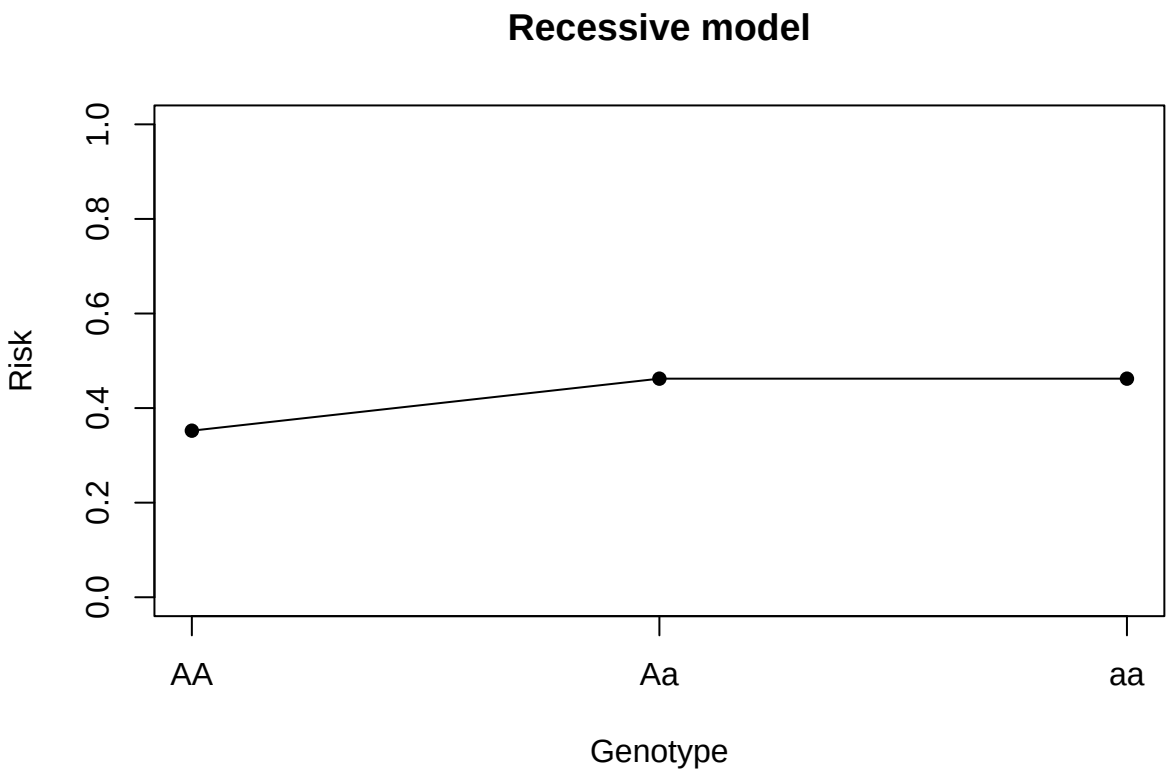

```
proportions_dom_cases <- X.dom["Cases", ] / colSums(X.dom)

proportions_dom_cases <- c(proportions_dom_cases[1], proportions_dom_cases)

plot(c(0,1,2), proportions_dom_cases, pch = 16,
     main = "Dominant model",
     xlab = "Genotype", ylab = "Risk", ylim = c(0, 1), xaxt = "n")
lines(c(0,1,2), proportions_dom_cases, type = "l")
axis(1, at = 0:2, labels = c("AA", "Aa", "aa"))
```




```
proportions_rec_cases <- X.rec["Cases", ] / colSums(X.rec)

proportions_rec_cases <- c(proportions_rec_cases, proportions_rec_cases[2])

plot(c(0,1,2), proportions_rec_cases, pch = 16,
     main = "Recessive model",
     xlab = "Genotype", ylab = "Risk", ylim = c(0, 1), xaxt = "n")
lines(c(0,1,2), proportions_rec_cases, type = "l")
axis(1, at = 0:2, labels = c("AA", "Aa", "aa"))
```




```
proportions_rec_cases <- X.rec["Cases", ] / colSums(X.rec)

proportions_rec_cases <- c(proportions_rec_cases, proportions_rec_cases[2])

plot(c(0,1,2), proportions_rec_cases, pch = 16,
     main = "Recessive model",
     xlab = "Genotype", ylab = "Risk", ylim = c(0, 1), xaxt = "n")
lines(c(0,1,2), proportions_rec_cases, type = "l")
axis(1, at = 0:2, labels = c("AA", "Aa", "aa"))
```



Question 4: Perform Armitage trend test for this data set. Does the null hypothesis $\beta_1 = 0$ hold? Comment on your response.

```
trend_test <- prop.trend.test(X[1,], colSums(X))

cat("Chi-squared Test Statistic:", trend_test$statistic, "\n")
```

```
## Chi-squared Test Statistic: 13.83108
```

```
cat("P-value:", trend_test$p.value, "\n")
```

```
## P-value: 0.0002000008
```

P-value is lower than 0.05 : we reject the null hypothesis that the probability of disease is the same with different number of A.

Question 5: Is there evidence for association of this marker with the disease? Argument your response.

The result of chi-square test and odds ratio of fisher test confirmed that there is an association between the marker and the disease, and that the odds of disease with allele a are higher than with allele A.

NOT FINISHED