Practical5 - Statistical Genetics

Anton Ickler Sara Montese

2023-12-12

1. (2p) Perform the alleles test for this data set. Provide the p-value and the odds ratio and comment on the 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")
X
```

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

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

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

```
chisq_test <- chisq.test(Y,correct=FALSE)
cat("\np-value:", chisq_test$p.value)
```

```
##
## p-value: 0.0002100232
```

```
fisher_test <- fisher.test(Y)
cat("\nodds ratio:", fisher_test$estimate)
```

```
##
## odds ratio: 0.7406221
```

```
#or <- Y[1,1]*Y[2,2] / (Y[2,1]* Y[1,2])
fisher_test <- fisher.test(Y)
cat("\nodds ratio:", fisher_test$estimate)
```

```
##
## odds ratio: 0.7406221
```

The low p-value is highly significant indicating strong evidence to reject the null hypothesis of no association between the alleles and Alzheimer's disease.

The odds ratio less than 1, suggests that the presence of the alleles under consideration is associated with a reduced odds of having Alzheimer's disease In other words, individuals with these alleles are less likely to have Alzheimer's disease compared to those without these alleles.

2. (2p) 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
p_value_codom_chi <- chisq.test(X)$p.value
p_value_codom_fish <- fisher.test(X)$p.value

# dominant
Y_dom <- cbind(X[,1], X[,2]+X[,3])
colnames(Y_dom) <- c("AA", "Aa or aa")
rownames(Y_dom) <- c("Cases", "Control")
p_value_dom_chi <- chisq.test(Y_dom)$p.value
p_value_dom_fish <- fisher.test(Y_dom)$p.value

# recessive
Y_rec <- cbind(X[,1]+X[,2],X[,3])
colnames(Y_rec) <- c("AA or Aa", "aa")
rownames(Y_rec) <- c("Cases", "Control")
p_value_rec_chi <- chisq.test(Y_rec)$p.value
p_value_rec_fish <- fisher.test(Y_rec)$p.value

cat("P-value of Chi-squared test for codominant model:", p_value_codom_chi, "\n")
```

```
## P-value of Chi-squared test for codominant model: 0.0008085403
```

```
cat("P-value of Fisher's test for codominant model:", p_value_codom_fish, "\n\n")
```

```
## P-value of Fisher's test for codominant model: 0.0007782748
```

```
cat("P-value of Chi-squared test for dominant model:", p_value_dom_chi, "\n")
```

```
## P-value of Chi-squared test for dominant model: 0.0008108124
```

```
cat("P-value of Fisher's test for dominant model:", p_value_dom_fish, "\n\n")
```

```
## P-value of Fisher's test for dominant model: 0.0006464142
```

```
cat("P-value of Chi-squared testfor recessive model:", p_value_rec_chi, "\n")
```

```
## P-value of Chi-squared testfor recessive model: 0.009952705
```

```r
cat("P-value of Fisher's test for recessive model:", p_value_rec_fish, "\n")
```

```
## P-value of Fisher's test for recessive model: 0.009094037
```

P-values for the codominant model indicate strong evidence against the null hypothesis, suggesting a significant association between Alzheimer's disease and the genetic variables.

Also p-values for the dominant model suggest a strong association between the disease and the genetic variables. We reject the null hypothesis that the probability of the disease does not depend on the allele 'A'.

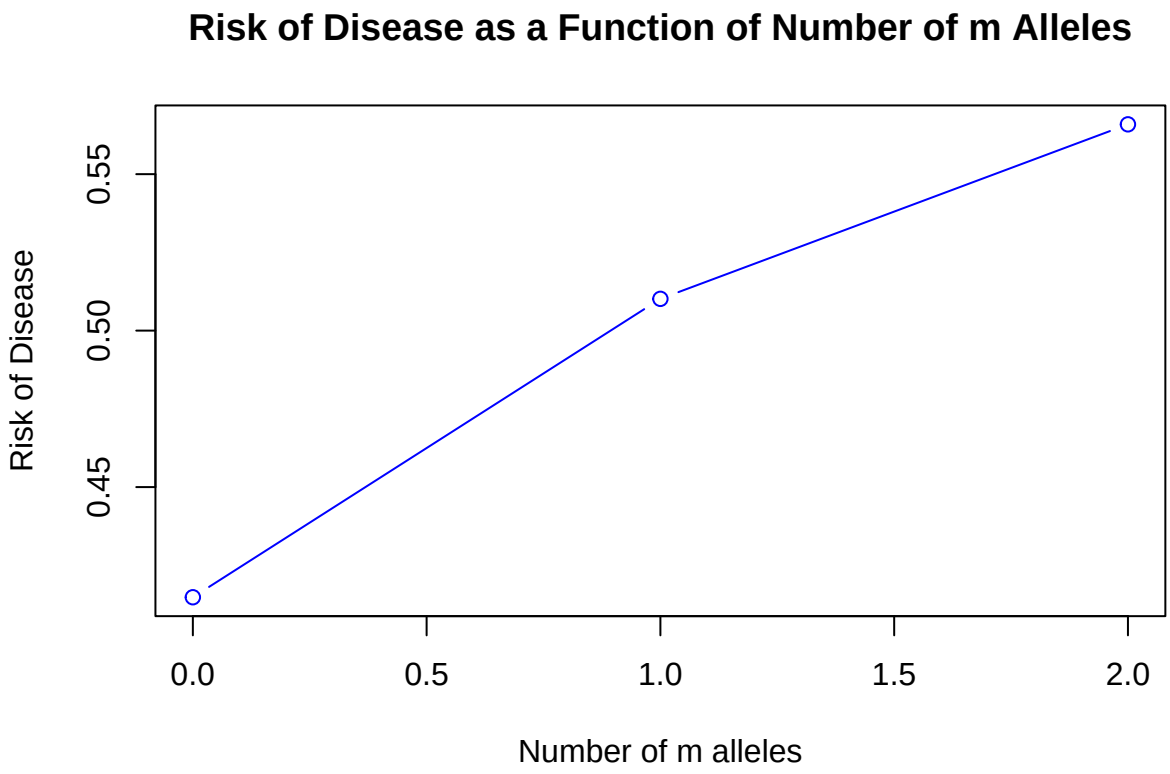
P-values for the recessive model are higher that those for the previous models but still suggest a statistically significant association under the recessive model. The relatively higher p-value in this model compared to the others suggests that the recessive model may be a less strong fit for the data, but there is still evidence to suggest that the 'aa' genotype is associated with Alzheimer's.

3. (2p) Plot the risk of disease as a function of the number of m alleles. Comment on the results. Which model seems most appropriate?

```r
# Calculate risk
allele_count <- c(0, 1, 2)
# Assuming m allele is represented by '2' in mm, '1' in Mm, and '0' in MM

cases <- rowSums(X)["Cases"]
controls <- rowSums(X)["Controls"]
risk <- (X[1,] / cases) / ((X[2,] / controls) + (X[1,] / cases))

# Plotting
plot(allele_count, risk, type="b", col="blue", xlab="Number of m alleles", ylab="Risk of Disease",
     main="Risk of Disease as a Function of Number of m Alleles")
```



The additive model seems appropriate, because the risk is rising with the amount of "m" alleles.

4. (2p) Perform Armitage trend test for this data set. Does the null hypothesis beta1 = 0 hold? Comment on your response.

```
n <- sum(X)
Cases <- X[1,]
Controls <- X[2,]
cas <- rep(c(0,1,2),Cases)
con <- rep(c(0,1,2),Controls)
y <- c(rep(1,sum(Cases)), rep(0,sum(Controls)))
x <- c(cas,con)
r <- cor(x,y)
A <- n*(r^2)
pvalue <- pchisq(A,df=1,lower.tail=FALSE)
pvalue
```

```
## [1] 0.0002000008
```

The results indicate a statistically significant trend, as evidenced by the low p-value. It suggests that there is strong evidence to reject the null hypothesis. In this context, the null hypothesis would be that the risk of having Alzheimer's disease does not change as the genotype changes. In practical terms, this means that there is reason to believe that the risk of Alzheimer's disease is influenced by the genotype under investigation

5. (2p) Is there evidence for association of this marker with the disease? Argument your response.

Yes, there is strong evidence for an association of this marker with Alzheimer's disease. The low p-values from various tests (alleles test, different genetic models, Armitage trend test) consistently reject the null hypothesis of no association. This suggests that the genotype under investigation is likely to influence the risk of Alzheimer's disease.