Practical 5 Statistical Genetics

Antonin Rosa

Hendrik Wellmann

Question 1

```
X <- matrix(c(112, 278, 150, 206, 348, 150), nrow = 2, byrow = TRUE)
colnames(X) <- c("AA","Aa","aa")
rownames(X) <- c("Cases","Controls")

Y <- cbind(2*X[,1]+X[,2],2*X[,3]+X[,2])
colnames(Y) <- c("A","a")
chi_square_result <- chisq.test(Y,correct=FALSE)
p_value <- chi_square_result$p.value
odds_ratio <- Y[1, 1] * Y[2, 2] / (Y[1, 2] * Y[2, 1])
print(odds_ratio)
```

```
## [1] 0.7405209
```

```
# Print the results
cat("Chi-square test results:\n")
```

```
## Chi-square test results:
```

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

```
## P-value: 0.0002100232
```

```
cat("Odds Ratio:", odds_ratio, "\n")
```

```
## Odds Ratio: 0.7405209
```

The small p-value suggests a strong association of the alleles with the disease. The odds ratio of 0.74 tells us that the disease is more common for people with the allele a.

Question 2 and 3

```
# codominant test
test <- chisq.test(X)
p_value <- test$p.value
cat("Codomminant test results:\n")
```

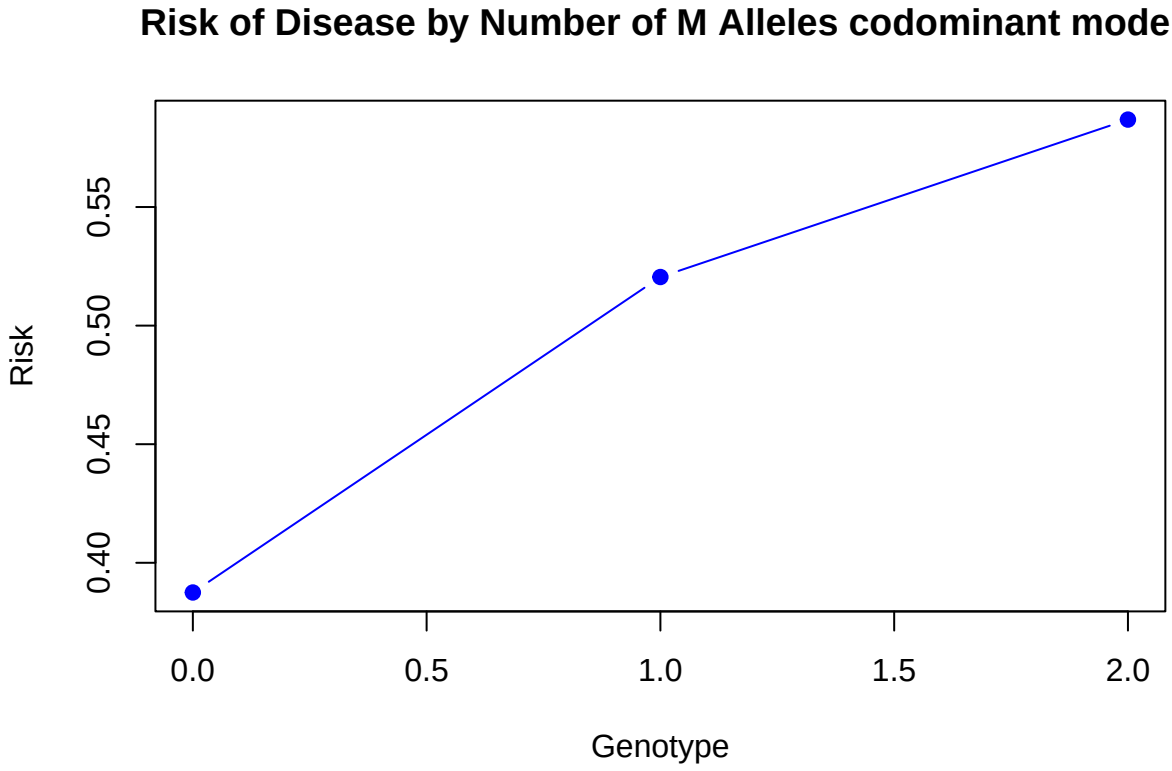
```
## Codomminant test results:
```

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

```
## P-value: 0.0008085403
```

```r
odds_ratios <- c((X[1, 1] / (X[1, 2] + X[1, 3])) / (X[2, 1] / (X[2, 2] + X[2, 3])))
odds_ratios <- c(odds_ratios, (X[1, 2] / (X[1, 1] + X[1, 3])) / (X[2, 2] / (X[2, 1] + X[2, 3])))
odds_ratios <- c(odds_ratios, (X[1, 3] / (X[1, 1] + X[1, 2])) / (X[2, 3] / (X[2, 1] + X[2, 2])))
risk <- odds_ratios / (1 + odds_ratios)

allele_counts <- c(0, 1, 2)
plot(allele_counts, risk, type = "b", pch = 19, col = "blue", main = "Risk of Disease by Number of M Al
```




```r
# Dominant test
Y <- cbind(X[,1],X[,2]+X[,3])
colnames(Y) <- c("AA","Aa or aa")
rownames(Y) <- c("Cases","Control")
test <- chisq.test(Y)
p_value <- test$p.value
cat("Dominant test results:\n")
```

```
## Dominant test results:
```

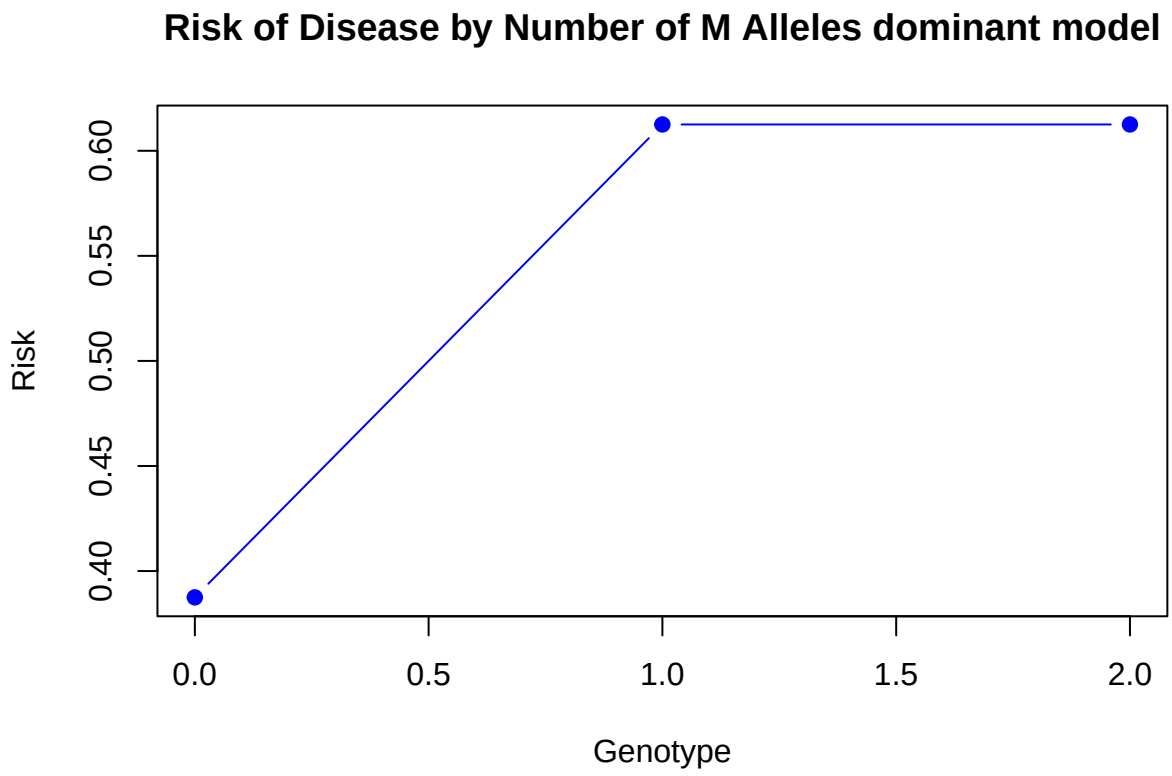
```r
cat("P-value:", p_value, "\n")
```

```
## P-value: 0.0008108124
```

```r
odds_ratios <- c((X[1, 1] / (X[1, 2] + X[1, 3])) / (X[2, 1] / (X[2, 2] + X[2, 3])))
odds_ratios <- c(odds_ratios, (Y[1, 2] / Y[1, 1]) / (Y[2, 2] / Y[2, 1]))
odds_ratios <- c(odds_ratios, (Y[1, 2] / Y[1, 1]) / (Y[2, 2] / Y[2, 1]))

risk <- odds_ratios / (1 + odds_ratios)

allele_counts <- c(0, 1, 2)
plot(allele_counts, risk, type = "b", pch = 19, col = "blue", main = "Risk of Disease by Number of M Al
```




```r
# Recessive test
Y <- cbind(X[,1]+X[,2],X[,3])
colnames(Y) <- c("AA or Aa","aa")
rownames(Y) <- c("Cases","Control")
test <- chisq.test(Y)
p_value <- test$p.value
cat("Recessive test results:\n")
```

```
## Recessive test results:
```

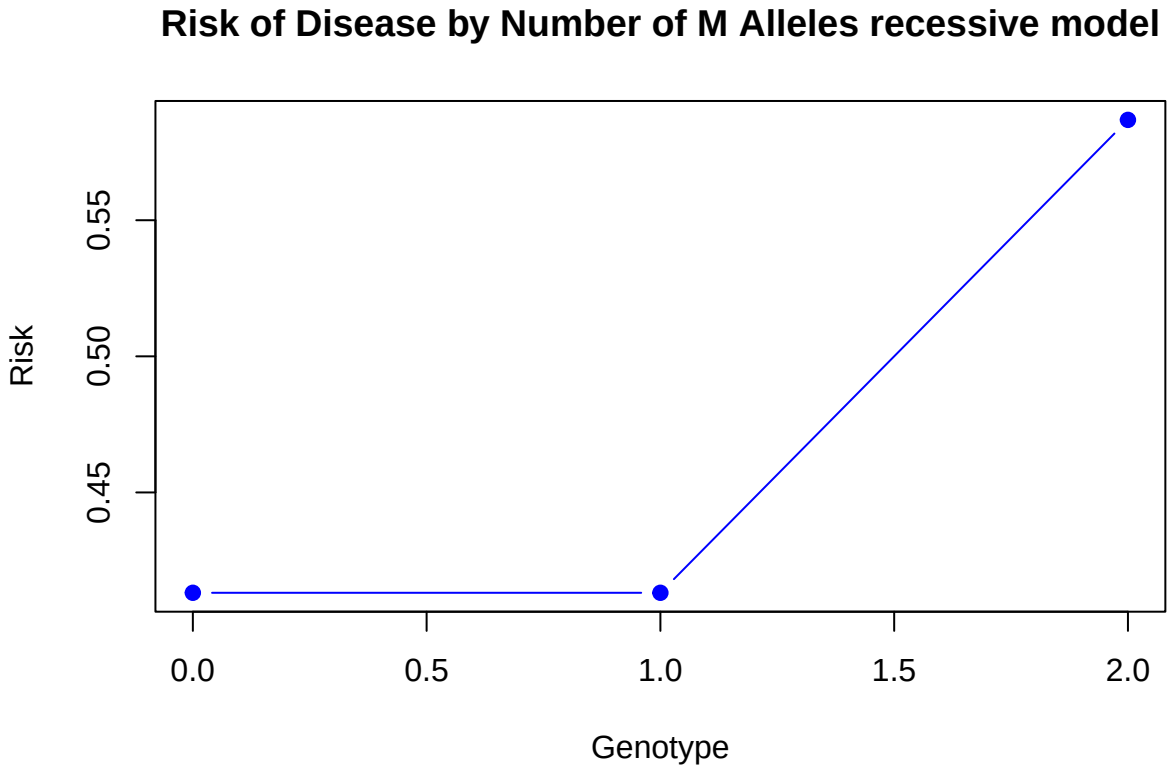
```r
cat("P-value:", p_value, "\n")
```

```
## P-value: 0.009952705
```

```r
odds_ratios <- c((Y[1, 1] / Y[1, 2]) / (Y[2, 1] / Y[2, 2]))
odds_ratios <- c(odds_ratios, (Y[1, 1] / Y[1, 2]) / (Y[2, 1] / Y[2, 2]))
odds_ratios <- c(odds_ratios, (Y[1, 2] / Y[1, 1]) / (Y[2, 2] / Y[2, 1]))

risk <- odds_ratios / (1 + odds_ratios)

allele_counts <- c(0, 1, 2)
plot(allele_counts, risk, type = "b", pch = 19, col = "blue", main = "Risk of Disease by Number of M Al
```



Based on the p-value for the codominant model we can reject the null hypothesis that the risk is the same for all genotypes. For the dominant model there is a p-value as low as for the co-dominant model and we can reject the null hypothesis that the risk of disease doesn't depend on the presence of a as well. The p-value for the recessive model is also very small but factor 10 bigger than the p-values for the other models. But it is still small enough that we can also reject the null hypothesis that the risk does not depend on being homozygote.

All models have significantly low p-values, indicating a strong association between the genetic variant and the studied effect. The codominant and dominant models appear to show slightly stronger associations than the recessive model.

If the codominant model fitted perfectly with our data, it should form a straight line with a null slope. Additionally, in the dominant model, a significant difference in risk is observed when at least one allele a is present. Hence, the dominant model appears to be the better fit.

Question 4

```
Cases <- c(112, 278, 150)
Controls <- c(206, 348, 150)
X <- rbind(Cases,Controls)
rownames(X) <- c("Cases","Controls")
colnames(X) <- c("AA","Aa","aa")
n <- sum(X)
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)

cat("Armitage Trend Tes:\n")
```

```
## Armitage Trend Tes:
```

```
cat("P-value:", pvalue, "\n")
```

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

The p-value of ~0.0002 suggest that we should reject the null hypothesis.

In other words, the significant trend observed in the distribution of genotypes suggests that the different genotypes may be significantly associated with the risk of Alzheimer's disease.

Question 5

The results got we the Armitage trend test proved there is an association for this marker on the disease. Moreover, like we told before the best model was the dominant. We can declare, with the low p-value of the Armitage trend test, the association does exist and that the risk of being sick increases a lot if you have an allele a.