Table 15.2 Genotype and Allele Counts in a Case Control Sample

	Genotypes				Alleles		
	A1A1	*A1A2*	*A2A2*	Total	*A1*	*A2*	Total
Cases	46	34	20	100	126	74	200
Controls	60	30	10	100	150	50	200

parameterization to the linear regression (Section 15.2) in terms of the coding of the genetic information at the locus. The same extensions described above for linear regression apply to logistic regression (e.g., multiallelic markers and covariates).

Example 2: Testing for association using a case-control design

In this example, we test a range of case-control datasets. The total dataset is comprised of one hundred cases and one hundred controls. We assume that the genotype counts are known for cases and controls. The genotype frequencies and allele frequencies in this sample are displayed in *Table 15.2*.

A sequence of *R* scripts is presented below outlining various association tests for case-control data.

```
# To test for differences in genotype frequencies between
# groups,
# read in the matrix with genotype (columns)
# by group (rows) frequencies, in this case a 2 by 3 matrix

geno=matrix(c(46,34,20,60,30,10),2,3, byrow=T)

# A chi-square test can then be used to assess whether the
# genotype frequencies
# differ significantly between the groups.

chisq.test(geno)

X-squared = 5.4324, df = 2, p-value = 0.06613
```

R is storing the counts from *Table 15.2* in a matrix called '*geno*'. Then, the *R* function *chisq.test* calculates a Pearson χ^2 test on the full table. This test is fully saturated and equivalent to a model for both additive and dominance effects. The test for the full model is not significant ($p = 0.066$).

One can also impose a dominance or recessive Mendelian model for the effect of the locus. For dominance effects, subgroups with genotypes A_1A_1 and A_1A_2 are combined, while for recessive effects, genotypes A_1A_2 and A_2A_2 are combined.


```
# To test for dominance, collapse the columns corresponding
# to A1A1 and A1A2
# across groups and perform a chi-square test on this new
# table

dom=matrix(c(geno[1,1],geno[1,2]+geno[1,3],geno[2,1],geno
[2,2] +geno[2,3]),2,2,byrow=T)
chisq.test(dom)

        X-squared = 3.3922, df = 1, p-value = 0.0655

# To test for recessive effect, collapse the columns
# corresponding to A1A2 and A2A2
# across groups and perform a chi-square test on this new
# table

rec=matrix(c(geno[1,1]+geno[1,2],geno[1,3],geno[2,1]+
  geno[2,2],geno[2,3]),2,2, byrow=T)
chisq.test(rec)

        X-squared = 3.1765, df = 1, p-value = 0.0747

# Note that with 2x2 contingency tables, R uses the Yates'
# correction by default.
# This correction is used to prevent overestimation of
# significance in small datasets.
# Other statistical packages, like SPSS, don't use this
# correction.
# The correction can be switched off as follows:

chisq.test(dom,correct=F)

        X-squared = 3.9342, df = 1, p-value = 0.04731

chisq.test(rec,correct=F)

        X-squared = 3.9216, df = 1, p-value = 0.04767
```

The tests for dominance or recessive effects are also not significant ($p = 0.066$ and $p = 0.075$, respectively), unless Yates' correction is switched off, in which case both tests are *just* nominally significant at $\alpha = 0.05$ ($p = 0.047$ and $p = 0.048$, respectively).

Finally, we compare the allele frequencies across groups. To generate the data for this, we sum the number of A_1 and A_2 for cases and controls.