13.7 Concluding Remarks

This chapter has introduced two-way classifications, which are the simplest form of contingency tables. For higher-order classifications, which frequently occur in practice, a greater variety of forms of dependence arise. For example, for a three-way table, the factors of which are denoted A, B, and C, we might consider testing whether, conditionally on C, A and B are independent.

Dependencies can be specified by means of **log linear models.** If the row and column classifications in a two-way table are independent, then

$$\pi_{ij} = \pi_i \pi_j$$

or

$$\log \pi_{ij} = \log \pi_i + \log \pi_j$$

We can denote $\log \pi_i$ by α_i and $\log \pi_j$ by β_j. Then, if there is dependence, $\log \pi_{ij}$ may be written as

$$\log \pi_{ij} = \alpha_i + \beta_j + \gamma_{ij}$$

This form mimics the additive analysis of variance models introduced in Chapter 12. The idea can readily be extended to higher-order tables. For example, a possible model for a three-way table is

$$\log \pi_{ijk} = \alpha_i + \beta_j + \delta_{ij} + \epsilon_{ik} + \gamma_{jk}$$

which allows second-order dependencies, but no third-order dependencies. The parameters of log-linear models may be estimated by mle's and likelihood ratio tests may be employed. Agresti (1996) treats these and other topics in the analysis of categorical data.

13.8 Problems

1. Adult-onset diabetes is known to be highly genetically determined. A study was done comparing frequencies of a particular allele in a sample of such diabetics and a sample of nondiabetics. The data are shown in the following table:

	Diabetic	Normal
Bb or *bb*	12	4
BB	39	49

   Are the relative frequencies of the alleles significantly different in the two groups?

2. Phillips and Smith (1990) conducted a study to investigate whether people could briefly postpone their deaths until after the occurrence of a significant occasion. The senior woman of the household plays a central ceremonial role in the Chinese