8

Categorical Data

In many experiments and in almost all surveys, many if not all the results fall into discrete categories rather than being measurable on a continuous scale: e.g., male vs. female, African–American vs. Hispanic vs. Asian vs. white, in favor vs. against vs. undecided. The corresponding hypotheses concern proportions: ("African–Americans are as likely to be Democrats as they are to be Republicans." "The dominant genotype 'spotted shell' occurs with three times the frequency of the recessive." "The data from all 14 treatment sites may be combined as the effects of treatment are identical at each site." In this chapter you will learn methods for testing these hypotheses. The techniques you will learn also are applicable when your measurements are ordinal but not metric, as with preference ratings: e.g. "Do you prefer the acting of Toni Colette to Nicole Kidman? Strongly prefer? Slightly prefer? Indifferent?"

8.1 Fisher's Exact Test

Suppose, upon examining the cancer registry in a hospital, we uncover the data that we put in the form of a 2×2 contingency table, Table 8.1.

The 9 denotes the number of males who survived cancer, the 1 denotes the number of males who died from the disease, and so forth. The four marginal totals or *marginals* are 10, 14, 13, and 11. The total number of men in the study is 10, while 14 denotes the total number of women, and so forth.

We see in this table an apparent difference in the survival rates for men and women: Only 1 of 10 men died following treatment, but 10 of the 14 women failed to survive. Is this difference statistically significant?

The answer is "yes" if the data represent a random sample of cancer patients. Let's see why, using the same line of reasoning that Fisher advanced at the annual Christmas meeting of the Royal Statistical Society in 1934. After Fisher's talk was concluded, incidentally, a seconding speaker compared Fisher's talk to "the braying of the Golden Ass." I hope the reader will take

Table 8.1.

	Survived	Died	Total
Men	9	1	10
Women	4	10	14
Total	13	11	24

more kindly to my own explanation. Fisher's test maximizes the minimum power [Tocher, 1950].

The marginals in this table are fixed because, indisputably, there are 11 dead bodies among the 24 persons in the study and 14 women. Suppose that before completing the table, we lost the subject identification labels so that we could no longer identify which subject belonged in which category. Imagine you are given two sets of 24 labels. The first set has 14 labels with the word "woman" and 10 labels with the word "man." The second set of labels has 11 labels with the word "dead" and 13 labels with the word "alive." Under the null hypothesis, you are allowed to distribute the labels to subjects independently of one another, one label from each of the two sets per subject.

The following two tables are the result of this relabeling procedure. The first of these tables could make a strong case for the superior fitness of the male, stronger even than our original observations. In the second table, the survival rates for men and women are more alike than they were in our original table.

There are a total of $N = \sum_{x=0}^{10} \binom{13}{x}\binom{11}{10-x} = \binom{24}{10}$ ways you could hand out the labels. $\binom{14}{10}\binom{10}{1}$ of the assignments result in tables that are as extreme as our original table (that is, in which 9 of the men survive), and $\binom{14}{11}\binom{10}{0}$ in tables that are more extreme (all 10 of the men survive). This is a very small fraction of the total, so we conclude that a difference in survival rates of the two sexes as extreme as the difference we observed in our

Table 8.2a.

	Survived	Died	Total
Men	10	0	10
Women	3	11	14
Total	13	11	24

Table 8.2b.

	Survived	Died	Total
Men	8	2	10
Women	5	9	14
Total	13	11	24