CHAPTER 1

Randomization

1.1 The Idea of a Randomization Test

Many hypotheses of interest in science can be regarded as alternatives to null hypotheses of randomness. That is, the hypothesis under investigation suggests that there will be a tendency for a certain type of pattern to appear in data, whereas the null hypothesis says that if this pattern is present, then this is a purely chance effect of observations in a random order.

Randomization testing is a way of determining whether the null hypothesis is reasonable in this type of situation. A statistic S is chosen to measure the extent to which data show the pattern in question. The value s of S for the observed data is then compared with the distribution of S that is obtained by randomly reordering the data. The argument made is that if the null hypothesis is true, then all possible orders for the data were equally likely to have occurred. The observed data order is then just one of the equally likely orders, and s should appear as a typical value from the randomization distribution of S. If this does not seem to be the case (so that s is significant), then the null hypothesis is discredited to some extent and, by implication, the alternative hypothesis is considered more reasonable.

The significance level of s is the proportion or percentage of values that are as extreme or more extreme than this value in the randomization distribution. This can be interpreted in the same way as for conventional tests of significance. If s is less than 5%, then this provides some evidence that the null hypothesis is not true; if s is less than 1%, then it provides strong evidence that the null hypothesis is not true; and if s is less than 0.1%, then it provides very strong evidence that the null hypothesis is not true. To avoid the characterization of belonging to "that group of people whose aim in life is to be wrong 5% of the time" (Kempthorne and Doerfler, 1969), it is better to regard the level of significance as a measure of the strength of evidence