Failing to account for or balance extraneous factors can lead to major errors in interpretation. In this chapter, you’ll learn to block or measure all factors that are under your control and to use random assignment to balance the effects of those you cannot. You’ll learn to design experiments to investigate multiple factors simultaneously, thus obtaining the maximum amount of information while using the minimum number of samples.

8.1. Noise in the Data

In Chapter 2, we encountered a woman anxious to demonstrate her abilities as a tea taster. We argued in that chapter that by removing all extraneous sources of variation—e.g., the appearance of the cup, the temperature of the water, the expression on the experimenter’s face—we could focus more narrowly on the factor to be tested.

Eliminating or reducing extraneous variation is the first of several preventive measures we use each time we design an experiment or survey. We strive to conduct our experiments in a biosphere with atmosphere and environment totally under our control. And when we can’t—which is almost always the case—we record the values of the extraneous variables to use them either as blocking units or as covariates.

8.1.1. Blocking

Although the significance level of a permutation test may be distribution-free, its power (defined in Section 6.1.3) strongly depends on the underlying distribution.
In Figure 8.1, we see that the more variable our observations, the less the power of our tests and our ability to detect the alternative. One way to reduce the variance is to block the experiment, that is, to subdivide the population into more homogeneous subpopulations and to take separate independent samples from each.

Suppose you were designing a survey on the effect of income level on the respondent's attitude toward compulsory pregnancy. Obviously, the views of men and women differ markedly on this controversial topic. To reduce the variance and increase the power of your tests, block the experiment, interviewing and reporting on men and women separately. A physician would want to block by gender in a medical study, and probably by age and race as well. An agronomist would want to distinguish among clay soil, sand, and sandy-loam.

Whenever a population can be subdivided into distinguishable subpopulations, you can reduce the variance of your observations and increase the power of your statistical tests by blocking or stratifying your sample.

Suppose we have agreed to divide our sample into two blocks—one for men and one for women. If this is an experiment, rather than a survey, we would then assign subjects to treatments separately and independently within each block. In a study that involves 2 treatments and 10 experimental subjects, 4 men and 6 women, we would first assign the men to treatment and then the women. We could assign the men in any of $\binom{4}{2} = 6$ ways and the women in any of $\binom{6}{2} = 20$ ways, for a total of $6 \times 20 = 120$ possible random assignments.

When we come to analyze the results of our experiment, we use the permutation approach to ensure that we analyze the way the experiment was designed. Our test statistic is a natural extension of that used for the two-sample comparison in that