against the null hypothesis, rather than showing whether the data are significant at a certain level.

In comparison with more standard statistical methods, randomization tests have two main advantages. First, they are valid even without random samples. Second, it is often relatively easy to take into account the peculiarities of the situation of interest and use nonstandard test statistics.

There is a disadvantage with randomization tests that may appear at first sight to be severe: it is not necessarily possible to generalize the conclusions from a randomization test to a population of interest. What a randomization test tells us is that a certain pattern in data is or is not likely to have arisen by chance. This is completely specific to the data at hand. The concept of a population from which other samples could be taken is not needed, which is why random sampling is not required.

Some statisticians argue that the lack of a theory for generalizing the results of randomization tests to populations means that these tests have very little value, if any, in comparison to more standard tests for which well-developed methods of statistical inference exist. Others, however, suggest that in reality samples are often not really random at all, but simply consist of items that happen to be readily available. The generalization of results then rests on the assumption that the sample obtained is effectively the same as a random sample. This nonstatistical judgment is similar to the type of judgment that is made when deciding that the result of a randomization test is what can generally be expected for data collected in a particular way.

As an example, suppose that a physiologist wishes to see whether drinking alcohol in moderation has an effect on reaction times of subjects aged 20. Rather than take a random sample of all possible subjects of this age (which leads to considerable difficulties about the definition of the population, and is in any case impossible), he uses all the 20-year-old students in a university class in physiology. These are divided at random into two groups: one has reaction times measured after taking a drink with a small amount of alcohol, and the other has reaction times measured after taking an alcohol-free drink.

Various methods can be used to analyze the results of an experiment of this type. For example, if a mean difference between the test scores for the two groups is of interest, then a conventional t-test can be used to determine whether the observed difference is significantly different from zero. However, whatever the outcome of such a test is, using it to draw conclusions about the effect of alcohol on all 20-year-olds is only valid on the assumption that the 20-year-olds in the