physiology class are equivalent to a random sample of all 20-year-olds with respect to the measurement of reaction times that is used. Hence, any such generalization has to be questionable, and requires a judgment as to whether the same type of result is likely to occur again if a different group of subjects is tested.

It seems clear that one experiment of this type will not give a definitive result, no matter how many subjects are used. However, if the experiment is repeated on other groups (law students, factory workers, office workers, etc.) and the results always come out about the same, then most people would believe that the effect (or lack of an effect) seen is common to all 20-year-olds. In other words, in the absence of truly random samples, convincing evidence of an effect requires it to be demonstrated consistently at different times in different places. This is a nonstatistical type of inference that works equally well with conventional and randomization tests.

Another point is that in many situations either the concept of a population is irrelevant or the data can be considered as representing the whole population. Thus, an example that is considered in the next section concerns the relationship between the world distribution of earwigs and the positions of the continents. There is only one distribution of earwigs that exists, and one set of continents, so the data are not samples from populations except in a most unrealistic sense. Another example in the next section addresses the question of whether there is a cycle in the times of mass extinctions of animals and plants in the geologic past. Here only one extinction record exists, and the concept of this being a random record from a population of possible records is again artificial.

Although random samples are not necessarily required to justify randomization tests, there are times when they do provide the justification. For example, in the reaction time experiment there would be no need to divide the subjects at random into two groups if initially there were two random samples available from the population of 20-year-olds. In that case, either group could be the one given the alcohol, and a valid comparison between the test scores of the two groups to examine the effect of alcohol could be made using a randomization test or a more conventional alternative.

Randomization tests are most easily justified if either the samples being analyzed are random or the experimental design itself justifies randomization testing. This has led some authors (e.g., Kempthorne and Doerfler, 1969) to use the description permutation tests for situations where random samples justify the calculations, and randomization tests for situations where the experimental design provides