

Investigating the Randomness of Numbers

Introduction

Setting the background: A random sample of 50 students from a school with a population of 1000 is to be selected. Each student is assigned a random number from 1 to 1000, and a random sample of those numbers is drawn without replacement.

Random selection is an important aspect of statistical tests. Tables of random numbers and programs for generating random numbers are commonly used in the process of random selection. Using statistical tests, is it possible to ascertain whether such numbers are likely represent a random selection? The key to assessing whether randomness is likely to exist lies in determining whether the outcome aligns with expectations. In this and the following investigation you will use statistical techniques to evaluate the reasonableness of assuming whether a selection of numbers purported to be random is likely to be so.

The following is a trick question: If randomly selecting four numbers without replacement from the set of numbers $\{1, 2, 3, \dots, 10\}$, which of the following outcomes is *more* likely to occur: $\{1, 2, 3, 4\}$, or $\{7, 4, 9, 2\}$? Some may perceive the first set to be too contrived or that the second set appeals to their general notion of randomness. In either case, they may judge that the second set is more likely to occur. This is a trick question because the outcomes are equally likely. No single outcome of four numbers is more or likely to occur than any other.

The following is *not* a trick question: If randomly selecting four numbers without replacement from the set of numbers $\{1, 2, 3, \dots, 10\}$, which of the following outcomes is *more* likely to occur: all numbers are less than 5, or *not* all numbers are less than 5? (Notice that the second outcome is not the same as “all numbers not less than 5.”)

The probability of the former is $(4/10) \times (3/9) \times (2/8) \times (1/7)$, or $1/210$. The probability of the latter is $1 - 1/210$, or $209/210$. The latter condition can be satisfied by many more outcomes than the former and is thus much more likely to occur. When assessing the randomness of numbers we are not trying to determine whether a specific outcome is likely to have occurred but whether that outcome is representative of those outcomes *most likely* to occur when numbers are truly selected randomly.

Sampling numbers randomly is no *guarantee* that the sample is similar to the population. However, if a sample is truly random it is reasonable to *expect* that the sample is similar to the population in some respects. If not, the nature of the selection method may be suspect. Whether expectations are met will be assessed using two different statistical tests. It will be shown that neither test alone may be sufficient to assess the reasonableness of a decision regarding randomness and that students should not perform tests in a vacuum. Human observation may reveal what tests alone conceal.