JOURNAL OF APPLIED STATISTICS
2019, VOL. 46, NO. 12, 2164–2179
https://doi.org/10.1080/02664763.2019.1581147

Taylor & Francis
Taylor & Francis Group

Generalized run tests for statistical process control

Young H. Chun

Department of Information Systems and Decision Sciences, E. J. Ourso College of Business, Louisiana State University, Baton Rouge, LA, USA

ABSTRACT

In a sequence of elements, a *run* is defined as a maximal subsequence of like elements. The number of runs or the length of the longest run has been widely used to test the *randomness* of an ordered sequence. Based on two different sampling methods and two types of test statistics used, run tests can be classified into one of four cases. Numerous researchers have derived the probability distributions in many different ways, treating each case separately. In the paper, we propose a *unified* approach which is based on recurrence arguments of two mutually exclusive sub-sequences. We also consider the sequence of nominal data that has more than two classes. Thus, the traditional run tests for a binary sequence are special cases of our generalized run tests. We finally show that the generalized run tests can be applied to many quality management areas, such as testing changes in process variation, developing non-parametric multivariate control charts, and comparing the shapes and locations of more than two process distributions.

ARTICLE HISTORY
Received 25 February 2018
Accepted 1 February 2019

KEYWORDS
Non-parametric statistics; run test; statistical process control

1. Introduction

In many experiments or observational studies, each element in the sample space can assume only one of two possible outcomes, such as heads or tails, success or failure, or up or down. The order in which the elements of the sample were drawn is frequently available. In any ordered sequence of n elements of two kinds, a 'run' is defined as a succession of similar events preceded and succeeded by different elements. The number of elements in a run is referred to as its length. We can also count the total number of runs in an ordered binary sequence.

Consider, for example, the ordered sequence of 10 binary numbers, 1 1 1 1 0 1 1 0 0 1. The total number of *runs* in the sequence is five (i.e. 1 1 1 1 0 1 1 0 0 1), among which the number of 1 runs is three and the number of 0 runs is two. In the same sequence, the *length* of the longest 1 run is four (i.e. 1 1 1 1), whereas the *length* of the longest 0 run is two (i.e. 0 0). Thus, the *maximum run length* (i.e, the length of the longest 1 or 0 run) is four in the binary sequence.

Various types of *run tests* have been proposed to examine whether or not a two-valued data sequence is randomly generated. The test statistic is often derived from (i) the number

CONTACT Young H. Chun chun@lsu.edu Department of Information Systems and Decision Sciences, E. J. Ourso College of Business, Louisiana State University, Baton Rouge, LA 70803, USA