1.3.5.13. Runs Test for Detecting Non-randomness

Purpose: Detect Non-Randomness

The runs test (Bradley, 1968) can be used to decide if a data set is from a random process.

A run is defined as a series of increasing values or a series of decreasing values. The number of increasing, or decreasing, values is the length of the run. In a random data set, the probability that the (*I*+1)th value is larger or smaller than the *I*th value follows a binomial distribution, which forms the basis of the runs test.

Typical Analysis and Test Statistics

The first step in the runs test is to count the number of runs in the data sequence. There are several ways to define runs in the literature, however, in all cases the formulation must produce a dichotomous sequence of values. For example, a series of 20 coin tosses might produce the following sequence of heads (H) and tails (T).

H H T T H T H H H H T H H T T T T T H H

The number of runs for this series is nine. There are 11 heads and 9 tails in the sequence.

Definition

We will code values above the median as positive and values below the median as negative. A run is defined as a series of consecutive positive (or negative) values. The runs test is defined as:

H_0:	the sequence was produced in a random manner
H_a:	the sequence was not produced in a random manner
Test Statistic:	The test statistic is

$$Z = \frac{R - \bar{R}}{s_R}$$

where R is the observed number of runs, $\bar{R}$, is the expected number of runs, and s_R is the standard deviation of the number of runs. The values of $\bar{R}$ and s_R are computed as follows:

$$\bar{R} = \frac{2n_1n_2}{n_1+n_2} + 1$$

$$s_R^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$$

with n_1 and n_2 denoting the number of positive and negative values in the series.

Significance Level: α

Critical Region: The runs test rejects the null hypothesis if

$|\boldsymbol{Z}| > \boldsymbol{Z}_{1-\alpha/2}$

For a large-sample runs test (where $n_1 > 10$ and $n_2 > 10$), the test statistic is compared to a standard normal table. That is, at the 5 % significance level, a test statistic with an absolute value greater than 1.96 indicates non-randomness. For a small-sample runs test, there are tables to determine critical values that depend on values of n_1 and n_2 (Mendenhall, 1982).

Runs Test Example

A runs test was performed for 200 measurements of beam deflection contained in the LEW.DAT data set.

```
H0:  the sequence was produced in a random manner
Ha:  the sequence was not produced in a random manner


Test statistic:  Z = 2.6938
Significance level:  α = 0.05
Critical value (upper tail):  Z1-α/2 = 1.96
Critical region: Reject H0 if |Z| > 1.96
```

Since the test statistic is greater than the critical value, we conclude that the data are not random at the 0.05 significance level.

Question

The runs test can be used to answer the following question:

- Were these sample data generated from a random process?

Importance

Randomness is one of the key assumptions in determining if a univariate statistical process is in control. If the assumptions of constant location and scale, randomness, and fixed distribution are reasonable, then the univariate process can be modeled as:

$$Y_i = A_0 + E_i$$

where $\boldsymbol{E_i}$ is an error term.

If the randomness assumption is not valid, then a different model needs to be used. This will typically be either a times series model or a non-linear model (with time as the independent variable).

Related Techniques	Autocorrelation Run Sequence Plot Lag Plot
Case Study	Heat flow meter data
Software	Most general purpose statistical software programs support a runs test. Both Dataplot code and R code can be used to generate the analyses in this section. These scripts use the LEW.DAT data file.