Runs Test for Serial Randomness of Nominal Data

Equations taken from Zar, 1984

Example: the runs test is used to determine for serial randomness: whether or not observations occur in a sequence in time or over space. In geographic studies the runs test is most often used to determine whether observations are random along a transect or other linear feature. In the example below fish were sampled along a river at equal intervals, resulting in the data set:

A A B B A A B B B B A A A B B B B A A B B B

Where **A** denotes rainbow trout and **B** denotes brown trout. We are interested in determining whether the order of the two species of trout is random or not, as opposed to the species forming groups such as:

A A A A A A A B B B B B B B

or members of one species shunning other members of that species, such as:

A B A B A B A B A B A B A B

Unlike other tests there is no equation for the runs test unless the sample size of either group is greater than 30*. One only needs to count the number of runs (***u***), a run being a series of the same nominal value when counting from left to right.

A A A B B A A ...

This is a run

Two Tailed Runs Test

H_0 : The distribution of trout along the stream is random.
H_a : The distribution of trout along the stream is not random.

A A B B A A B B B B A A A B B B B A A B B B

$n_1 = 9$ ← there are 9 occurrences of the value **A**.
$n_2 = 13$ ← there are 13 occurrences of the value **B**.
$\boldsymbol{u} = 8$ ← there are 8 runs.

$\alpha = 0.05$
$\boldsymbol{u}_{\text{Critical}} = 6, 17$ ← there are 2 critical values of u, if the calculated value falls between these then H_0 is accepted.

Since $6 < 8 < 17$ accept H_0

The distribution of trout along the stream is random ($\boldsymbol{u}_8$, $0.50 > p > 0.20$).

* In which case a z score can be calculated using the following equations:

$$\bar{x} = \frac{2n_1 n_2}{n_1 + n_2} + 1 \qquad s = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}} \qquad z = \frac{|u - \bar{x}| - 0.5}{s}$$

This z score can then be compared the probabilities found in the z table.

If a one tailed runs test is used, we can determine whether the data are either random, non-random due to clustering, or non-random due to uniformity. Again, there are 2 critical values…

If u < the lower $u_{Critical}$ then the data are non-random due to clustering.
If u > the upper $u_{Critical}$ then the data are non-random due to uniformity.
If u falls between the lower and upper $u_{Critical}$ then the data are random.

One Tailed Runs Test

H_0 : The distribution of trout along the stream is random.
H_a : The distribution of trout along the stream is not random.

AAAAABBBBAAAAABBBAAAAAA

$n_1 = 16$ ← there are 16 occurrences of the value **A**.
$n_2 = 7$ ← there are 7 occurrences of the value **B**.
$\boldsymbol{u} = 5$ ← there are 5 runs.

$\alpha = 0.05$
$\boldsymbol{u}_{Critical} = 6, 15$ ← there are 2 critical values of u, if the calculated value falls between these then H_0 is accepted.

Since $5 < 6 < 15$ reject H_0

The distribution of trout along the stream is non-random due to clustering ($\boldsymbol{u}_5$, $p < 0.025$).