Tests for Randomness-The Runs Test

The simplest time series is a random model, in which the observations vary around a constant mean, have a constant variance, and are probabilistically independent. In other words, a random time series has not time series pattern. Observations do not trend upwards or downwards, the variance does not increase over time, the observations do not tend to be bigger in some periods than in other periods.

A random model can be written as

$$Y_t = m + e_t$$

Here m is a constant, the average of the $Y_t\text{'}s$, and e_t is the residual (or error) term which is assumed to have a zero mean, a constant variance, and to be probabilistically independent.

There are two situations where a random time series occur.

1. The first is when the original time series in random.
2. The second is when we fit a model to a time series to obtain an equation like:
   $Y_t =$ fitted part + residual

The second situation is the most common. What we wish to do is to model Y_t as a fitted part plus noise, where the fitted part includes any forecastable pattern in the series, and the noise is impossible to model any further.

How do we tell whether or not a series is random?

1. Plot it and see whether or not there is any pattern.
2. Control charts-individual observations are plotted, with a centerline equal to $\bar{Y}$ and the control limits usually set at $\pm 3s$. If the series is in control then there should not be any points beyond the control limits nor should there be any patterns.
3. The Runs Test.
   For each observation associate a 1 if $Y_t > \bar{Y}$ and a 0 otherwise. The series then has an associated series of 1's and 0's. A run is a consecutive sequence of 0's or 1's. A run's test check if the number of runs is the correct number for a series that is random.

To figure this out let T be the number of observations, T_A be the number above the mean and T_B be the number below the mean. Let R be the observed number of runs. Then using combinatorial methods, the probability $P(R)$ can be established and the mean and variance of R can be derived:

$$E(R) = \frac{T + 2T_A T_B}{T}$$

$$V(R) = \frac{2T_A T_B (2T_A T_B - T)}{T^2 (T - 1)}$$

When T is relatively large (>20) the distribution of R is approximately normal and thus

$$Z = \frac{R - E(R)}{Stdev(R)} \sim N(0,1)$$