measurements of the celestial coordinates of a star are unlikely to have been made by the same observer with the same instrumentation in the sa me observing conditions, so the results were probably formed of two or more superimposed normal distributions with similar means but di erent variances. Gross errors might occasionally arise too,

impossible to prove or disprove with much conviction that a data sample is taken from a normal population if the number of measurements is small.

Significance tests for normality

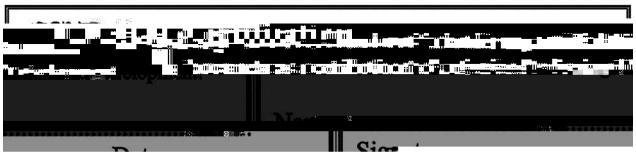
Despite the assurances provided by the CLT there are many occasions when it is desirable to test to see whether data sets might be normally distributed. Important examples include the testing of residuals found in regression and analysis of variance (ANOVA) calculations. If the residuals derived from a regression plot are not randomly and normally distributed, the model used (e.g, linear, quadratic) is probably inappropriate. In ANOVA the calculated probability values rely on the assumption that the random errors re ected in the residuals are normally distributed.

Several established methods for testing for normality are available in many software packages. A simple approach is the use of normal probability paper, in which the individual measurements are plotted against their cumulative frequencies, the latter being on a non-linear scale derived from the percentage points of the normal distribution. Normally distributed data should yield a straight line plot: the Ryan-Joiner (RJ) test provides a correlation coe cient to evaluate the linearity. Fig. 2 shows a normal probability plot for a set of 8 measurements: the points lie

Further reading

A clear treatment of methods of testing for the normal distribution is provided in Practical Statistics for Medical Research D. G.Altman, Chapman and Hall, 1991.

An example of a Shapiro–Wilk calculation using Excel® is provided at http://www.real-statistics.com. Downloaded February 2nd 2017.



James N. Miller

This Technical Brief was reviewed for the Statistical Subcommittee and approved by the Analytical Methods Committee on 11/07/17.