Announcements and correspondence

Errata

The editors apologise for minor errors in **'Comments on the Spirit Level Controversy'** by *Hugh Noble.*

In the 3rd paragraph on page 52 #104, the letter 's' was omitted as shown below.

> However, most curve fitting procedures are based on the assumption that residuals (the difference**s** between the actual position**s** of plotted points and the corresponding position**s** on the regression curve) are due to random errors of measurement, and that the magnitudes of these errors have a normal distribution. That is the rationale behind the method of least squares approach which is fundamental to most curve-fitting procedures. The justification of that assumption is based on the Theorem of Central Limits. That theorem and its conclusions assume that the total error of any measurement is the sum total of a very large number of very small random errors. In these circumstances a binomial distribution can be assumed. In practical circumstances and for large numbers of trials, the binomial distribution is so close to a normal distribution that we can ignore the discrepancy.

Correspondence

The following contribution has been received from Roger Boyle that relates to the same article.

> I enjoyed Hugh Noble's piece on the correlation issues behind the "Spirit Level" debate [Newsletter 104].
>
> Line and curve fitting are almost always approached on a Least-Squares basis – there's a very nice article [Sorensen 1970] putting this into good historical perspective (nice one Gauss, again). Outliers, of course, are very neatly considered by RANSAC [Fischler & Bolles, 1980].
>
> Natural [sic] systems are actually unusual if they demonstrate linearity ($y = ax+b$). On the other hand, many such systems do demonstrate log-log linearity [a power law, $y = \alpha x^{-\beta}$], at least in usable ranges. A wide range of examples exists – scholarly co-

authorship, Zipf's Law, actor co-appearance, WWW page "popularity" are obvious examples [Wasserman and Faust, 1994]. Graph theorists have only relatively recently modelled this phenomenon successfully [Barabasi and Albert, 1999].

A lovely paper came out of Northwestern University (IL.) recently in which Hecht [Hecht, 2009] data-mined Wikipedia to "evidence" Tobler's First Law of Geography – *Everything is related to everything else, but near things are more related than distant things* [Tobler, 1970] – as just such a power law.

References:

A-L Barabasi and R Albert, Emergence of Scaling in Random Networks, Science, 286(5439), 509-512, 1999.

M A Fischler and R C Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, Comms. of the ACM 24 (6): 381–395, 1981.

B Hecht and E Moxley, Terabytes of Tobler: Evaluating the first law in a massive, domain-neutral representation of world knowledge, Proc. 9th Intl. Conf. Spatial Information Theory, (eds. K S Hornby et al.), COSIT 2009, Aber Wrac'h, LNCS5756, 88-105, Springer, September 2009.

H W Sorenson, Least-squares estimation: from Gauss to Kalman, IEEE Spectrum, 7, 63-68, 1970.

W Tobler, A computer movie simulating urban growth in the Detroit region, Economic Geography, 46(2), 234-240, 1970

S Wasserman and K Faust, Social Network Analysis, Cambridge Univ. Press, 1994.