Cellular Oncology 31 (2009) 501
DOI 10.3233/CLO-2009-0506
IOS Press

Letter to the Editor

How to calculate the goodness-of-fit of a fractal dimension

To the Editor,

Karyometric data are helpful for differential diagnosis and prognosis [1–4]. A new morphometric variable, the coefficient of determination of the Minkowski fractal dimensions (the R^2 value between the y of the regression lines and y values of the real data) has recently been introduced [1,2]. Adam et al. [1] described a negative correlation between the fractal dimension (FD) and the R^2 values. Yet at least part of this effect may under certain assumptions be expected for theoretical reasons.

If the residuals of the regression are the y-residuals (i.e., log(fractal area)) – as is common in Least Squares regression – R^2 will vary systematically as the slope of the regression line (i.e., FD) changes, even though the distribution of observed points around the regression line may be the same. However, if the residuals were computed orthogonally to the regression line, at least some of this effect would vanish.

A simple simulation of a point distribution can be made: homogeneous random on an interval $[-L, L]$ in the x-direction, normal random in the y-direction, $N(0, \sigma)$. Then rotate the point distribution around the origin to a given mean slope, $-\phi$, and alter the slope stepwise by an angle $\Delta\phi$ around the mean slope, and compute the R^2 value for each step. The numerical value of the correlation between R^2 and $FD = 2 + \mathrm{tg}(\phi)$ depends on the parameters used, σ/L and the range of ϕ. For $FD < 2.5$ it is positive, for $FD > 2.5$ it is negative, since R^2 peaks at a slope of 45 degrees. This is an obvious result if all the rotated distributions are identical except for the rotation angle. But the correlation is still there – although the numerical value is lower – if the distributions at each separate angle are allowed to be independent.

I do not deny that the coefficient of determination of the Minkowski fractal dimensions might be a statistically significant prognostic variable, or that there may be a real variation in the distribution around the regression line as FD varies. I am simply commenting that the variation of R^2 described in [1,2] may be partially due to a projection effect in the residuals, and that an improved measure could be obtained by a regression based on orthogonal residuals.

It is worth pointing out that orthogonal regression or Total Least Squares is generally appropriate when there is no natural distinction between predictor and response variables, or when all variables are measured with error, as is usually the case when estimating fractal dimension in digital images. This in contrast to the usual regression assumption that predictor variables are measured exactly, and only the response variable has an uncertainty or error component.

Professor Fritz Albregtsen
Department of Informatics, University of Oslo,
P.O. Box 1080 Blindern, N-0316 Oslo, Norway
E-mail: fritz@ifi.uio.no

References

[1] R.L. Adam, R.C. Silva, F.G. Pereira, N.J. Leite, I. Lorand-Metze and K. Metze, The fractal dimension of nuclear chromatin as a prognostic factor in acute precursor B lymphoblastic leukemia, *Cellular Oncology* **28** (2006), 55–59.

[2] R.C. Ferreira, P.S. de Matos, R.L. Adam, N.J. Leite and K. Metze, Application of the Minkowski–Bouligand fractal dimension for the differential diagnosis of thyroid follicular neoplasias, *Cellular Oncology* **28** (2006), 331–333.

[3] B. Nielsen, F. Albregtsen and H.E. Danielsen, A review of statistical nuclear texture analysis in cancer research, *Cellular Oncology* **30** (2008), 165.

[4] B. Nielsen, F. Albregtsen, W. Kildal and H.E. Danielsen, Prognostic classification of early ovarian cancer based on very low dimensionality adaptive texture feature vectors from cell nuclei from monolayers and histological sections, *Analytical Cellular Pathology* **23** (2001), 75–88.