Parabola Volume 46, Issue 1 (2010)

Randomness and Clustering

Peter Donovan and John Steele[1]

Some years ago an office building in Brisbane acquired considerable notoriety for the rate at which cancer had been detected in those who had worked in it. This phenomenon may have been caused by some carcinogenic agent, presumably introduced in the construction phase, or it may have been an unusually large cluster that originated from a random distribution of the health problem.

This note cannot enter into an analysis of the epidemiology of the building or even attempt to define randomness. However modern mathematical computing facilities do have the capacity to generate random (perhaps pseudo-random would be a more appropriate word) numbers in a given range. For example, 100,000 points may be placed uniformly spaced on a circle and the computer could then be used to select 40 of them. Likewise the machine can be used to select points randomly on discs of the same radius.

This task was carried out four times for circles and four times for discs to produce the output shown in the following diagrams. All show clustering: the points do not spread out uniformly but tend to clump together.

As a check, the programs were rerun with a different random number generator. The diagrams so obtained had clustering similar to those displayed.

The point here is that since clustering occurs in mathematical models of random behaviour, we can expect it to occur in nature. Deciding whether a cluster is a statistical effect or something sinister is part of the discipline of statistical inference.

[1]Dr Peter Donovan is a Visiting Fellow and Dr John Steele is a Senior Lecturer in the School of Mathematics and Statistics at the University of New South Wales.

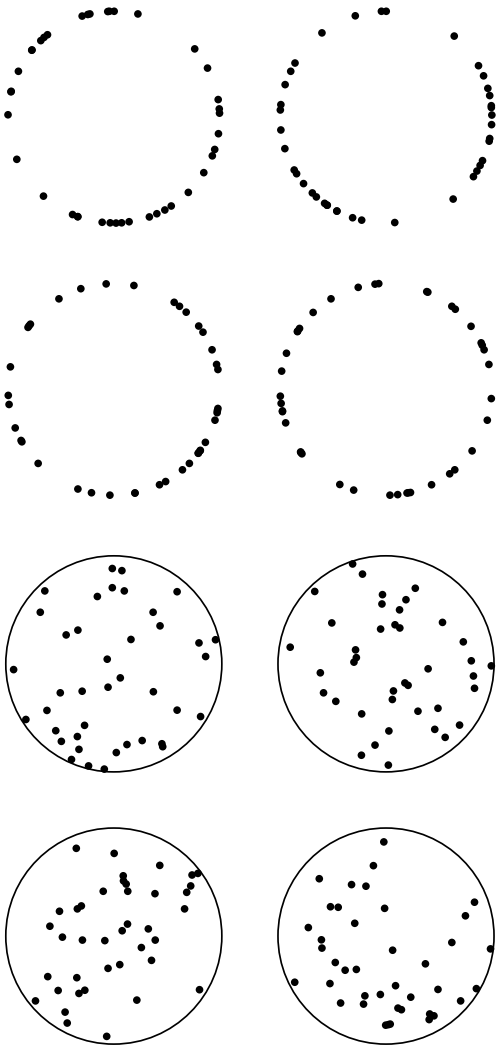

Figure 1: The figure shows clustering in random distributions of points on a circle and in a disc.