

Mathematicians develop new statistical indicator

August 9 2019



Credit: CC0 Public Domain

Most of us know this phenomenon only too well: when it's hot outside, you feel the urge for a cooling ice cream. But would you have thought that mathematics could be involved?

Let us explain: The rising temperatures and rising ice consumption are two statistical variables in linear dependence; they are correlated.

In statistics, correlations are important for predicting the future behaviour of variables. Such scientific forecasts are frequently requested by the media, be it for football or election results.

To measure linear dependence, scientists use the so-called correlation coefficient, which was first introduced by the British natural scientist Sir Francis Galton (1822-1911) in the 1870s. Shortly afterwards, the mathematician Karl Pearson provided a formal mathematical justification for the correlation coefficient. Therefore, mathematicians also speak of the "Pearson product-moment correlation" or the "Pearson correlation."

If, however, the dependence between the variables is non-linear, the correlation coefficient is no longer a suitable measure for their dependence.

René Schilling, Professor of Probability at TU Dresden, emphasises "Up to now, it has taken a great deal of computational effort to detect dependencies between more than two high-dimensional variables, in particular when complicated non-linear relationships are involved. We have now found an efficient and practical solution to this problem."

Dr. Björn Böttcher, Prof. Martin Keller-Ressel and Prof. René Schilling from TU Dresden's Institute of Mathematical Stochastics have developed a dependence measure called "distance multivariate." The definition of this new measure and the underlying mathematical theory were published in the leading international journal *Annals of Statistics* under the title "Distance Multivariate: New Dependence Measures for Random Vectors."

Martin Keller-Ressel explains: "To calculate the dependence measure, not only the values of the observed variables themselves, but also their mutual distances are recorded and from these distance matrices, the distance multivariance is calculated. This intermediate step allows for the detection of complex dependencies, which the usual correlation coefficient would simply ignore. Our method can be applied to questions in bioinformatics, where big data sets need to be analysed."

In a follow-up study, it was shown that the classical correlation coefficient and other known dependence measures can be regained as borderline cases from the distance multivariance.

Björn Böttcher concludes by pointing out: "We provide all necessary functions in the package 'multivariance' for the free statistics software R, so that all interested parties can test the application of the new dependence measure."

More information: Björn Böttcher et al. Distance multivariance: New dependence measures for random vectors, *The Annals of Statistics* (2019). [DOI: 10.1214/18-AOS1764](https://doi.org/10.1214/18-AOS1764)

Provided by Dresden University of Technology

Citation: Mathematicians develop new statistical indicator (2019, August 9) retrieved 16 April 2024 from <https://phys.org/news/2019-08-mathematicians-statistical-indicator.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--