BIOSTAT 2017
23rd International Scientific Symposium on Biometrics
Šibenik, Croatia, 7 - 10 June 2017

KEYNOTE LECTURE

Janez Stare
University of Ljubljana, Faculty of Medicine,
Institute for Biostatistics and Medical Informatics,
Ljubljana, Slovenia

janez.stare@mf.uni-lj.si
http://ibmi.mf.uni-lj.si/janez/index.php?lang=en-GB

Topic:

R^2 – a simple statistic simply misunderstood. Still.

Abstract:

The recent misuse of R^2 in a paper in Science is a good reminder of the fact that this, apparently simple, statistic is still often misunderstood.
The Science paper is an example of reading too much into R^2, and the same goes for its interpretation as a measure of goodness of fit. The other side of misunderstanding of R^2 is represented by its many criticisms, most of them unfounded, and those left could be addressed to almost any other statistic. Adding to these the fact that all software packages report a meaningless R^2 in the case of linear regression with zero constant, it becomes obvious that there is a lot of confusion out there when it comes to understanding the message communicated by R^2. The most persistent criticism is about the fact that R^2 depends on distribution of covariates, and the most persisting misunderstanding is about R^2 measuring goodness of fit. In this talk I will briefly revisit the criticisms directed towards R^2 and explain the falseness of the accusations, and then focus on the two points, mentioned above. My message is: a) dependence on the distribution of covariates is not a property, specific to R^2, and b) R^2 is not a measure of goodness of fit. And, to further emphasize this last point, I will conclude with a brief explanation of how to formally test goodness of fit in linear regression.