(Mis)Use of Statistics in Science – Interview with Dr. Rainer Wanke

Dr. Rainer Wanke[1] is a physicist working in the field of experimental particle physics at the University of Mainz, Germany. He is working on the NA62 experiment at the European Organization for Nuclear Research (CERN) in Geneva, Switzerland, which measures ultra-rare K-meson decays. This involves both, particle detector development and the analysis of data taken with those particle detectors. He furthermore teaches statistics for undergraduate students in Mainz.

[1]email: wanke@uni-mainz.de

JUnQ: In your view, how important is statistics in your discipline, also compared to other disciplines, e.g., physics, chemistry, biology, psychology?

Dr. Wanke: A very good knowledge of statistics is indispensable for data analysis in particle physics. As all processes are quantum mechanical, i.e. intrinsically random, statistics is the only way to describe them. It is used for determining probabilities, efficiencies, and limit setting for very rare processes, and last but not least in Monte Carlo simulations of physics processes as decays and detector interactions of particles. The statistical methods used in particle physics are very often highly sophisticated and discussed within the community, with still many papers being published on specific topics as, e.g. limit setting. Since I'm a physicist and not working in another discipline, it is a bit difficult for me to compare the statistics used in particle physics to those in other disciplines. At a glance, it seems to me that other disciplines – in particular medicine and social sciences – do not use nor need as highly specialized methods. However, they also do have other needs like, e.g. determination of regressions, which is of not much use in particle physics.

JUnQ: In your opinion, is it important that only mathematicians teach statistics?

Dr. Wanke: In particle physics, all more recent relevant text books and publications have been written by particle or nuclear physicists. In my opinion, mathematicians would not have the overview nor the experience of the special topics of statistics needed in particle physics. In fact, I do not know any institute, where (higher) statistics for physics data analysis is taught by mathematicians.

JUnQ: Should students from different fields have the same statistics training or should this be individually adjusted for different fields (e.g. physics vs. biology)?

Dr. Wanke: Since the requirements are very different between the different disciplines, there should also be different statistics courses, at least for the more advanced methods. Basic knowledge as means and variances, error propagation, least-squares method, etc., which is already relevant for beginners lab courses, of course can easily be taught together.

JUnQ: Do you teach statistics in your own discipline / in other disciplines? If yes, in which phase of the students' education?

Dr. Wanke: I am teaching statistics and data analysis for the Master of Science in Physics.

JUnQ: To your mind, is statistics in general taught in an appropriate way in your discipline so that, e.g., PhD students are capable of applying statistics to the interpretation of their data in a correct way?

Dr. Wanke: When I was studying (in the 80's), statistics courses for physicists were very rare or non-existent and I had to learn it myself during my PhD from the only existing textbook. Today it looks much better and I believe that any serious Physics Master of Science program involves a course in advanced statistics. For PhD students in particle physics, there even exist schools of one or two weeks, which teach advanced topics of statistics used in particle physics. Still, the situation for basic statistics as, e.g. needed in lab courses, does not always seem to be sufficient: very often basic statistical methods are just taught in an introductory lesson of the lab courses, only aiming on the direct application but not at any deeper understanding of the underlying principles.

JUnQ: What is the usual sample size you are dealing with?

Dr. Wanke: Very different: from 0 events to several billions.

JUnQ: Do you think that it's common practice to exclude "outliers"?

Dr. Wanke: No. However, in particle physics we are in a bit of a privileged situation: Usually we take millions and billions of data events under the very same conditions, therefore we usually do not have good arguments to exclude single "outliers".

JUnQ: Have you ever experienced misuse of statistics to make data appear better than it actually is?

Dr. Wanke: I do not remember any case in my personal environment in my whole career, that statistics were intentionally misused. However, unintentional misuse happens quite frequently in my opinion. The reason is practically always that selection criteria are chosen by looking at the data. This is in most cases (and all cases I know) not bad intention, but a mixture of laziness and unawareness. Performing a "blind analysis" usually requires a lot of effort and, in addition, the need of a blind analysis is very often not realized.

JUnQ: What do you think can be done on the educational level to improve good scientific practice regarding statistics?

Dr. Wanke: Statistics courses should probably be more applied and less theoretical. In particular, the problem mentioned in the previous answer needs to be addressed and emphasized.

JUnQ: Thank you very much for this interview!

—Kai Litzius