USEFUL “LAWS” OF RANDOMNESS: TEACHING EXPERIMENTS WITH EXCEL

Manfred Borovcnik
University of Klagenfurt, Austria
Manfred.Borovcnik@uniklu.ac.at

Only at first sight law and randomness are antithetic. The special feature of laws of randomness is thus difficult to understand. Some simulation experiments help for orientation. EXCEL will be used to introduce the idea of probabilistic laws in teaching with students at secondary school and with business students at university.

REGRESSION ANALYSIS AND THE MEANING OF CORRELATION

The correlation coefficient “measures” how well a line fits data points: With a slider in EXCEL for both the x- and y-values for the sixth data point you can dynamically demonstrate what this really means.

Variation in data - corresponding to their unpredictability - is measured by sums of squares. By simple descriptive statistics applied to the dependent variable y, to predicted values for y based on a linear model and to the residuals of these predictions from actual values you can illustrate various important relations such as: There is a split of variation in data y into the variation of predicted values and variation of the residuals. $1-R^2$ is the factor by which variation of residuals – the remaining “unpredictability” in data after using the linear model – is smaller than original variation. This Analysis of Variance split is understandable at that very basic level.

LAW OF LARGE NUMBERS AND THE MEANING OF REMAINING RISK

Occasionally the law of large numbers is over-interpreted as convergence of a specific series of relative frequencies towards the unknown probability. With a spreadsheet it is easy to discuss at more depth the character of that convergence. Matters are demonstrated better by repeating the whole series (how small it may be) than by continuing to observe the series in its convergence. With a fixed sample size of say 10, the proportions of an event – with fictionally known probability of 0.3 – to occur will lie between 0.0 and 0.7 related to 95% of repeated sampling. There is a risk to get a sample with a proportion of that event outside this interval of 5%. With 100 data, the sample proportion will lie between 0.20 and 0.40 – with the same remaining risk. This illustrates the fact that larger samples – if taken randomly – “measure” the unknown probability with more accuracy supposed the remaining risk is put at the same level. In EXCEL, this is easily implemented; the whole sampling process of repeated samples may easily be repeated, which will manifest a true meaning of remaining risk – as it fluctuates.

STATISTICAL INFERENCE BY RESAMPLING METHODS

It is surprising how easy concepts are in the setting of resampling. In the classical approach the value of one sample is judged by a theoretical distribution for the repeated procedure of taking samples. Some knowledge about probability distributions is necessary to understand the required steps. The resampling approach replaces the theoretical consideration what would occur if the sample is repeated by a practical experiment: What does occur if the sample is repeated. And it is repeated by simply re-sampling the already available sample. In the poster some of the procedures are implemented in EXCEL.

DISCUSSION

Students liked the active approach. They were surprised that sophisticated concepts may have a simple meaning. More information: http://www.uni-klu.ac.at/stochastik.schule.