

# USING MICROSOFT EXCEL VBA IN A BIOMETRIC CASE STUDY

György Hampel

Institute of Economics and Rural Development, Faculty of Engineering, University of Szeged, Szeged,  
Hungary

*hampel@mk.u-szeged.hu*

## Abstract

Spreadsheets can be used to calculate values from data and display the results in tables or charts. Microsoft Excel spreadsheet software supports the use of Visual Basic for Application (VBA) development environment which supports the creation of the users' own programs with user-defined functions. This feature is particularly useful when executing complex and repetitive calculations.

Sometimes, before calculating statistical data, we need to check whether our data (the random sample) is normally distributed and for numerical data type, one of the most effective methods is the Shapiro-Wilk test.

Excel can be used to execute the Shapiro-Wilk test in a user-friendly, programmable way with a little knowledge of VBA. The task requires complicated computing; thus, it is advisable to create a universal method which can be readily re-used with datasets with a different number of elements. Excel also supports the creation of a user-friendly, intuitive interface.

The formula of the Shapiro-Wilk test gives a value which is compared to a critical value (dependent on the sample size and the Type I error probability). Smaller Shapiro-Wilk values than the critical value indicate that the sample is not normally distributed.

In the research, the samples originated from the measurement of a type of crop with 44 chemical components per sample. These samples were available from three types of crop, for three consecutive years and according to two cultivation methods. Three parallel measurements were made in the same conditions (including crop type, year and cultivation mode). As a result,  $44 \times 3 \times 3 \times 2 \times 3 = 2.376$  numerical data had to be evaluated.

At first, a worksheet was programmed with the necessary calculations which were hidden from the user. Then the interface was created for the user in another worksheet to input the data, to set the Type I error probability and then display the result based on the calculations on the hidden worksheet: whether the input data is normally distributed or not.

Since the data input into the spreadsheet could be automated, several datasets could be evaluated one after another. As a result, samples gathered for a biometric study could be verified very quickly before executing further statistical calculations.

*Key words: Excel, VBA, statistical evaluation*