Per Vejrup-Hansen

Excel for STATISTICS

How to Organize Data

Per Vejrup-Hansen

Excel for Statistics
How to Organize Data

Samfunds
litteratur

Per Vejrup-Hansen
Excel for Statistics. How to Organize Data

Preface

This brief text is a handbook on statistical software, not on statistical theory and methodology as such. The focus is on how to create input for Excel's statistical functions and data analyses when using 'raw' data. Also, a presentation of basic Excel targeted at statistics is included. In conclusion, output by statistical functions and data analyses is briefly interpreted.

The need for organizing data applies to survey data in particular. Survey data are individual observations, not aggregates as is the case of accountancy data on for example sales. Also in contrast to experimental data, you cannot determine the structure of data beforehand when using survey data.

The necessity of organizing data stems from the fact that Excel is a spreadsheet program. Variables are arranged in columns and rows, and inputs into statistical tools are cell ranges. Another aspect is that Excel's statistical functions on tests often are based on ready-made figures. Input is for example the standard error which, however, is a measure to be derived from underlying data.

In this presentation, Excel is plain Excel without reference to supplementary add-ins. The idea is that you can proceed with standard Excel software which is readily available. However, the Data Analysis component (Analysis ToolPak) is missing in the Mac Excel version. But this essential part of standard Windows PC Excel can be included by a free add-in called 'Statplus:mac LE'. The add-in is provided by analystsoft.com, and can be downloaded at *analystsoft.com/en/products/statplusmacle*.

The first chapter gives a brief account of basic Excel targeted at statistical analyses. For example, the presentation of how to create a Pivot table is targeted at making a contingency table, and for making charts the Scatter Plot is selected. Also, when describing how to move columns, sort rows etc., the aim is to create input for statistical analyses. Due to this targeted approach, the presentation is delimited. Included are common procedures on how to format cells, use dialog boxes, copy formulas by Auto Fill, etc. Prior experience with Excel is not presupposed.

A methodological assumption also contributes to delimiting the text: the t-distribution is only applied for tests about means. Accordingly, it is assumed that surveys are samples, and that the population variance is unknown.