Electronic resources on the statistical functions of the program Excel

Aghamehdi Mehdialiyev

Academy of Public Administration under the President of the Republic of Azerbaijan

In this report the functions of MICROSOFT EXCEL have been considered with its indispensable role being highlighted. Statistical functions are grouped on the functional proximity and purpose. In each slide created by the author, there are links to the slides of other functions in the same group. In most examples a solution was found through two ways, i.e. by "manual" way and using EXCEL, which allows the user to understand better the essence of the considered function.

Keywords: Statistics, statistical analysis, statistical functions

Introduction

Nowadays the interest in statistical methods is growing due to their important role as a means of decision making. It is difficult to study and forecast socio-economic processes without the use of mathematical statistics and econometric models. It is known that laws of both centralized and market economy are based on the theory of probability and mathematical statistics. With the development of society, economic system becomes complicated meanwhile the statistical nature of laws describing the socio-economic processes enhances. This requires from a specialist in-depth knowledge on the theory of probability and statistics in addition to an ability to implement these skills with the help of computer programs. For the statistical analysis two types of instruments are used: universal (spreadsheets, database management systems, and others) and specialized (program SPSS, MathCad, MathLab, and others). Specialized packages are usually narrower and require in-depth knowledge of the field while spreadsheets are more flexible and simple. The most popular representative of the spreadsheet programs is a MICROSOFT EXCEL, which is a powerful and versatile enough spreadsheet. The strength of it is that it has many additional add-ins and a large number of internal analytical and computing functions. Usually in high schools and at EXCEL computer courses these features are not given due attention with providing an overview of software add-ins and built-in functions at maximum. As regards Computer Science course syllabus, required number of hours is not given for details of EXCEL features. Therefore, the average person has to ask help from the system (as experience shows users are reluctant even to do this). Sometimes even experienced users have a sense of frustration after reading the contents of reference. This is because that some of software upgrades and many built-in EXCEL (except for trivial operations of summation, multiplication, calculating trigonometric functions, etc.) require a mathematical background. Therefore, in order to correctly apply any add or built-in function user should refer to special literature and know physical nature of an appropriate method and apparatus.

Among built-in function categories of EXCEL "Statistical" category occupies a special place due to its numerous functions. It must be said that in the Help EXCEL there is summarized information on add-in "Data Analysis" and on each statistical functions with examples for practical use. This is believed to be

insufficient because in order to have a good understanding of the essence of each statistical function, user must feel the essence of the considered mathematical functions. For this reason brief examples are considered in the current paper and they are solved in two ways, namely by manual mode and by using appropriate statistical functions of EXCEL (Mehdialiyev and Aslanzade, 2008; Mehdialiyev, 2010).

While investigating statistical functions original slides for each statistical function have been created and the following information was reflected on them:

- The syntax;
- Result;
- Function arguments;
- Notes;
- Mathematical and statistical interpretation, and
- Example (link).

The first four elements of the slides are organized on the basis of background information on EXCEL. Section of Mathematical and statistical interpretation of the aggregate is made on the basis of a variety of sources of mathematical statistics. Mathematical formulas and ideas that are the basis for the definition of the function are provided as well. Each slide contains a link to interesting examples on this statistical function. An example is the discovery of the nature of statistical functions and is solved "by hand" and using the appropriate statistical function EXCEL.

Attempts are made to classify statistical functions of EXCEL by various criteria in this paper. One of the criteria used is the proximity on the destination. For example, functions **VAR, VARA, VARP, VARPA, AVEDEV, DEVSQ, STDEV, STDEVA, STDEVP, STDEVPA** grouped together, and within the group are connected to each other by hyperlinks. This grouping allows the user to move easily from one function to another and see similarities and differences between them. For example, note that the functions **VAR** and **VARA** unbiased estimate of the variance is calculated by the formula:

$$\hat{S}^2 = \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

or

$$\hat{S}^2 = \hat{\sigma}^2 = \frac{n \sum_{i=1}^{n} x_i{}^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)},$$

where n - sample size;

x_i - values of sample;

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$ - the average of the sample.

And function **VARP** and **VARPA** the offset variance estimate for the population by the formulas:

$$S^2 = \tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

or

$$S^2 = \tilde{\sigma}^2 = \frac{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2}{n^2}.$$

It should be noted that the allocation of statistical functions into groups was carried out in the work and has the additional load. Consider, for example, the grouping: **STDEV, STDEVA, STDEVP, STDEVPA.** Here **STDEV** is the root word, names of other functions differ from it only with one or two letters. Function **STDEV** assumes that its arguments are a sample of the population. In addition, the arguments of function **STDEV** can be either numbers or names containing numbers, arrays, or references, logical values and text representations of numbers that you type directly into the list of arguments. To include logical values and text representations of numbers in a reference as part of the calculation, the **STDEVA** should be used. The same is true for group functions **VAR, VARA, VARP, VARPA.**

Having two different formulas for estimating the variance explained by the need to eliminate biased sampling error (part of the error representation). It is known that as a permanent part of the representativeness error, offset error increases with increasing sample size.

Among of statistical functions EXCEL has a number mutual-inverse functions such as **NORMDIST (NORMINV), NORMSDIST (NORMSINV), BETADIST (BETAINV), LOGNORMDIST (LOGINV), FDIST (FINV), TDIST (TINV), CHIDIST (CHIINV), FISHER (FISHERINV).** The concept of reversibility of the two statistical functions is similar to the concept of inverse functions in mathematics. For example, between mutually inverse statistical functions; **NORMDIST (x, mean, standard_dev, cumulative)** and **NORMINV (probability, mean, standard_dev)** is performed the following relations:

x = NORMINV (NORMDIST (x, mean, standard_dev, cumulative), mean, standard_dev),

probability = NORMDIST (NORMINV probability, mean, standard_dev), mean, standard_dev, cumulative).

To analyze the relative position of characteristic values in EXCEL, there are the following statistical functions: **RANK, PERCENTRANK, PERCENTILE.**

- The **RANK** determines rank of given number in the list of numbers.
- Function **PERCENTRANK** determines percent rate of given value in the data set.
- **PERCENTILE** defines the value of a specified percentage rank.

We note some properties of statistical functions **RANK, PERCENTRANK,** and **PERCENTILE.**

1. RANK (number, ref, 0) + RANK (number, ref, 1) = n + 1, for non-recurring values in an array of data, where n indicates the number of elements in the array.

2. RANK (number, ref, 0) + RANK (number, ref, 1) = n, for duplicate values in the array data.

3. PERCENTRANK (array, ROUND (PERCENTILE (array, PERCENTRANK (array, x, 0)) = PERCENTRANK (array, x, 0)

4. RANK (ROUND (PERCENTILE (array, PERCENTRANK (array, x, 0)), array, 0) = RANK (x, ref, 0)

5. ROUND (PERCENTILE (array, PERCENTRANK (array, x, 0) = x

Here one would like to mention the democratic nature of program EXCEL. The function **RANK** assigns duplicate numbers the same rank. This is important when ranking of participant of sports events is evaluated, ranking performance of different content of socio-economic processes and phenomena. It is worth noting that if an interested result does not match with any of the values of the array, the function **PERCENTRANK** interpolate and determines the relative position of this result, which is very useful in mathematical and economic analysis of the real problems of life and society.

Moreover, there are such functions as **MAX** and **MIN** in EXCEL. For defining extreme data set, there are additional two functions:

- LARGE (array, k) determines the k-th largest value in a data set as measured in the descending order of the maximum value;
- SMALL (array, k) determines the k-th largest value in a data set as measured in the ascending order of the minimum value.

For these functions, we have:

- LARGE (array, n) = MIN (array)
- SMALL (array, n) = MAX (array)
- LARGE (array, k) = SMALL (array, n-k +1)

Some statistical functions have a logical argument, known as cumulative. The implications of this argument for all functions take the same values: **TRUE** and **FALSE.** However, these values perform certain functions in certain situations:

- If in the function **NORMDIST** (x, mean, standard_dev, cumulative) cumulative is **TRUE**, it returns a cumulative distribution function, if this argument is set to **FALSE**, the function returns the density function. So it is with the functions **GAMMADIST** (x, alpha, beta, cumulative), **EXPONDIST**(x, lambda, cumulative), **WEIBULL**(x, alpha, beta, cumulative).
- For function **BINOMDIST**(number_s, trials, probability_s, cumulative) this means: if the argument is "cumulative" is set to **TRUE**, the function returns the probability that the number of successful tests is not less than the argument "number_s"; if this argument is set to **FALSE**, it returns the probability that the number of successes equal to the number "number_s".

For function **POISSON**(x, mean, cumulative), this means that if the argument is "cumulative" is set to **TRUE**, then the function **POISSON** returns the cumulative. Poisson distribution is like the number of random events will be in the range from 0 to x, inclusive. If this argument is set to **FALSE**, it returns the density function **POISSON**, there is a possibility of an exact equality of the events to the value x.

Further studies will be conducted to collect new facts about other new mathematical features of EXCEL's statistical functions and their relationship with other functions from various categories in future (Mehdialiyev, 2011a; Mehdialiyev, 2011b; Mehdialiyev, 2012). As a logical continuation of research, a blog of "Statistical functions of EXCEL program" has been created (Mehdialiyev, 2012). The URL address is http://www.statistical-functions-aga.blogspot.com. This blog is an electronic service sample built on the basis of statistical functions MS EXCEL 2007 to teach "Statistics" subject according to Bologna credit system and requirements of State Educational standards. This electronic resource is intended for students, master's students, research scientists, economists, engineers, and specialists with various backgrounds, who process statistical information in their daily work. A blog is constantly updated, adding new content on the statistics and program EXCEL.

References

Mehdialiyev, A., Aslanzade, N., 2008. Add-in functions EXCEL [Vstroennye funkcii EXCEL], in Russian, Baku

Mehdialiyev, A., 2010. Statistical functions [Statistik funksiyalar], in Azerbaijan, Baku

Mehdialiyev, A., 2011a. "Some mathematical properties of statistical functions EXCEL," Electronic resources in continuing education ("Erno-2011"): Proceedings of the II International scientific-methodological symposium [Mezhdunarodnyj nauchno-metodicheskiy simpozium "Jelektronnye resursy v nepreryvnom obrazovanii" ("ERNO-2011"): Trudy II Mezhdunarodnogo nauchno-metodicheskogo simpoziuma], in Russian, 18-21 September, Anapa, Rostov-na-Donu, pp.85-87

Mehdialiyev, A., 2011b. "About of an experience of the teaching Statistics by using EXCEL," 5th International Conference on Application of information and communication technologies [5-ya Mezhdunarodnaya konferenciya po ispolzovaniyu informacionno-kommunikacionnyh tehnologiy], in Russian, 12-14 October, Baku, pp.288-93

Mehdialiyev, A., 2012. Statistical functions of Excel http://www.statistical-functions-aga.blogspot.com