Todor Kaloyanov, Assotiated Professor, Ph. D.

A POSSIBILITY FOR A GRAPHIC REPRESENTATION OF THE INTERRELATIONS AMONG THE PARAMETERS OF STATISTICAL DISTRIBUTIONS

In some of the cases when statistical methods of research on inter-relationships and interdependencies are applied to distributions different from the normal, the results can be disputed. One of the reasons is that the construction of these methods is based only on the first moment, the second central and mixed moments. The measures based on moments of higher order, such as the estimators for asymmetry and excess, are usually ignored despite the fact that they carry information which very often is extremely valuable. The article presents an idea for the simultaneous use of the moments from the first to the fourth order - the mean, standard deviation, and the asymmetry and excess coefficients for comparing several distributions to the same moment and for investigating the dynamics of distributions.

JEL: C10, C46, G11, J10, J13

Normal distribution plays a fundamental role in statistical methodology. This role is mostly determined by two factors:

1. Part of the empyrical statistical distributions either follow the normal distribution or asymptotically incline towards it. This makes possible the use of the Central Limit Theorem's effects and the application of the apparatus of the sample statistical research in its two modes - statistical estimation of parameters and statistical hypotheses testing.

2. Normal distribution is explicitely defined by means of two parameters only - the mean (μ) and standard deviation (σ). This characteristic of the normal distribution facilitates the application of the method of scientific abstraction - in this case by eliminating the impact of one of the parameters we can determine the impact of the other distribution parameter. As a rule, in statistical research the focus is mostly on the differences of the values of the mean. These differences are determined by eliminating the change /difference in the variance, in other words, the standard deviation is considered constant, invariable or the same. This assumption underlies the requirements which must be met for the correct application of the basic statistical methods – analysis of varince, regression analysis and corellation analysis. This requirement enables us to identify the presence or absence of any differences in the values of the mean. Apparently, if the distribution is not normal and if it is defined by means of three or more parameters, the task becomes more complicated since it demands the elimination not only of one parameter, as it is with the normal distribution, but of all other parameters.

Let us study a diagram showing the application of the scientific abstraction method when working with normal distribution.[1]

[1] The mean and the standard deviation can swap places, which does not result in any substantial changes. See Body, Z., A. Cain, A. G. Marcus. Investments. Sofia: Naturella, 2000, p. 130.

Figure 1

Values of the mean at a constant variance

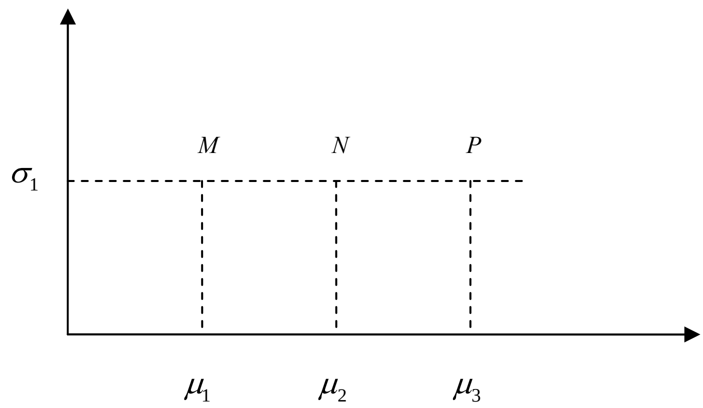


Figure 1 presents a case when the values of the mean vary, which results in a change of the coordinates of the respective points. In the case above the three points *M, N, and P* are situated on a line determined by the value of the standard deviation. The difference in their positions is determined only by the values of the mean.[2] Usually, when examining the factor influence and in the case of time series analysis it is assumed that the most important changes occur in the value of the mean. The presumption is that variation is constant and that the result's variable distribution is normal.Therefore, for the purpose of the research it is sufficient to study the differences in the mean.

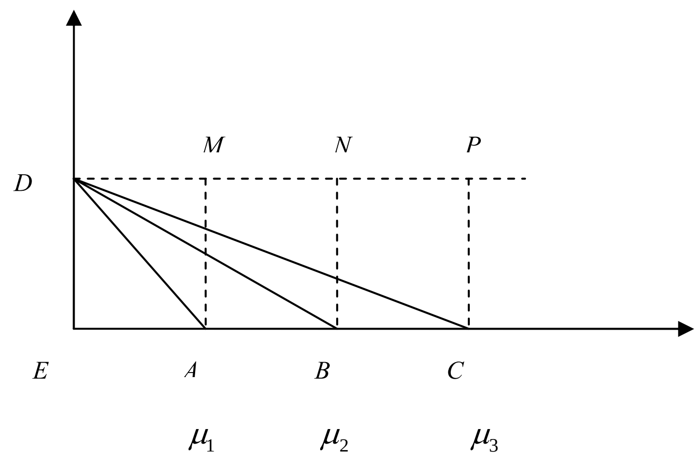
Let us go back to Figure 1 and ask the following question: is it possible to compare not only the co-ordinates of the respective points *M, N, and P,* but also the areas of the respective rectangles that are also functions of these two parameters? In this particular case the differences between the areas are functions of the mean only. We believe that these areas can be compared. Similarly, it is possible to compare the perimeters of the rectangles as well as the areas of the triangles formed by connecting the points corresponding to the respective values of the arithmetic mean on the x-axis and on the y-axis as shown on Figure 2. These are the triangles with a common point D, corresponding to the value of the standard deviation, and point E at the beginning of the coordinate system, and the third point corresponding to the value of the mean. In this particular case three triangles are shown - *ADE, BDE, and CDE.* The difference in the areas of these triangles is apparantly determined by the different values of the mean.

---

[2] As already mentioned, it is possible to eliminate the variations (differences) in the mean and to research the variations (differences) in the values of the standard deviation. This task, as a rule, is rarely set.

Figure 2

The triangles formed with the different values of the mean when the value of the standard deviation is constant



We can conclude that there are differences in the value of the mean from the perimeter of every triangle, as well as from the length of the hypotenuse or from the angle between the abscissa and the triangles' hypotenuses. As already mentioned, these options are available due to the existence of normal distribution. As stated by Campbell R. Harvey, John C. Liechty, Merrill W. Liechty and Peter Muller, it is well-known that the distributions of the returns from investments in stocks and bonds are, in many cases, not normal.[3] This means that these distributions have either left or right asymmetry or positive or negative excess, or both at the same time. Therefore, the distributions are defined by three or four parameters, i.e. the chart presented in diagrams one and two is not sufficient.[4]

We believe that the diagram can be extended and can be used to compare distributions different from the normal. The objective when performing these comparisons is defined on the basis of the tasks set: to determine a variance/ difference between the distributions of the units of two or more sets at the same moment, or at two or more subsequent moments, to estimate the influence of factors.

---

[3] *Campbell R. H., J. C. Liechty, M. W. Liechty and P. Muller*. Portfolio Selection with Higher Moments. SSRN, Electronic Paper Collection, December 13, 2004, p. 5.

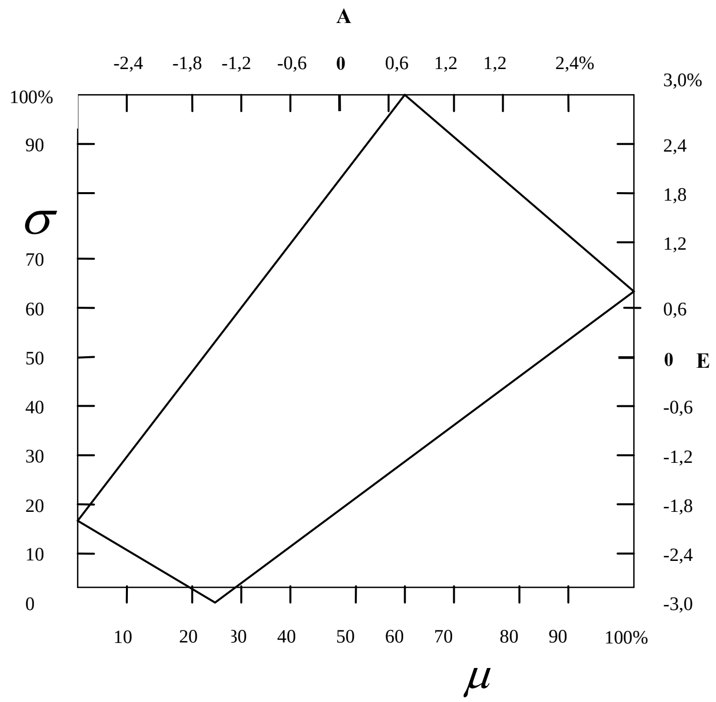
[4] *Feldstein, M. S.* Mean-Variance Analysis in the Theory of Liquidity Preference and Portfolio Selection. - Review of Economic Studies, 1969, 36; *Samuelson, P. A.* The Fundamental Approximation Theorem of Portfolio Analysis in Terms of Means, Variances, and Higher Moments. - Review of Economic Studies, 1970, 37; *Tsiang, S.C.* The Rationale of the Mean-Standard Deviation Analysis Skewness Preference, and the Demand for Money. - American Economic Review, June 1972, 62, p. 354-371; *Byme, P., S. Lee*. Real estate portfolio analysis under conditions of non-normality: The case of NCREIF. - Journal of Real estate Portfolio Management, 1997; *Sugarev, Z*. Demographic statistics. Sofia: "Nauka i Izkustvo", 1975.

At the same time, for the graphic representation of the parameters of a particular distribution based on the first/initial, second, third and fourth central moments we can use a square whose sides are the respective axes. We mark the values of the first/initial moment, or mean (μ), on the base of the square, which lies on the abscissa; on the left side, the y-axis, we mark the values of the standard deviation (σ); on the third side, which is parallel to the basis, we mark the values for the asymmetry coefficient A; and on the last side we mark the values of the excess coefficient E - see Figure 3. A zero value on the y-axis means the lack of difference/variance of the units of the characteristic observed. In the middle of the third side we mark/ plot zero value which corresponds to the symmetrical distribution, and the end values are - 3, and +3 - the range within which the values of the asymmetry coefficient for the particular distributions extend. The fourth side of the square is divided in a similar way: in the middle is the zero value which corresponds to the normal excess, and the end values are -3 and +3. It is known that these values do not have theoretically determined upper and lower limits.

After the respective values of the distribution parameters have been marked on the sides, the resulting points are connected by means of straight lines. Thus, a rectangle with a specific area and shape is formed. This diagram is represented to Figures 1 and 2.

In general, the value of the four parameters is preferred to be in percentages, as shown in Figure 3.

Figure 3



By means of the so constructed diagram we achieve the following:

1. Visual representation of the proximity(difference) between two or more empirical distributions simultaneously for the four parameters, constructed not only on first or second moments but also third and fourth moments - (μ),(σ), *A and E.* The representation of the normal distribution is a specific case when the coefficients of asymmetry and excess are equal to zero. Subsequently, the differences in the shapes will be due to differences in the values of the mean and the standard deviation. Various combinations are possible. The first of these is a combination between different means and the same standard deviation.

Then the shape will change/vary only along the x-axis and the two lines connecting the mean with the standard deviation and the excess respectively will vary their in lengths and movement along the x-axis. A second combination is between a constant mean and a varying standard deviation; this will result in the variation /change of the length of the lines connecting the x-axis to the axis of the coefficient of asymmetry. The third combination is when we have both the mean and the standard deviation varying at the same time. The commonest case is when the values of the four parameters - (μ), (σ), *A and E -* vary/ change.

2. By calculating the areas of the resulting figures and comparing them we can determine the size of the variance/differences between the respective distributions. Both direct comparison and relative comparison of the areas is possible. In order to achieve this, first we calculate the area of the square, and then we find the ratios of each of the areas to the area of the square. The resulting quotients will always be less than 1 and can be represented both as a coefficient or as a percentage, i.e. in a form easy to interpret.

3. On the graph it is possible to represent the respective standard distribution and then in relation to it to determine the differences/variations. The fact that all four moments are used makes it possible to work with distributions different from the normal distribution. The comparisons can be done for several distributions of different sets/series whose units are observed in relation to the same characteristic or for the distributions of a single set in relation to several discrete moments, in other words, for accomplishing a specific stochastic process for several consecutive moments.

4. The separate areas - triangles formed by moving the values of the different parameters along the axes - can also be the subject of analysis. As a result, we can look for an estimate of /try to estimate the contribution to the variation/ difference of each parameter. A similar estimate can be obtained by analysing the angles formed by each straight line and the respective axis. The perimeters of the different resulting figures can also be compared.

The interpretation of the diagram, the resulting figures and the differences among them depends on the character of the phenomenon being studied.

To illustrate this idea we will be using specific distributions constructed on the basis of data provided by demographic statistics for Bulgaria. These are the distributions according to age coefficients for women’s birth-giving in our country -

for all births and the birth of a first, second or third child for the period between 1985 and 2005.[5]

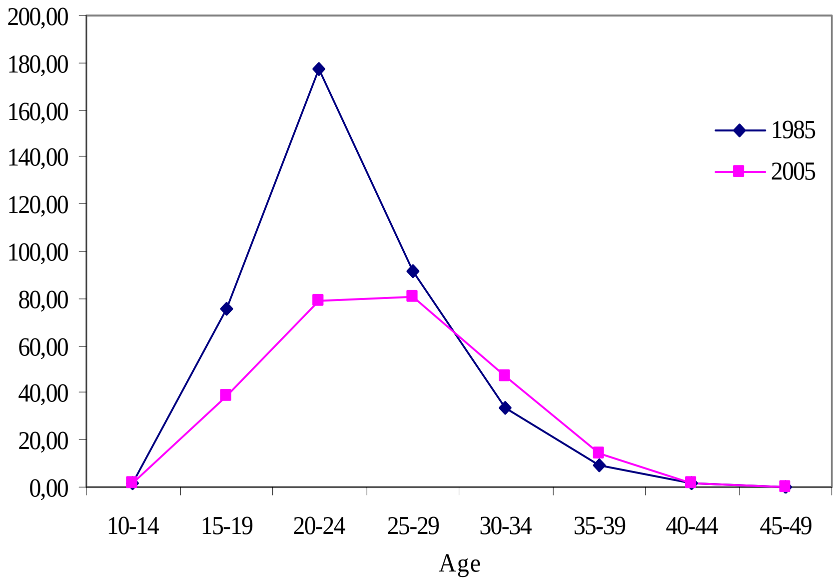
Table 1

Parameters of the distribution of age coefficients for birth-giving of women in Bulgaria - for all births given 1985 and 2005

Years	Parameters of the distribution of the age coefficients for all births			
	Average age μ	Standard deviation σ	Asymmetry coefficien *A*	Excess coefficient *E*
1985	23.52	4.96	0.82	0.77
2005	25.28	5.63	0.26	-0.28
Difference	1.76	0.67	-0.56	-1.05

Figure 4

Fertility age coefficients for the women in Bulgaria for all births in1985 and 2005



Calculated on the basis of data provided by the National Statistical Institute.

[5] The author is grateful to the specialists in the NSI department of Demographic statistics for the help offered to provide the information needed.

The value of the parameters in percentage points are shown in Table 2 and on the basis of these Figure 5 is constructed. In this particular case, the zero on the abscisa corresponds to 10 years of age and 100 on the same axis corresponds to 49 years . These are the bounds determining the child-bearing age of a woman.

Table 2

Parameters for the distribution of the age coefficients for female fertility for the women in Bulgaria – for all births for 1985 and 2005 (%)

Years	Parameters for the distribution of the age coefficients for female fertility for all births			
	Average age μ	Standard deviation σ	Asymmetry coefficien *A*	Excess coefficient *E*
1985	33.30	12.79	0.79	0.76
2005	38.49	14.79	0.27	-0.34
Difference	5.19	2.00	-0.59	-1.10

Figure 5

Diagram for the parameters for the distribution of the age coefficients for the fertility of women in Bulgaria – for all births in 1985 and 2005

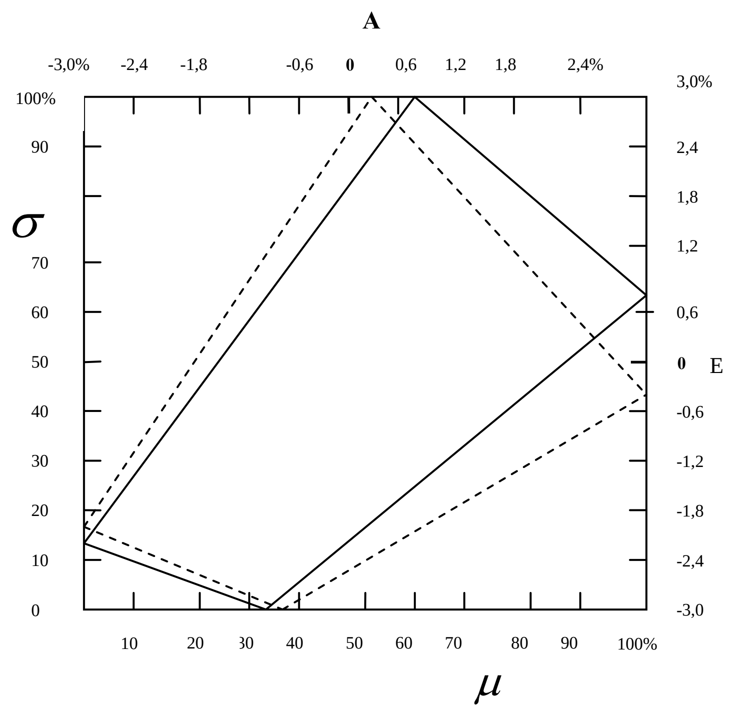


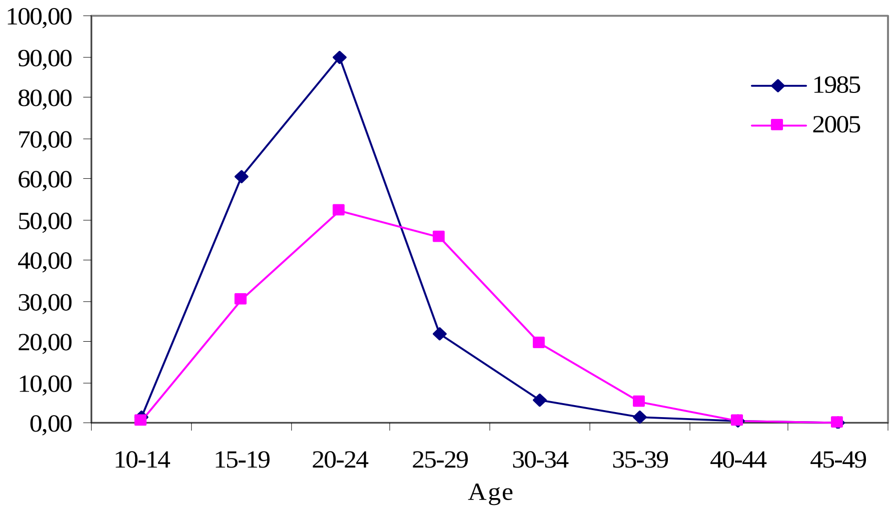
Table 3

Parameters of the distribution of the age coefficients for the fertility of Bulgarian women – for having a first child in 1985 and 2005

Years	Parameters for the distribution of the age coefficients for a first child birth			
	Average age μ	Standard deviation σ	Asymmetry coefficien *A*	Excess coefficient *E*
1985	21.35	4.19	1.08	2.27
2005	25.14	5.54	-0.04	-0.51
Difference	3.79	1.35	-1.12	-2.78

Figure 6

Age coefficients for female fertility for women in Bulgaria - for a first child birth in 1985 and 2005



Calculated on the basis of data provided by the national Statistical Institute

Table 4

Parameters for the distribution of age coefficients for female fertility for the women in Bulgaria for having a first child in 1985 and 2005 (%)

Years	Parameters for the distribution of age coefficients for first child birth			
	Average age μ	Standard deviation σ	Asymmetry coefficien *A*	Excess coefficient *E*
1985	27.85	10.74	1.09	2.26
2005	35.41	13.80	0.43	-0.19
Difference	7.56	3.06	-0.66	-2.45

Figure 7

Diagram for the parameters for the distribution of age coefficients for female fertility of the women in Bulgaria having a first child in 1985 and in 2005

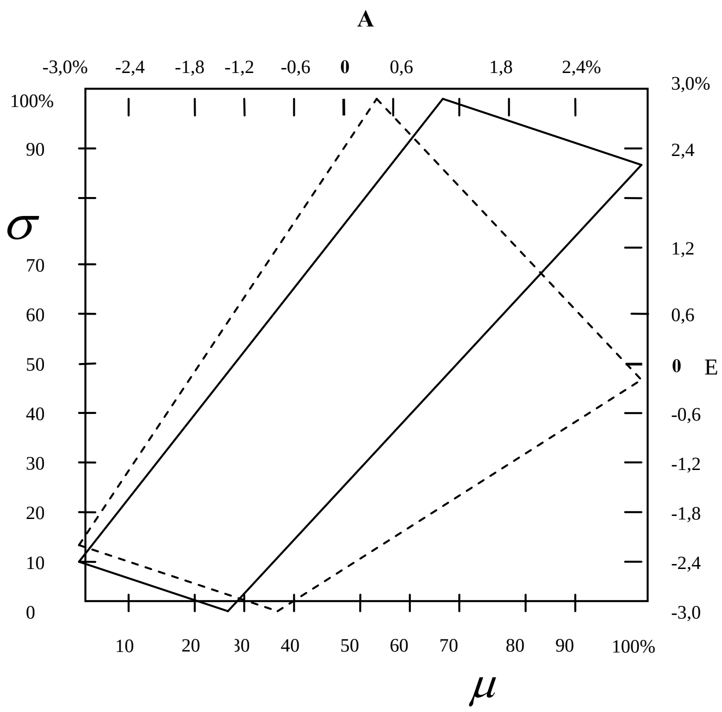


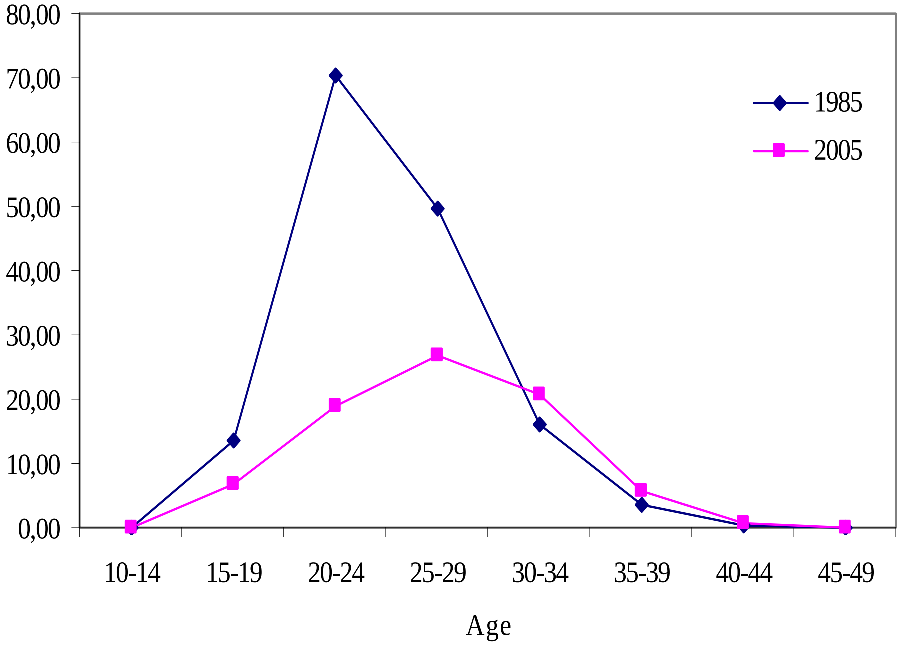
Table 5

Parameters for the distribution of age coefficients for female fertility of the women in Bulgaria having a second child in 1985 and in 2005

Years	Parameters of the distribution of the age coefficients for all births			
	Average age μ	Standard deviation σ	Asymmetry coefficien *A*	Excess coefficient *E*
1985	24.62	4.49	0.63	0.56
2005	27.05	5.49	0.32	-0.42
Difference	2.43	1.00	-0.31	-0.98

Figure 8

Age coefficients for the fertility of Bulgarian women having a second child in 1985 and 2005



Calculated on the basis of data provided by National Statistical Institute.

Table 6

Parameters for the distribution of age coefficients for the fertility of women in Bulgaria – for a second child birth in 1985 and 2005 (%)

Years	Parameters for the distribution of age coefficients for a second child birth			
	Average age μ	Standard deviation σ	Asymmetry coefficien *A*	Excess coefficient *E*
1985	36.25	11.52	0.63	0.54
2005	42.46	14.06	0.03	-0.42
Difference	6.21	2.54	-0.60	-0.96

Figure 9

Diagram for the parameters for the distribution of age coefficients for the fertility of women in Bulgaria – for a second child birth in 1985 and 2005

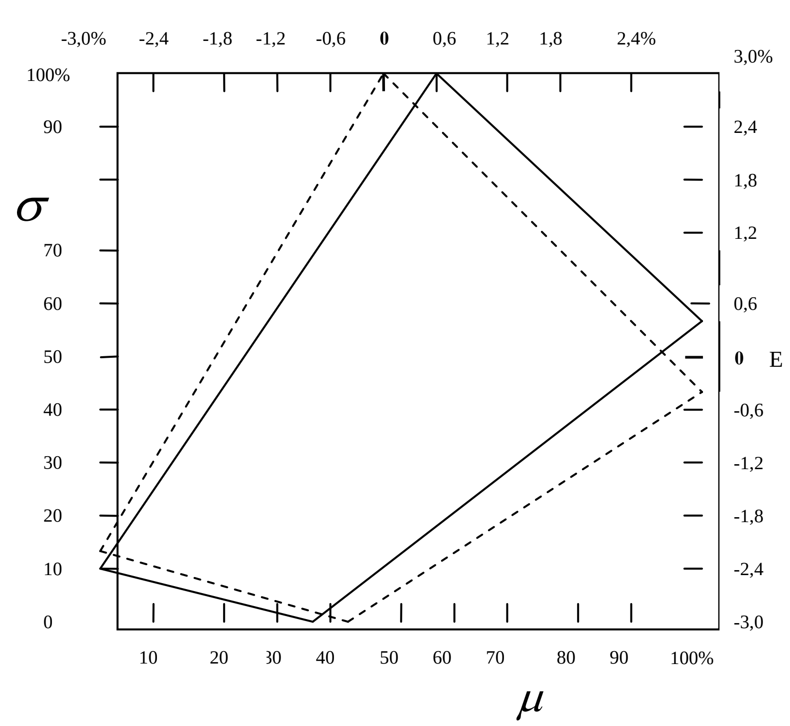


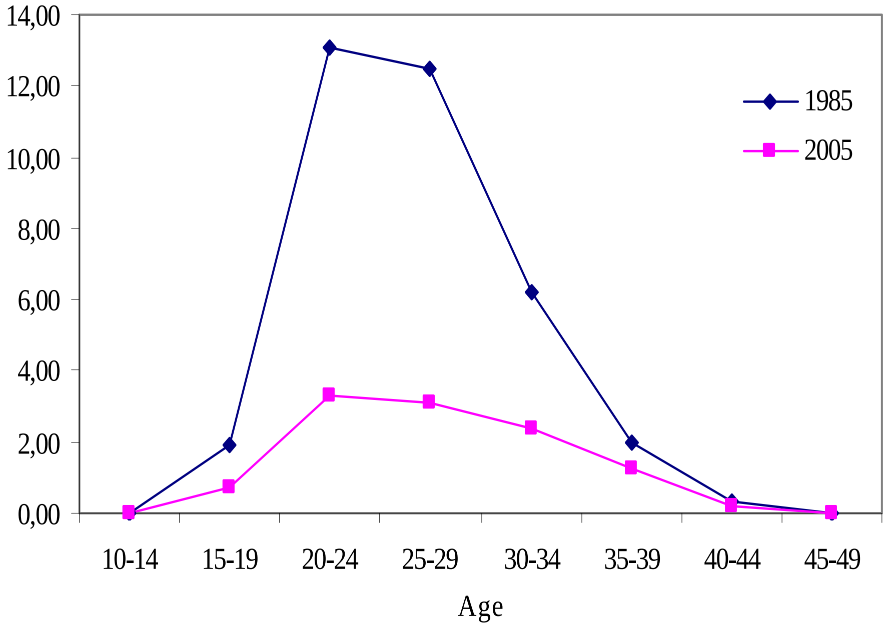
Table 7

Parameters for the distribution of age coefficients for the fertility of women in Bulgaria – for a third child birth in 1985 and 2005

Years	Parameters for the distribution of age coefficients for thr birth of a third child			
	Average age μ	Standard deviation σ	Asymmetry coefficien *A*	Excess coefficient *E*
1985	26.22	5.09	0.58	0.20
2005	27.35	5.78	0.26	-0.32
Difference	1.13	0.69	-0.32	-0.52

Figure 10

Age coefficients for the fertility of women in Bulgaria – for the birth of a third child in 1985 and 2005



Calculated on the basis of data provided by National Statistic Institute.

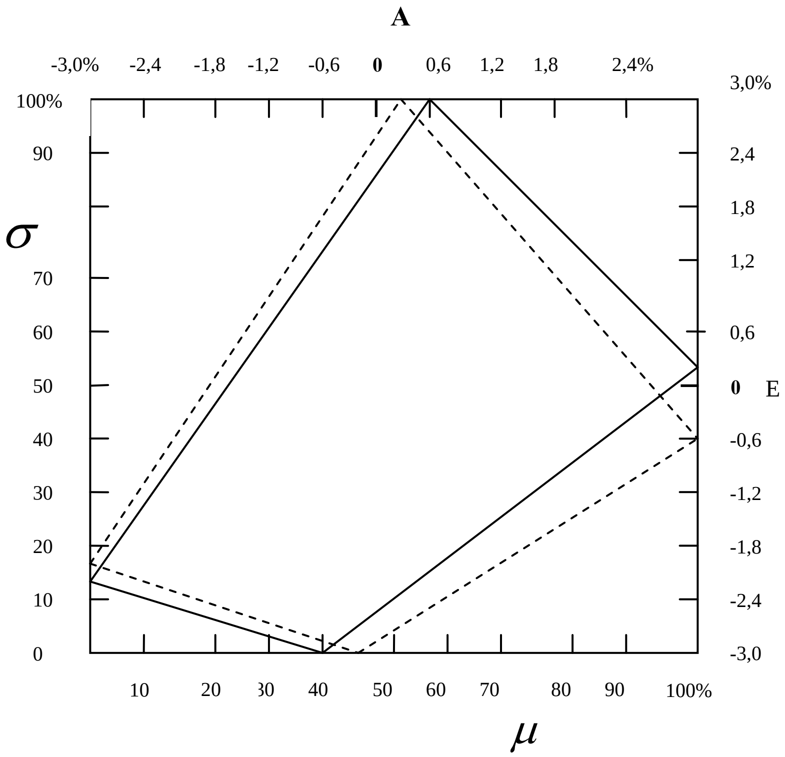
Table 8

Parameters for the distribution of age coefficients for the fertility of women in Bulgaria – for the birth of a third child in 1985 and 2005 (%)

Years	Parameters for the distribution of age coefficients for the birth of a third child			
	Average age μ	Standard deviation σ	Asymmetry coefficien A	Excess coefficient E
1985	40.33	13.04	0.58	0.20
2005	43.23	15.28	0.32	-0.56
Difference	2.29	2.24	-0.26	-0.76

Figure 11

Diagramme of the parameters of the distribution of age coefficients for female fertility for the women in Bulgaria giving birth of a third child for 1985 and 2005



Diagrammes 5, 7, 9 and 11 the figures drawn by solid line are for 1985 and the dotted lines form the figures for 2005. It can be seen that there are changes in all four parameters for all distributions, and as a result there are changes in the respective rectangles, which can be interpreted in the following ways: [6]

[6] For the interpretation and application of the asymmetry coefficient and the excess coefficient see *Borodachev, N. A*. Validations of the Methodology of Admissible and Mistakes Calculation of Size and Kinematic Chains. Moscow, 1946; *Borodachev, N. A*. Main Issues in the Precision of Production Theory. Moscow, 1950; *Venetskii, I. G., V. I. Venetskaya*. Genetal Mathematic-Statistical Concepts and Formulas in Econocmic Analysis. Moscow, 1979; *Kaloyanov, T*. Regarding the Cognitive Meaning of Assymetry and Excess Coefficients. - Alternatives, 1998, 10/11; *Стефанов, И., A. U. Totev*. Theory of Statistics. Sofia, 1960; *Tzonev, V*. Foundations of the Representative Study. Sofia, 1971; *Alexiev, J.* The Impact of Higher Moments on Hedge Fund Risk Exposure. - The Journal of Alternative Investments, Spring 2005.

• The rectangle's side connecting the average age with the standard deviation extends due to the growth of the average age and the diffusion registered both in the total number of births and in the birth of a first or second or third child. This means a delay of giving birth and probably giving birth later in life. The rise in the average age at which a woman gives birth is accompanied with an increase in the differences in the behavior of women from different age groups which is registered by the standard deviation. These changes are different in value but the tendency is the same – towards growth. The average age marks the highest growth with the birth of a first child – by 7.56 points. The slightest change is registered for the age of giving birth to a third child – only by 2.29 points. The changes in the dimensions of the standard deviations are similar. This could be explained by the fact that the changes in the moral values affect most significantly the young, and they are the ones most likely to give birth to a first child. The importance of the child in the family has also changed. However, a delayed birth of a first, and sometimes a second, child, does not lead to a substantial delay in the birth of a third child, if the family has decided to have one. The growth of the diffusion of the distribution of the age coefficients is evidence for the greater differences in women's reproductive behavior. This probably means that the impact of factors of diverse type is increasing and also new factors are likely to appear – factors which form the respective behavior in women of different education, income, etc.

• The line connecting the standard deviation and the asymmetry parameter gets shorter due to the changes in the measurements of both parameters. In this particular case it should be noticed that along with the growth of the variance, the distributions tend towards the symmetry and they have all been with positive asymmetry. With regard to asymmetry too, the biggest changes registered are these for the birth of a first child – from 1.09 to 0.43%, which is a difference of minus 0.66 points. The change in the measurements of the asymmetry means that the likelihood for having high fertility at a low average age diminishes a lot. This fact can be interpreted in another way: the benefit of giving birth to a child earlier in life is replaced by other benefits such as completing one's education, starting work, building a career, achieving financial independence, etc. In this particular case, the asymmetry reflects the changes in the relationship between benefits and losses resulting from changes in women's set of moral value.

• This extension of the line connecting the asymmetry with the excess is similarly due to changes of both parameters. Of great importance here is the fact that the transition from positive to negative excess is accompanied with growth of the standard deviation. For the analysis, this most probably means that there is an ongoing process of forming groups of women with different characteristic traits and more diverse reproductive behavior. There is no dominant factor to compel women to have similar behavior, in other words – to follow the same pattern of reproductive behavior. Nowadays they can choose from more options, or, at least, there are factors whose influence has increased, as a result of which various behavior patters arise.

The changes in the measurements of the four parameters are reflected in the areas of the corresponding rectangles.

Table 9

Areas of the rectangles illustrating the relationship among the parameters of the distributions of the age coefficients of the women in Bulgaria – for all births and for the respective order of the birth in the sequence for 1985 and 2005 (%)

Year	Total of births	Firstborn child	Second child	Third child
1985	42.08	34.07	43.20	48.90
2005	48.48	46.72	49.50	48.50
Difference	6.40	12.66	6.30	-0.40

The differences in the areas of the rectangles for 1985 and 2005 reflect the *aggregate changes* in the measurements of the four parameters of the distributions under research. It has been established that among the resulting changes the biggest is the change in the distribution for the firstborn child. This change is 12.66% - from 34.07% to 46.72% (the total area of the rectangle is 100%), with the area and the shape of the respective figure approaching these for a second child in 2005. This again confirms the conclusion that the major changes are these concerning the firstborn child, and the changes for a second child are only half as big. There are almost no changes in the area corresponding to the distribution of a third-born child.

On the basis of the diagram showing the relationship among the four parameters we can conclude that the changes in the reproductive behavior of the women for the period 1985 – 2005 involve mostly the firstborn child and to a lesser extent the second child.

The fact, for the analysis of the demographic events we have used four parameters for the distribution of the units rather than only one or two has made it possible to go deeper into the ongoing processes and to reveal the specific influences. With the help of these we have established that the increase in the age when women have a first and a second child has led to decrease in the asymmetry. The growing diffusion and the transition from positive to negative excess – a sign for over-dispersion/diffusion of the units – indicate that there is an ongoing process of the formation of a variety of behavior patterns among the women in Bulgaria. These patterns are most probably related to the appearance of groups of women with similar socio-economic characteristics. Over the last several years a certain part of the women within the fertile age have been receiving a high income, have taken important management positions and have committed themselves to climbing up the professional ladder, which has affected their plans for having children. The family and children have a relatively smaller proportion and fall back in the hierarchy within the women's set of moral values. The formation of groups of women with similar characteristics leads to different patterns of reproductive behavior, and this determines the growth in the diffusion/dispersal and the negative excess. In order to single out these patterns it is necessary to use groups formed on the basis of women's education, income, profession and other socio-economic features.

*

We believe that the advantages of the diagram presented here can be noticed when there are distributions different from the normal distribution. In these cases the areas will not be a function of only two parameters (which most often are the mean and the standard deviation) but will also be a function of the coefficients of asymmetry and excess. Examples of similar situations are presented in figures 5 – 8, where the measurements of all four parameters vary. With this diagram one can determine the individual contribution of each parameter for the respective change (difference). Similarly to the pure regression coefficients in the multiple regression analysis, here, by retaining three of the parameters one can estimate the contribution of each of the four parameters of the statistical distribution. The first result will show an estimate for the effect of the aggregate changes in the four parameters and the second result will indicate the effect from the changes in the individual parameters. Furthermore, it is possible to simulate a variety of situations with different combinations of the measurements of the parameters and to evaluate the respective results. The estimates of the variation/difference can be carried out in relation to a specific standard distribution which has been determined theoretically or empirically. It is possible to establish differences between two or more empirical distributions. Virtually, there are no restrictions for the comparative analyses that can be performed.

We would like to emphasize that this paper presents only the basic ideas for constructing and analyzing a diagram which reveals the relationship among four of the parameters of the statistical distribution.

20.XI.2007