Dr. I. TEICH
Ministry of Mines, Oil and Geology

A Statistical-Mathematical Interpretation of Man-Power Fluctuation Rate in Mining Industry

"The point is precisely to find such a model that its theory may success fully be applied to the actual situation"

Warren Weaver:

"Lady Luck: The Theory of Probability"

Man-power fluctuation in mining industry is characterized by double personnel shifting: inside shifting, brought on by workers migrating among mining units, and outside shifting brought on by workers migrating from this industrial configuration towards other branches of national economy.

This shifting process with all its economic and sociologic implications was recently discussed in a scientific symposium *).

Alongside of the analysis carried out by this symposium, the author of these lines submits a statistical-mathematical interpretation of the fluctuation rate of man-power migrating outside and its characteristic model for a definite time interval. Particular stress is laid on this class of workers directly affecting the qualitative structure of the investigated effectives.

To this end, we shall consider a representative example, on the analogy of the fluctuation rate in one of the main coal fields in our country. It is assumed that the monthly rate of the phenomenon that called our attention has got a configuration like în Fig. 1.

*) The scientific symposium "Man-power stability problems" was held at the Petroşani Mining Institute on October 6—7, 1972 under the auspices of the Sociological Department of the Academy of Social and Political Sciences.

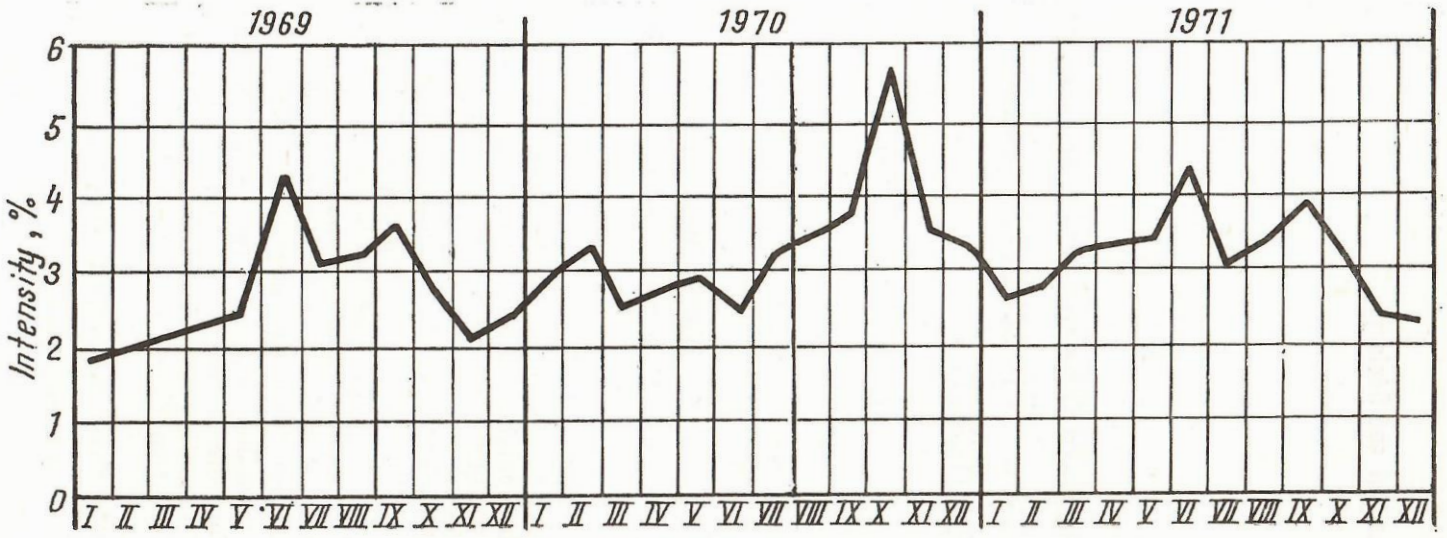
1969
1970
1971
Intensity, %
0
1
2
3
4
5
6
I II III IV V VI VII VIII IX X XI XII

Fig. 1

The general configuration of this curve is extremely complex and, for this reason, we cannot directly compare it to any of the curves known in mathematical statistics in view of identifying the model after which the phenomenon under survey is taking place. Therefore, we must process the data that have underlain the drawing up of the diagram in Fig. 1, so as to make obvious the existence of possible regularity, stability or evolution conditions in agreement with certain laws of fluctuation rate in the coal field under consideration. Such conditions shall favour a realistic interpretation of the phenomenon we deal with, in keeping with the present requirements of applying mathematics in economy.

In the foreseen processing, we shall in the first place use the statistical grouping method and also the least squares method. Some other procedures of mathematical investigation shall further be used.

In order to secure uniformity we shall adopt the following elements:

f_u — leaves occurrence (the ratio between the number of workers who have left and the total amount of this class of manpower in the entirety of the observation intervals at stage u);

a — lower limit of grouping intervals;

n — number of grouping intervals;

h — length of grouping intervals;

$a + nh$ — upper limit of grouping intervals;

$a + \left(i + \frac{1}{2}\right)h$ — central value of interval of rank i (yielded by observations comprised between $a + ih$ and $a + (i + 1)h$;

f_{iu} — a number of observation data comprised in the interval $a + ih$; $a + (i + 1)h$ at stage u (these data were computed in accordance with monthly occurrences in Table I)

$f_i = \sum_{u=1}^{k} f_{iu}$; k is the number of stages on which observation was carried out;

$p_i = \frac{f_i}{\sum_{1}^{n} f_j}$ fluctuation coefficient per total number of observation stages.

1. *The statistical grouping method.* The data underlying the drawing up of diagram in Fig. 1 are supplied in Table I.

Table I (%)

year	MONTH											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1969	1.88	1.95	2.09	2.27	2.49	4.34	3.07	3.19	3.57	2.92	2.19	2.30
1970	2.86	3.33	2.54	2.78	2.82	2.44	3.21	3.45	3.74	5.72	3.56	3.32
1971	2.66	2.75	3.14	3.34	3.36	4.23	3.03	3.39	3.87	3.10	2.32	2.26

In the group we mean to achieve, the intervals which have been chosen so as to emphasize representative data.

Thus, it is obvious that the fluctuation rate varied throughout the referred interval from 1.88 per cent to 5.72 per cent.

Depending on such percentage values, we assume in Table II: $a = 1.70$ and $a + (n-1)h = 4.40$. It should be noted that the values exceeding this latter value have conventionally been grouped in interval $a + (n-1)h$; $a + nh$, because in so doing the gross situation of data expressing the phenomenon under consideration is very little distorted.

Table II

Grouping intervals	$a + \left(i + \frac{1}{2}\right) h$	f_{iu}			f_i
		1969	1970	1971	
1.70—2,00	1.85	2	—	—	2
2.01—2.30	2.15	4	—	1	5
2.31—2.60	2.45	1	2	1	4
2.61—2.90	2.75	—	3	2	5
2.91—3.20	3.05	3	—	3	6
3.21—3.50	3.45	—	4	3	7
3.51—3.80	3.75	1	2	—	3
3.81—4.10	4.05	—	—	1	1
4.11—4.40	4.35	1	—	1	2
over 4.40	—	—	1	—	1
TOTAL	—	12	12	12	$\sum_1^n f_j = 36$

With reference to column f_i, we shall remark that most occurrences are comprised between 2.01 and 3.50 (representing 75 per cent of the total amount of occurrences). In the same interval we also meet the values corresponding to monthly means (1969 = 2 . 71; 1970 = 3 · 31; 1971 = 3.12). Likewise, in the interval 2.31 — 3.50 a certain regularity is noticeable, and so we shall take it into account in our interpretation.

It shall also be remarked that the representative values of f_i in the interval 3.51 — over 4.40 are far smaller than in the interval 1.70 — 3.50, leading to the impression of some random situations, that is to say non-significant for the phenomenon under survey. However, on plotting about 20 per cent of the total amount of occurrences, their impact on the final results of research shall not be disregarded. For this reason, in Table III we carry out a regrouping of the occurrences in the last intervals, so that the respective data in Table II become:

We see that no essential changes have taken place, since the only alteration is related to the partial sum in the last interval. The data regrouping, as well as the regularities remarked in the interval 2.31—3.50, suggest a fluctuation rate mathematical model in arithmetical progression.

We are to corroborate if such a model mirrors indeed the actual situation as reflected by f_i values. To this end, on knowing primarily the trend of the phenomenon under consideration, as shown by the statistical grouping

Table III

Grouping intervals	$a + \left(i + \frac{1}{2}\right) h$	f_{iu}			f_i regrouped
		1969	1970	1971	
1.70—2.00	1.85	2	—	—	2
2.01—2.30	2.15	4	—	1	5
2.31—2.60	2.45	1	2	1	4
2.61—2.90	2.75	—	3	2	5
2.91—3.20	3.05	3	—	3	6
3.21—3.50	3.45	—	4	3	7
over 3,50	3.75	2	3	2	7
TOTAL	—	12	12	12	$\sum_1^n f_j = 36$

method, we assume that the terms of series $f_i (i = 1, 2, \ldots, 7)$ can be approximated by the terms of a theoretical series in arithmetical progression $f'_i (i = 1, 2, \ldots, 7)$.

We have then to find such a series so that its terms shall, as a whole, be closest to terms f_i. At the same time, we assume as an optimum condition for adopting a series of this type that the sum of square deviations between the actual fluctuation coefficients p_i and the fluctuation coefficients resulting from the theoretical series shall be a minimum.

There are several possibilities for drawing out a theoretical approximation series in arithmetical progression. However, we shall insist in the series that can be inferred through the least squares method and we shall further include also other series in arithmetical progression in the calculus in order to complete the analysis.

2. *The least squares method.* The calculus of the theoretical series in this way consists in finding the growth rate and, indirectly, its original value *).

Consequently, consider:

$$f'_1,\ f'_1 + r;\ f'_1 + 2r;\ \ldots\ ;f'_1 + 6r$$

Since the sum of this arithmetical progression is 36 and the number of its terms is $n = 7$, we can write:

$$f'_1 + 3r = \frac{36}{7}\ ;\quad f'_1 = \frac{36}{7} - 3r$$

*) The method we make use of is a variant of the one applied by academician G. Mihoc and prof. dr. G. Theiler in view of finding the development rate through the parabolic mean. (See "Revista de Statistică", No. 1, 1963).

In order to find the value of r we make use of the least squares method through which we effect the sum of the square deviations between f_i and f_i' terms:

$$\left[2-\left(\frac{36}{7}+3r\right)\right]^2+\left[5-\left(\frac{36}{7}+2r\right)\right]^2+\left[4-\frac{36}{7}+r\right]^2+$$

$$+\left[5-\frac{36}{7}\right]^2+\left[6-\left(\frac{36}{7}-3r\right)\right]^2+\left[7-\left(\frac{36}{7}-2r\right)\right]^2+$$

$$+\left[7-\left(\frac{36}{7}-r\right)\right]^2=\left[3r-\frac{22}{7}\right]^2+\left[2r-\frac{1}{7}\right]^2+\left[r-\frac{8}{7}\right]^2+$$

$$+\left[-\frac{1}{7}\right]^2+\left[\frac{6}{7}-r\right]^2+\left[\frac{13}{7}-2r\right]^2$$

After effecting the due calculus we get the following equation:

$$3r-\frac{22}{7}+2r-\frac{1}{7}+r-\frac{8}{7}-\frac{6}{7}+2r-\frac{13}{7}=0$$

Hence:

$$9r-\frac{50}{7}=0;\quad r=\frac{50}{63}=0.79365,$$

so that the original value of the theoretical series will be:

$$f_1'=\frac{36}{7}-3\times 0.79365=5\cdot 14285$$

Table IV shows the calculus of the terms of the theoretical series f_1' in agreement with the above findings:

Table IV

Calculus elements for variant f_i'				
f_i'	t'	r	$r\cdot t'$	$f_i'=f_1'\pm r\cdot t'$
5.14285	—3	0.79365	—2.38085	2.76
5.14285	—2	0.79365	—1.58730	3.56
5.14285	—1	0.79365	—0.79365	4.35
5.14285	0	0.79365	0	5.14
5.14285	+1	0.79365	+0.79365	5.94
5.14285	+2	0.79365	+1.58730	6.73
5.14285	+3	0.79365	+2.38085	7.52
Total	—	—	—	$\sum_1^n f_j'=36$

Similarly, we also consider other two theoretical series in arithmetical progression with a conventional ratio, such as we have had in view:

f_i''	2	3	4	5	6	7	8
f_i'''	$2\frac{1}{7}$	$3\frac{1}{7}$	$4\frac{1}{7}$	$5\frac{1}{7}$	$6\frac{1}{7}$	$7\frac{1}{7}$	$8\frac{1}{7}$

where:

f_i'' is the arithmetical progression with $\sum_1^7 f_i'' = 35 \neq \sum_1^7 f_i = 36$ and

f_i''' is the arithmetical progression with $\sum_1^7 f_i''' = 36 = \sum_1^7 f_i = 36$

Now we have sufficient data to verify the optimum condition. We do this in Table V by calculating for the regrouped series and for the theoretical series the coefficients of the form p_i, as well as the adequate square deviations.

Table V

p_i	Variants of approximation			Square deviations		
	p_i'	p_i''	p_i'''	$(p-p')^2$	$(p-p'')^2$	$(p-p''')^2$
$\frac{2}{36}$	$\frac{2.76}{36}$	$\frac{2}{35}$	$\frac{2\frac{1}{7}}{36}$	0.000441	0.000026	0.000016
$\frac{5}{36}$	$\frac{3.56}{36}$	$\frac{3}{35}$	$\frac{3\frac{1}{7}}{36}$	0.000961	0.002809	0.002601
$\frac{4}{36}$	$\frac{4.35}{36}$	$\frac{4}{35}$	$\frac{4\frac{1}{7}}{36}$	0.000081	0.000009	0.000016
$\frac{5}{36}$	$\frac{5.14}{36}$	$\frac{5}{35}$	$\frac{5\frac{1}{7}}{36}$	0.000016	0.000015	0.000016
$\frac{6}{36}$	$\frac{5.94}{36}$	$\frac{6}{35}$	$\frac{6\frac{1}{7}}{36}$	0.000004	0.000023	0.000016
$\frac{7}{36}$	$\frac{6.73}{36}$	$\frac{7}{35}$	$\frac{7\frac{1}{7}}{36}$	0.000049	0.000030	0.000016
$\frac{7}{36}$	$\frac{7.52}{36}$	$\frac{8}{35}$	$\frac{8\frac{1}{7}}{36}$	0.000196	0.000224	0.001024
TOTAL —		—	—	0.001748	0.003136	0.003705

It is obvious that the variant calculated through the least squares method can be assumed to be the best. The sum of deviations is about 50 per cent smaller than in the other two cases. This result supplies an additional reason for assuming that the evolution of fluctuation rate in the referred coal field can be plotted by a mathematical model in arithmetical progression. Such a model makes obvious a relatively marked man-power shifting process that usually bears consequences on certain technical-economic indicators in the units. The problem involves a prompt and adequate decision concerning the replacement of the leaving workers so as to go on fulfilling the current production assignments. This means a continuous and many-sided improvement of stable personnel, capable of carrying on complex tasks, the discovery of specific capabilities in maintaining labour productivity at the planned level, a rapid integration of recently employed workers, etc.

Under such terms — that are already being provided for on large scale — it will be possible to favourably influence not only the plan targets but also man-power stability itself. Thus, it will be possibile to achieve a limited shifting process, particularly with unskilled workers, within tolerable limits only.

3. *The development of investigation.* The regularization of occurrences and the ascertainment of an optimum like in the previous example enable the achievement of a parallelism with the distribution functions in the probability calculus. In this respect we shall consider frequencies f_i successively cumulated, that is: f_1; $f_1 + f_2$; $f_1 + f_2 + f_3$;. . ; $f_1 + f_2 + f_3 + f_4 + f_5 + f_6$ If, also here, insted of exact data, we consider the data calculated through the least squares method and successively cumulate them, we shall achieve a 2nd order arithmetic progression.

In such a case the distribution function will assume the form of a 2nd degree trinomial, namely: $Ax^2 + Bx + C$.

Assuming

$$f(x) = Ax^2 + Bx + C \tag{3.1}$$

we can write:

$$f(a + 2h) - 2f(a + h) + fa = 2Ah.$$

Similarly:

$$f(a + 3h) - 2f(a + 2h) + f(a + h) = 2Ah$$

and so on, bearing in mind that $2\ Ah = r$, which is the ratio of 2nd order arithmetical progression.

Because A, B, C are parameters of a distribution function and $f(x) = f\left(a + \frac{h}{2}\right)$, we shall further write equation in (3.1) under the form:

$$f\left(a + \frac{h}{2}\right) = A\left(a + \frac{h}{2}\right)^2 + B\left(a + \frac{h}{2}\right) + c \tag{3.2}$$

The value of parameters A, B, C can be found either through the least squares method or through the probability laws.

3.1. *The first solving possibility.* For this purpose we shall proceed as follows:

$$\left[f\left(a+\frac{h}{2}\right)-p_1\right]^2 \Rightarrow \left[\left[A\left(a+\frac{h}{2}\right)^2+B\left(a+\frac{h}{2}\right)+C\right]-p_1\right]^2$$

$$\left[f\left(a+\frac{3h}{2}\right)-p_2\right]^2 \Rightarrow \left[\left[A\left(a+\frac{3h}{2}\right)^2+B\left(a+\frac{3h}{2}\right)+C\right]-p_2\right]^2$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\left[f\left(a+\frac{2n-1}{2}-p_n\right]^2 \Rightarrow \left[\left[A\left(a+\frac{2n-1}{2}\right)^2+B\left(a+\frac{2n-1}{2}\right)+C\right]-p_2\right]^2$$

and, further, on applying the principle of the least squares method:

$$\left[\left[A\left(a+\frac{h}{2}\right)^2+B\left(a+\frac{h}{2}\right)+C\right]-p_1\right]^2+\left[\left[A\left(a+\frac{3h}{2}\right)^2+B\left(a+\right.\right.$$

$$\left.\left.+\frac{3h}{2}\right)+C\right]-p_2\right]^2+\cdots+\left[\left[A\left(a+\frac{2n-1}{2}\right)^2+B\left(a+\frac{2n-1}{2}\right)+\right.\right.$$

$$\left.\left.+C\right]-p_n\right]^2=\textit{minimum} \qquad (3.1.1.)$$

which, for the example calculated in Table IV, will be:

$$\left[[A(1.85)^2+B(1.85)+C]-\frac{2.76}{36}\right]^2+\left[[A(2\cdot 15)^2+B(2.15)+C]-\frac{3.66}{36}\right]^2+$$

$$+\left[[A(2.45)^2+B(2.45)+C]-\frac{4.35}{36}\right]^2+\left[[A(2.75)^2+B(2.75)+C]-\frac{5.14}{36}\right]^2+$$

$$+\left[[A(3.05)^2+B(3.05)+C]-\frac{5.94}{36}\right]^2+\left[[A(3.45)^2+B(3.45)+C]-\frac{5.75}{36}\right]^2+$$

$$+\left[[A(3.75)^2+B(3.75)+C]-\frac{7.52}{36}\right]^2=\textit{minimum.}$$

The looked for minimum shall be ascertained also with regard to other variants (for instance, of the form considered in Table V) which are to be compared to the regrouped variant or to the exact variant if the phenomenon shows certain stability in the mathematical meaning of the word.

3.2. *The second solving possibility.* Another way for finding the values of parameters A, B, C consists in considering a distribution function of the type:

$$F(x)=\begin{cases}0 & \text{for } x \leqslant a\\ Ax^2+Bx+C & \text{for } a<x<a+nh\\ 1 & \text{for } x \geqslant a+nh\end{cases}$$

This requires a deeper examination of the stability conditions or the fluctuation rate regularity, as well as a larger volume of data covering a longer time period.

By way of information we point out that in order to find parameters A, B, C we must assume here $F(a) = 0$; $F(a + nh) = 1$, and $2Ah = r$, where r is the 2nd order constant difference.

The investigation of fluctuation rate in this manner may reveal some other aspects of the problem we have tackled. However, since this requires a number of details and sets certain conditions to be met in actual practice, we shall dwell on this problem in another paper. So much the more that the present paper is in fact a first attempt to investigate man-power fluctuation rate in mining industry by means of methods employed in mathematical statistics.

REFERENCES

1. YULE G. V., KENDALL M.G. *Introducere în teoria statisticii* (Introduction to Statistics Theory). Translated from the English language, Editura ştiinţifică, Bucharest, 1969.
2. BOEV G. P. *Teoria veroiantnostei Gosudarstvenoe Izddatelstvo*, Moscow, 1950.
3. KRAMER H: *Mathematical Methods of Statistics*. Princeton Universited Press, 1946.
4. THEILER G: *Cerinţele aplicării matematice în economie*. (The Requirements of Mathematics Appling in Economy). Studii şi cercetări de calcul economic şi cibernetică economică, no. 4. 1972.

INTERPRETATION STATISTIQUE MATHEMATIQUE DE L'INTENSITE DE LA FLUCTUATION DE LA FORCE DE TRAVAIL DANS L'INDUSTRIE MINIERE

Utilisant certaines méthodes de la statistique théorique et de la statistique mathématique, l'auteur poursuit l'identification du modèle de l'intensité de la fluctuation de la force de travail dans l'industrie minière. Envisagent un cas particulier de ce domaine industriel, mais suffisamment significatif pour le problème abordé, on a relevé que, pendant une période relativement courte — de trois à quatre ans, le modèle caractéristique cherché a la forme d'une progression arithmétique. Pour une période de temps plus longue, l'auteur propose que l'investigation respective soit réalisée à l'aide d'une fonction de répartition, dans un cas ayant un large domaine d'existence dans l'industrie minière, si l'intensité de la fluctuation présente certaines conditions de régularité — stabilité ou évolution selon certaines lois.

СТАТИСТИКО-МАТЕМАТИЧЕСКОЕ ТОЛКОВАНИЕ ИНТЕНСИВНОСТИ ТЕКУЧЕСТИ РАБОЧЕЙ СИЛЫ В ГОРНОЙ ПРОМЫШЛЕННОСТИ

Применяя методы теоретической и математической статистики, автор пытается выявить модель интенсивности текучести рабочей силы в горной промышленности. Рассматривая частный случай в этой отрасли промышленности, достаточно показательный для данного вопроса, автор выявляет, что для сравнительно короткого периода — примерно 3—4 года искомая характерная модель имеет форму арифметической прогрессии. Для более длительного периода времени в статье предлагается проводить соответствующие исследования при помощи функции распределения в случае, часто встречающемся в горной промышленности, согда интенсивность текучести рабочей силы отмечает некоторые условия регулярности, ктабильности или эволюции по определенным закономерностям.