

AD619009

765-62825

APPLICATION OF THE STATISTICAL METHOD IN THE STUDY OF  
EPIDEMIOLOGICAL RELATIONS

Translation No. 1415

JUNE 1965

COPY	2	3	Q
POSTAGE		1.00	
REPRODUCTION		0.50	
			118

DDC  
RECEIVED  
AUG 16 1965  
DDC-IRA E

U. S. ARMY  
BIOLOGICAL LABORATORIES  
FORT DETRICK, FREDERICK, MARYLAND

ARCHIVE COPY

Translation No. 1415

DIS AVAILABILITY NOTICE

Qualified requestors may obtain copies of this document from DDC.

This publication has been translated from the open literature and is available to the general public. Non-DOD agencies may purchase this publication from Clearinghouse for Federal Scientific and Technical Information, U. S. Department of Commerce, Springfield, Va.

Technical Library Branch  
Technical Information Division

APPLICATION OF THE STATISTICAL METHOD IN THE STUDY OF EPIDEMIOLOGICAL RELATIONS

[Following is the translation of an article by M. N. Tkacheva, Gamaleya Institute of Epidemiology and Microbiology, AMN, USSR, appearing in the Russian-language periodical Zhurnal Mikrobiologii, Epidemiologii i Immunobiologii (Journal of Microbiology, Epidemiology and Immunobiology) #9, 1964, pages 60-64. It was submitted on 27 Jan 1964. Translation performed by Sp/7 Charles T. Ostertag Jr.]

The theory of epidemiology and the antiepidemic practice are promoting the necessity for developing a method for studying the relations which emerge during the development of the epidemic process. For obtaining objective criteria for evaluating the relationship it is necessary to apply the statistical method, which, up until the present time, has not yet occupied a proper place in epidemiological investigations.

In becoming acquainted with the appropriate literature, we were successful in encountering several works, where with greater or lesser success they applied the statistical analysis of the relation between phenomena.

Vanskaya (1947), when studying the number of *musca domestica* and, consequently, morbidity with intestinal infections under the conditions of urban living, made use of calculating the correlation coefficient, which confirmed the presence of a relationship between the level of morbidity and the number of flies.

Verbev (1961), when studying the dependency between mass immunization and morbidity with an inoculation form of epidemic hepatitis, made extensive use of determining the correlation coefficient between epidemic hepatitis incidence and the percentage of persons which had been immunized against typhoid, smallpox and diphtheria. The results obtained made it possible to confirm the hypothesis put forward concerning the absence of a link between mass immunization and morbidity with an inoculation form of epidemic hepatitis. Abou-Greeb (1960) studied the dependency of cholera morbidity in Calcutta on the density of the population and other factors. He was able to confirm a relation in the level of morbidity with the material well-being of the population and the organization of the water supply system in this or that sector of the city. A relationship between cholera morbidity and the density of the population was not confirmed statistically.

At first glance the results of comparing morbidity with the density of the population suggests doubts, however, even here an analysis of the material nature of the phenomena makes it possible to substantiate the results obtained by the statistical method. In the cited work there is talk about the density of the population, that is, the number of persons living in an area of 1 km<sup>2</sup>, which in our opinion can in no way be identified with the concept of congestion, which is actually related to the level of incidence with various infections. As an example it is possible to cite Belgium, where, with a very high population density, a very low level of infectious morbidity is observed (data from official statistics).

In the present work we have taken on the goal of acquainting the readers with elementary methods of studying relationships. To readers desiring a deeper acquaintanceship with this method, we recommend the following textbooks: B. S. Bessmertny and M. N. Tkacheva "Statistical Methods in Epidemiology" (1961), P. F. Rozhitskiy "Bases of Variation Statistics for Biologists" (1961), B. Hill "Bases of Medical Statistics" (1958).

In an objective world they distinguish two forms of relationships. They call a relation functional if for each measure of one phenomenon there is the corresponding strictly specific measure of another. In epidemiology and branches of medicine adjacent to it, relations persist during which changes of the actuating factor and the results of its influence may not occur parallel. In other words, an intimate relation takes place when (following numerous observations) a certain average change in the value of one corresponds to a specific change of the other. These relations are called correlations and for their study a number of methods are used which we will talk about below.

A determination of the correlation coefficient serves as the most general statistical method for measuring such relationships.

We applied correlation analysis for the determination of the influence of the value of the seasonal wave on the level of annual dysentery morbidity in localities.

We determined the volume of the seasonal wave with the help of the formula

$$S = \frac{B - \frac{A-B}{12M} \cdot M}{A} \cdot 100,$$

suggested by Kheyfets and Khazanov (1959), having changed somewhat the authors' suggested interpretation of the values entering into the formula:

$S$  -- the value of the true seasonal rise (in percentages);  $A$  -- the number of cases in a year;  $B$  -- the number of cases registered during the period of the seasonal rise;  $M$  -- the number of months of the seasonal rise, determined with a consideration of the average climatic conditions of the territory on which the morbidity analysis is being carried out.

Values, characterizing changes in the criteria being compared, are arranged in the form of two series. Series  $X$  is morbidity with acute dysentery in conditional indices and series  $Y$  is the value  $S$  in percentages. For each series of numbers the average value is determined, then for each number its deviation from the series average is established. The deviations obtained are fabricated into a square and added up. Twin deviations of both series are multiplied by each other and the products also added. The correlation coefficient is computed according to the quoted formula, where  $r_{xy}$  is the correlation coefficient,  $d_x$  and  $d_y$  - the deviations of the variant from the average of the series,  $\Sigma d_x d_y$  - addition symbol. The correlation coefficient obtained, equal to 0.86, characterizing the relation between the seasonal rise of morbidity with this infection in the RSFSR testifies to the presence of a very intimate relation between these phenomena (table 1).

It must be stressed that the correlation coefficient, just as other criteria for measuring the intimacy of a relation, serves only as a confirmation of the hypothesis advanced by researchers concerning the presence or absence of a relation between the phenomena being compared. At the same time, a considerable magnitude of the correlation coefficient during an analysis of the values of two, it would seem, non-related features, may compel researchers to more deeply analyze the phenomena under study and lead to the concept concerning the presence of material foundations for the existence of a relation between them.

The epidemiologist often has to investigate phenomena, when the variation is exhausted by two incompatible possibilities (alternative variation). This takes place, for example, during an investigation of the effectiveness of inoculations (inoculated -- not inoculated, became ill -- did not become ill). Measurement of the relation in these cases is conducted according to the so-called method of constructing a four symbol table, by means of the subsequent calculation of the association coefficient. An example, illustrating the method of determining the association coefficient, is taken from the current work of our laboratory, which at the present time is engaged in a study of the epidemiological role of various manifestations of diphtheria infection and, in particular, a study of the roles characterizing the ability to carry the diphtheria bacillus. One of the trends in the investigation was the study of the relation of the non-specific resistance of the children investigated and the ability to carry the diphtheria bacillus.

As a conditional test we utilized the Nesterov test characterizing the C-vitamin saturation of the human organism. In our work, which was carried out earlier (Zenkevich and Tkacheva, 1960), we presented a detailed characterization of the Nesterov test as an index of nonspecific resistance, resting on a considerable number of observations.

The data, obtained during the investigation of children from one of the children's homes were arranged by us in the form of a four symbol table, where  $a$  -- number of carriers, possessing a sufficient level of nonspecific resistance;  $b$  -- number of children, not yielding the diphtheria bacillus but having a sufficient level of nonspecific resistance;  $c$  -- number of carriers among the children with a lowered nonspecific resistance;  $d$  -- number of children with a lowered nonspecific resistance, in which the diphtheria bacillus was not isolated (table 2).

After substituting our specific data into the formula we obtain the association coefficient, equal to

$$-0.86 : \frac{ad-bc}{ad+bc} = \frac{28-384}{28+384} = \frac{-356}{412} = -0.86.$$

The minus sign demonstrated the reverse relation (the more children there were in the collective who had an adequate nonspecific resistance, the less carriers there were in it). The value 0.86 characterizes the high degree of intimacy in the relationship between nonspecific resistance and the ability to carry the diphtheria bacillus.

The association coefficient is used solely for measuring the intimacy of the relation between criteria characterizing qualitative indices.

A very simple method is known for measuring the intimacy of relationships. It is applicable for analyzing the relationships between quantitative and also between qualitative criteria. This is the method of determining the coefficient of rank correlation.

We present the following example as an illustration. It is commonly known that in cities the indices of morbidity with acute dysentery are higher than in rural areas. Since the oblasts have a various percentage ratio of urban and rural population, it may be proposed that a relationship exists between the proportion of the urban population in the oblast and the incidence of acute dysentery.

We will introduce two series of values -- the proportion of the urban population and morbidity with acute dysentery (table 3).

We will arrange the oblasts in the order of decreasing morbidity indices. In this way each oblast receives a rank (serial frequency

number) based on the index of morbidity and on the proportion of the urban population. The determination of the coefficient of rank correlation stems from the diversity of ranks for each of the two designated criteria for each oblast. If the values obtained are substituted in the formula presented in table 3, then  $P$  amounts to 0.51, which testifies to the direct, average level of the relationship between the level of dysentery morbidity and the proportion of the urban population. In this manner, our supposition concerning the presence of a relationship between these criteria was confirmed.

The methods presented and the examples cited concern only a comparison of the two factors and primarily to an analysis of a causative relationship. The majority of phenomena investigated by epidemiologists are found under the simultaneous influence of many factors. As a result a new problem emerges -- to attempt to break up this multifactorial complex and qualitatively characterize the values of each of the factors in this complex.

The study of this problem should begin with the application of the method of fractional correlation and dispersion analysis to epidemiological phenomena. The set up for carrying out dispersion analysis is quite complex and it is impossible to present it in a summary. We used the method of dispersion analysis in its most elementary form when studying the problem concerning to what degree the Nesterov test may be used as a conditional test for the characteristics of nonspecific resistance. The results obtained showed that the dominating influence on the degree of nonspecific resistance proves to be the endurance by a child of an infectious or catarrhal disease not long before the investigation was set up. Living conditions exerted three times less influence, and accidental factors, unaccounted for by us, were 12 times weaker than endurance of the disease.

We know of the successful application of this method in the works of Zatspein (1959) and Akatov and Lebedeva (1961).

It is obvious that the determination of the place of this method in epidemiological investigations has been presented insufficiently. This method can receive acceptance only as the result of its wide application in research.

The numerical expression of the degree of common variability of phenomena imparts great definiteness to conclusions and judgments obtained as a result of investigations. It becomes possible to compare the degrees of intimacy in the relationships of the phenomenon being studied with various factors. This in turn makes it possible to determine the factor exerting the dominating influence on the criterium being investigated.

The methodological nature of the article does not demand special conclusions.

### Literature

- a. Akatov, A. K., Lebedeva, Z. I., Antibiotiki (Antibiotics), 1961, No 4, page 363.
- b. Vanskaya, R. A., Zh. mikrobiol. (Journal of Microbiology), 1947, No 10, page 461.
- c. Verbev, P. E., Vopr. virusol. (Problems of Virology, 1960, No 4, page 462.
- d. Zenkevich, M. V., Tkacheva, M. N., Zh. mikrobiol., 1960, No 11, page 63.
- e. Abou-Gareeb, A. N., J. Trop. Med. Hyg., 1960, V 63, page 17.
- f. Kheyfets, L. B., Khazanov, M. I., Methods and Organization of Epidemiological Tests Based on the Study of the Effectiveness of Vaccines Against Intestinal Infections, Moscow, 1959.



Table 1  
Data concerning the relation between S and the level of morbidity in the USSR for a period of ten years.

Year	Morbidity (in conditional indices) X	$\frac{S}{y}$	$d_x$	$d_y$	$d_x^2$	$d_y^2$	$d_x \cdot d_y$
1953	1.00	43.8	0.15	15.0	0.0225	225.00	2.250
1954	1.05	43.2	0.20	14.4	0.0400	207.36	2.880
1955	0.90	32.7	0.05	3.9	0.0025	15.21	0.195
1956	0.76	25.4	-0.09	-3.4	0.0081	11.56	0.306
1957	0.95	36.0	0.10	7.2	0.0100	51.84	0.720
1958	0.71	22.9	-0.14	-5.9	0.0196	34.81	0.826
1959	0.92	31.7	0.07	2.9	0.0049	8.41	0.203
1960	0.78	10.4	-0.07	-18.4	0.0049	338.56	1.288
1961	0.70	22.2	-0.15	-6.6	0.0225	43.56	0.990
1962	0.69	20.0	-0.16	-8.8	0.0256	77.44	0.408

$$n = 10 \quad \Sigma x = 8.46 \quad \Sigma y = 288.3 \quad \Sigma d_x^2 = 0.1606 ;$$

$$M_x = 0.85 \quad M_y = 28.8 \quad \Sigma d_y^2 = 1013.75 ;$$

$$\Sigma d_x \cdot d_y = 11.066$$

$$r_{xy} = \frac{\Sigma d_x \cdot d_y}{\sqrt{\Sigma d_x^2 \times \Sigma d_y^2}} = \frac{11.066}{\sqrt{0.1606 \times 1013.75}} = \frac{11.066}{13.76} = 0.86$$

Table 2

Processing the results from the investigation of children from one children's home with the help of the four symbol table

Ability to carry diphtheria bacteria Nesterov test	+	-	In all
+	2(a)	48(c)	50
-	8(b)	14(d)	22
In all	10	62	72

Relation between morbidity with acute dysentery and the make up of the population in the city and the country (data of 1960 based on a selected circle of oblasts in the RSFSR)

Table 3

Name of the ASSR, Oblast or Kray	Morbidity (in conditional indices)	Proportion of municipal population (%)	Serial frequency number		Diversity of numbers	
			Based on morbidity level	Based on proportion level	$d$	$d^2$
Karelskaya ASSR	6.54	63	1	8	7	49
Khabarovskiy Kray	6.36	74	2	5	3	9
Primorskiy "	6.34	67	3	6	3	9
Kurganskaya Oblast	1.67	33	32	30	2	4
Voronezhskaya "	1.38	35	33	29	2	16
Belgorodskaya "	1.00	18	34	34	0	0

$n = 34$

$\sum d^2 = 3222$

$$P = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 3222}{34 \cdot 1155} = \frac{19338}{39270} = 0.51$$