E3S Web of Conferences **258**, 09067 (2021) https://doi.org/10.1051/e3sconf/202125809067
UESF-2021

Stochastic correlation, regression probabilistic-statistical models of laminated composite structures with material-energy-saving polyurethane thermal insulation

Victor Bobryashov[1], and *Nikolay* Bushuev[2]*

[1]TSNIISK Named After Koucherenko V.A. Research Center of Construction Joint Stock Company, 109428, Moscow, Russia,
[2] Moscow State University of Civil Engineering, 129337, Moscow, Russia

Abstract. After testing according to the developed state methods and procedures for polyurethane foam thermal insulation of laminated composite structures with metal facings, the issues of correlation, probabilistic-statistical regression models between the signs of elasticity in tension Eshear and signs of elasticity in shear Gshear were investigated. It is shown that the test results are random stochastic values. Their variability, depending on the type of tests and the parameter under study, is in the acceptable average values. The presence of a significant number of each of the characteristics, for example, Eshear - Gshift, predetermined the use of correlation tables to establish the fact of a relationship. Modules of elasticity under tension and shear were determined. The correlation coefficient between Eshear and Gshift is 0.659. With the help of a computer program, correlation tables were built and the calculation of the probabilistic-statistical interaction of characteristics with different reference points was made. When considering one-dimensional aggregates, the laws of distribution of positive values of the characteristics of foams can be adopted of various types. The connection between phenomena can be not only linear, but also non-linear. In this case, nonlinear correlation can be realized in the form of parabolic and other equations of a certain degree. The calculation of the interaction is carried out on the basis of second-order parabolic equations. Approximation in the form of second-order equations does not greatly improve the convergence, but the possibilities for extrapolating statistical models are reduced.

1 Introduction

In works [1 - 3], the authors, on the basis of numerous tests of samples, including full-scale (one-span, two-span), according to the methods and procedures approved by state standards, reliably showed that the test results are random stochastic values. Their variability (variation) depending on the type of tests and physical parameters (strength,

* Corresponding author: bush_ni@mail.run

elasticity, creep, thermal conductivity, etc.). It is in acceptable averages ($c_i \leq 33\%$, where c_i is the variance with Sheppard's corrections). Almost homogeneous for each individual parameter R_i, λ_i, Δ_i, etc. This is due to a typical continuous technological process on the same equipment.

Correlation here is understood as the dependence of the variability (variation) of a trait on the surrounding conditions according to Galton, Pearson - a ratio, a correspondence, which can be functional, correlative.

2 Research methods

In what follows, we will consider only complex correlations, the study of which solves a wide range of issues [4 - 7]. A preliminary analysis of the simulated population is carried out, the presence of the fact of connection, its direction and form is established, the degree of tightness of the connection between the parameters is measured. The construction of a probabilistic-statistical regression model (analytical expression of the relationship) is carried out, an assessment of the adequacy of the model, its economic interpretation and practical use are given.

Let us emphasize once again the essential scientific significance of correlation and regression. The presence of a significant amount of each of the features, for example, elasticity $E_{shear} - G_{shift}$ at N = 68, predetermined the use of correlation tables to establish the fact of the presence of a relationship (Table 1). The factorial attribute of elasticity E_{shear}, i is located in columns, the attribute G_{shift}, i is located in rows. At the intersection of rows and columns, the frequency of repetition of the combination E_{shear}, G_{shift} is indicated, the total value of which Σn_E, Σn_G determines the limiting laws of the probability distribution. The values of the tensile modulus of elasticity $\bar{E}$ = 99.27 Pa · 10^5, the shear modulus $\bar{G}$ = 37.56 Pa · 10^5 and their root-mean-square values $S_{E_{shear}}$ = 20.19 and S_G = 8.64 have been reliably established. Determination of fluctuations (changes) of elastic characteristics from these data with c_{Ep} = 20.19 / 99.27 = 0.203; c_G = 8.64 / 37.56 = 0.23, we are convinced that the correlation of features E_{shear} and G_{shift} is Gaussian normal.

3 Research results

Let us refer to table 1 of the distributions of the probabilities of the tensile modulus of elasticity E_{shear} and the shear modulus G_{shift} with the number of interaction pairs equal to 68.

As you can see from table 1, the probabilities are located from right to left. They are centered around a single straight line. Therefore, it is possible to hypothesize about the linear form of the interaction between the elastic characteristics. An analytical expression must be found for this interaction. And so that the sum of the squares of the deviations from the straight line reaches the smallest value. This will be the best fit of the theoretical curve to the experimental results. The equation of a straight line of regression of E_{shear} on G_{shift} (the indices for Ep and Gsdv will be omitted for convenience in the future) can be found from the test results:

$$E_i - \bar{E} = r_{EG} \frac{S_E}{S_G} (G_i - \bar{G}) \tag{1}$$

with the number of equidistant intervals in i = (1, 2 .... 8) and j = (1, 2.... 8). For simplicity of calculations, let's move on to conditional options using the following formulas:

$$G'_i = (G_i - \Theta_1)/C_1; \qquad E'_i = (E_i - \Theta_2)/C_2 \ , \tag{2}$$

where Θ_1, Θ_2 - "false zeros" adopted for the intervals with the highest frequency - G_6 = 39.9 Pa·10^5 and E_7 = 112.0 Pa·10^5; C_1 and C_2 - reduction factors for *G* and *E*, respectively, equal to 11.9 Pa · 10^5 and 4.3 Pa · 10^5.

Table 1. Calculation of the linear probabilistic-statistical interaction between the characteristics of the elastic and shear moduli

	G'_i	- 5	- 4	- 3	- 2	- 1	0	1	2	Σn_E	$u=\sum n_{GE}\cdot G'$	$U\cdot E'_p$
E'_{pi}	G_i / E_{pi}	18.4	22.7	27.0	31.3	35.6	39.9	44.2	48.5			
- 6	40.7	2								2	-10	60
- 5	52.6	1	1	1						3	-12	60
- 4	64.5							1		1	1	-4
- 3	76.4			3						3	-9	27
- 2	88.2		3	2	1		6	2		14	-18	36
- 1	100.1		1		2	3	5	3	3	17	-2	2
0	112.0			1	2	1	7	1	4	16	1	0
1	123.9						3	6	3	12	12	12
Σn_G		3	5	7	5	4	21	13	10			
$V=\sum n_{GE}\cdot E'$		-17	-12	-18	-4	-3	-14	-5	0			
$\vartheta\cdot G'$		85	48	54	8	3	0	-5	0			

$$\bar{G}' = -0{,}54; \quad \bar{E}'_{shear} = -1{,}07; \; S_{G'} = 2{,}01; \quad S_{E'} = 1{,}70; \quad E_{shear} = 1{,}54G_i + 41{,}43\text{Па}\cdot 10^5; \quad r_{EG} = 0{,}659$$

$$E_i - \bar{E} = r_{EG}\frac{S_E}{S_G}(G_i - \bar{G}) \tag{3}$$

with the number of equidistant intervals in *i* = (1, 2 .... 8) and *j* = (1, 2.... 8). For simplicity of calculations, let's move on to conditional options using the following formulas:

$$G'_i = (G_i - \Theta_1)/C_1; \qquad E'_i = (E_i - \Theta_2)/C_2 \; , \tag{4}$$

where Θ_1, Θ_2 - "false zeros" adopted for the intervals with the highest frequency - G_6 = 39.9 Pa·10^5 and E_7 = 112.0 Pa·10^5; C_1 and C_2 - reduction factors for *G* and *E*, respectively, equal to 11.9 Pa · 10^5 and 4.3 Pa · 10^5.

For convenience, the correlation table, taking into account the "false zeros", is combined with the values and obtained by the formula (2), they are shown on the side panels of the table:

$$G'_i : \; -5, -4, -3, -2, -1, 0, 1, 2$$

$$E'_{i\ shear} : -6, -5, -4, -3, -2, -1, 0, 1.$$

We determine the average values of $\bar{G}'$ and $\bar{E}'$, taking into account the "false zeros" (where the dash above G and E means the average values):

$$[\;\overline{G}'\;] = \frac{\sum n_G \cdot G'}{n} = 3(-5)+(-4)+(-3)+5(-2)+4(-1)+21(0)+13(1)+10(2)/68=-0{,}544$$

$$[\bar{E}'] = \left(\frac{\sum n_E \cdot E'}{n}\right) = -1{,}079$$

Here n is the number of test pairs equal to 68. Let us find the auxiliary values $(\overline{G'})^2$ and $(\overline{E'})^2$:

$$(\overline{G}')^2 = \frac{\left[\sum n_G(\overline{G}')\right]^2}{n} = (3\cdot 25 + 5\cdot 16 + 7\cdot 9 + 5\cdot 4 + 4\cdot 1 + 13\cdot 1 + 10\cdot 4)/68 = 4{,}34$$

$$(\overline{E}')^2 = \frac{\left[\sum n_E(\overline{E}')\right]^2}{n} = 4{,}044$$

Let's determine the values of the standard deviations taking into account the "false zeros":

$$S_{G'} = \sqrt{(\overline{G})^2 - [\overline{G}']^2} = \sqrt{4{,}34 - 0{,}54^2} = 2{,}01$$

$$S_{E'} = \sqrt{(E')^2 - [\overline{E}']^2} = 1{,}70.$$

Then the values of the standard deviations are equal:
$S_G = S_{G'} \cdot C_1 = 2{,}01 \cdot 4{,}3 = 8{,}64$ Па·10^5; $S_E = S_{E'} \cdot C_2 = 1{,}7 \cdot 11{,}9 = 20{,}19$ Па·10^5
Here C_1, C_2 are the values of the reduction coefficients equal to 4.3 and 11.9 Pa ·10^5, respectively.
Let's calculate the correlation coefficient r_{EG}:

$$r_{EG} = \frac{(\sum n_{GE} G' E' - n_{GE} \bar{G}' \bar{E}')}{n_{GE} S_{G'} \cdot S_{E'}} \tag{5}$$

First, determine the value of $\sum n_{GE} G'E'$. To do this, compose the product of the sum of frequencies Σn_{GE} by the variants $G' - n_{GE} \cdot G'$ and write these products in the upper right corner of the cells containing the corresponding frequencies. Similarly, we compose the product Σn_{GE} by options $E' - n_{GE} \cdot E'$ and write in the lower left corner of the table. For example, for a cell n_{GE} is equal to 2, the product of $\Sigma n_{GE} \cdot G' = 2 \cdot (-5) = -10$ and $\Sigma n_{GE} \cdot E' = 2 \cdot (-6) = -12$. Summing the values of $\Sigma n_{GE} \cdot G'$ and $\Sigma n_{GE} \cdot E'$ over the rows and columns of the correlation table, we obtain the values of U and ϑ:

$$U = \Sigma n_{GE} G' \quad u \quad \vartheta = \Sigma n_{GE} E' \tag{6}$$

For example, for the first column: $\vartheta = \Sigma n_{GE} \cdot E' = -12 - 5 = -17$, for the second line:

$$U = \Sigma n_{GE} \cdot G' = -5 - 4 - 3 = -12.$$

These amounts are written in the sidebars of Table 1:

U: -10, -12, 1, -9, -18, -2, 1, 12
ϑ: -17, -12, -18, -4, -3, -14, -5, 0

Find the product of U by E' and ϑ by G' and obtain:

$U \cdot E'$ 60, 60, -4,0, 27, 36, 2, 0, 12;

$$\vartheta \cdot G' \quad 85, 48, 54, 8, 3, 0, -5, 0$$

Summing up the values of the obtained series taking into account the signs, we have:

$$\sum U \cdot E' = 193, \quad \sum \vartheta \cdot G' = 193$$

The convergence of the sums means that all calculations were performed correctly. Then the sought correlation coefficient according to the formula (3) will be equal to:

$$r_{EG} = \frac{193 - 68 \cdot (-0{,}544) \cdot (-1{,}073)}{68 \cdot 2{,}01 \cdot 1{,}70} = 0{,}659$$

The average values of the characteristics of elasticity in tension and shear are found by the formulas:

$$\bar{E} = \bar{E}'^{C_2} + \Theta_2 = (-1{,}073) \cdot 11{,}9 + 112{,}0 = 99{,}27 \text{ Па} \cdot 10^5$$
$$\bar{G} = \bar{G}'^{C_1} + \Theta_1 = (-0{,}544) \cdot 4{,}3 + 39{,}9 = 37{,}56 \text{ Па} \cdot 10^5$$

The interaction between E and G is found from the equation:

$$E_{sheai} - \bar{E} = r_{GE} \frac{S_E}{S_G} (G_i - \bar{G})$$

Substituting the values of the obtained parameters into the equation, we have:

$$E_{sheari} - 99{,}27 = 0{,}659 \cdot \frac{20{,}19}{8{,}64} (G_i - 37{,}56) = 1{,}54 G_i - 57{,}84$$
$$E_{shear\, i} = 1{,}54 G_i + 41{,}43 \text{ Па} \cdot 10^5.$$

So, we have obtained a formula showing the probabilistic-statistical interaction between the characteristics E_{shear} and G_{shift} with a linear correlation between them (Figure 1).

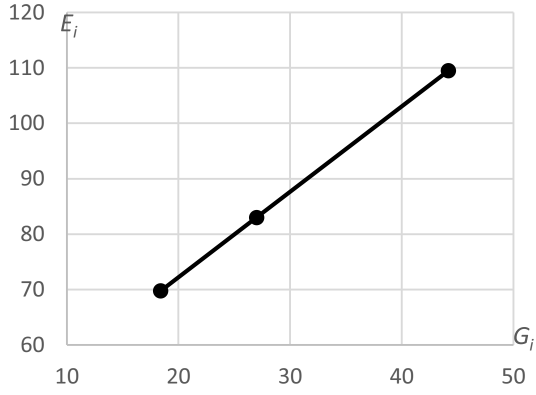

Fig. 1. Regression of elastic modulus E and shear modulus G

It meets the conditions for the best approximation to the experimental results:

E_{shear}^{theor}	76,4	83,0	89,6	96,3	102,8	109,4	116,1	Па·10 ^{5}
E_{shift}^{exper}	83,5	81,4	102,5	103,1	104,1	107,4	112	Па·10 ^{5}

Here E_{shear}^{theor} and E_{shift}^{exper} are the theoretical and experimental values of the tensile modulus.

The above algorithm was used to compose a computer program. According to the program, correlation tables were obtained and the calculation of the probabilistic-statistical interaction between characteristics with different reference points was made. At the same time, the values of the reduction factors varied within wide limits. This was achieved by varying the number of intervals for i and j. The calculation of the probabilistic-statistical interaction was carried out for both square and rectangular initial matrices with the number of intervals from five to twelve. In this case, as a rule, the parameters of equations (1 - 4) - mean values, standard deviations, correlation coefficients, free terms were somewhat different from each other. For example, for this calculation with a 5 x 5 matrix, they were equal: $G = 39.19$ Pa·10^5; $E = 102.3$ Pa·10^5; $S_G = 8.4$ Pa·10^5; $S_E = 20.32$ Pa·10^5; $r_E = 0.583$. The grouping of test results affects the final results, but not significantly. The calculation of the characteristics of interaction with a new reference point increases the reliability of the calculations, allows for a comprehensive assessment of the phenomena under study. It is especially needed when considering non-linear correlation in a limited number of trials.

As we have established when considering one-dimensional aggregates, the distribution laws of essentially positive quantities, which are all the characteristics of foamed plastics, can be adopted in various forms. This implies one important point - the correlation between phenomena can be not only linear (in the case of normal laws or those close to them), but also nonlinear. In this case, the nonlinear correlation can be realized in the form of parabolic equations of a certain degree: $y = a_o + a_1x^1 + a_2x^2 + \ldots a_k x^n$, non-parabolic correlation equations: power-law $y = ax^b$, exponential $y = ab^x$, logarithmic $y=a + b\lg x$, polynomials, etc. [8 - 12]. As noted in [1, 4, 13], the type of the correlation equation is determined by the nature of the probabilistic-statistical interaction between the studied random variables. As a rule, this character is determined by the normal correlation in the plane. In the case of nonlinear correlation, the problem will be solved according to the principle proposed by P.L. Chebyshev [4, 7] - a gradual increase in the order of the correlation equation. First, we calculate the probabilistic-statistical interaction using the first-order correlation equations «*n*». With their insufficient convergence, especially with insignificant correlation coefficients, we increase the order of the equation to $(n+1)$. If the interaction in this case, the approximation is not enough, we pass to equations of order $(n+2)$, etc. We will assume that if the results obtained by the theoretical calculation are "covered" by the experimental confidence limits, then the approximation at a given level of significance is $\alpha = 0{,}05$; $\alpha = 0{,}01$ is permissible and meets the assigned tasks

The decision on the probabilistic-statistical interaction between characteristics using first-order equations was given earlier. Consider the calculation of the interaction based on second-order parabolic equations:

$$y = ax^2 + bx + c \tag{7}$$

In this case, we will use the same experimental results as for linear correlation. The parameters of function (5) will be found from the condition of minimizing the mean square of deviations [5, 7, 12]:

$$Q = \sum_x n_x (\bar{y}_ч - c - bx - ax^2)^2 = \min \tag{8}$$

It follows from this condition that all partial derivatives of the sum Q with respect to the required parameters must be equal to zero:

$$\frac{dQ}{dc} = 0, \qquad \frac{dQ}{db} = 0, \qquad \frac{dQ}{da} = 0 \tag{9}$$

Differentiation and transformation of formulas (6 - 7) leads to a system of equations of the form [4, 6, 8, 10]:

$$\begin{aligned} &(\sum n_x x^4)\cdot a + (\sum n_x x^3)\cdot b + (\sum n_x x^2)\cdot c = \sum n_x \bar{y}_x \cdot x^2 \\ &(\sum n_x x^3)\cdot a + (\sum n_x x^2)\cdot b + (\sum n_x x\ )\cdot c = \sum n_x \bar{y}_x \cdot x \\ &(\sum n_x x^2)\cdot a + (\sum n_x x\ )\cdot b + n_x \cdot c = \sum n_x \bar{y}_x \cdot x^2 \end{aligned} \quad (10)$$

Let us estimate the probabilistic-statistical nonlinear interaction between the elastic and shear moduli. Using the results given in correlation table 1 and formulas (8). The calculation results are shown in Table 2.

Table 2. Calculation of the parameters of the nonlinear probabilistic-statistical interaction of the elastic and shear moduli

G	n_G	$\overline{G}$	$n_G \cdot G$	$n_G \cdot G^2$	$n_G \cdot G^3$	$n_G \cdot G^4$	$n_G \cdot \overline{G}$	$n_G \cdot \overline{G} G$	$n_G \cdot \overline{G} G^2$
18.4	3	4.4(1)	5.52(1)	9.72(2)	1.75(4)	3.15(5)	1.34(2)	4.47(3)	4.54(4)
22.7	5	8.35(1)	1.13(2)	2.64(3)	6.08(4)	1.4(6)	4.17(2)	9.48(3)	2.15(5)
27.0	7	8.15(1)	1.89(2)	5.1(3)	1.38(5)	3.72(6)	5.70(2)	1.54(4)	4.15(5)
31.3	5	1.02(2)	1.56(2)	4.8(3)	1.49(5)	4.62(6)	5.12(2)	1.6(4)	5.02(5)
36.6	4	1.03(2)	1.42(2)	5.2(3)	1.86(5)	0.71(6)	4.12(2)	1.47(4)	5.22(5)
39.9	21	1.04(2)	8.37(2)	3.36(4)	1.34(6)	5.37(7)	2.18(3)	8.72(4)	3.48(6)
44.2	13	1.07(2)	5.74(2)	2.51(4)	1.1(6)	4.87(7)	1.4(3)	6.17(4)	2.73(6)
48.5	10	1.12(2)	4.85(2)	2.4(4)	1.17(6)	5.76(7)	1.12(3)	5.43(4)	2.63(6)
∑	68	0	2.55(3)	1.05(5)	4.18(6)	1.77(8)	6.75(3)	2.61(5)	1.05(7)

Substituting the obtained values of the sums (the last row of Table 2) into equation (8), we have:

$$\begin{aligned} 1{,}77(8)a + 4{,}18(6)b + 1{,}01(5)c = 1{,}05 \\ 4{,}18(6)a + 1{,}01(5)b + 2{,}55(3)c = 2{,}61 \\ 1{,}01(5)a + 2{,}55(3)b + 68c = 6{,}75 \end{aligned}$$

Solving this equation using determinants or the Gaussian method, we obtain the unknown parameters a = - 0,043; b = 4,25; c = 3,61.

Therefore, the probabilistic-statistical interaction between E and G is determined by a second-order nonlinear equation: $E_i = -0{,}043G_i^2 + 4{,}25\ G_i + 3{,}61$.

A comparison of the experimental and theoretical results obtained using this formula is given below:

E^{theor}	77,93	87,06	94,51	100,42	104,73	107,46	108,63
E^{exper}	83,5	81,4	102,5	103,1	104,1	107,4	112,02

The convergence of experimental and theoretical results with a nonlinear correlation between E and G improved on average by 1 - 2% at each interval in comparison with the linear correlation. The obtained theoretical results are completely "covered" by the confidence limits for the experimental results with the coefficient of C_E variability – 15 - 20%. Taking into account that the approximation of theoretical results to experimental ones does not exceed 2% on average and is "covered" by confidence limits, taking into account random fluctuations of experimental results, we believe that further approximation in the form of parabolas of the third or fourth order does not make sense.

According to the given algorithm in the form (8), a computer program was compiled. The program is designed in such a way that at the beginning the parameters of the linear correlation equations are determined and the linear probabilistic-statistical interaction is calculated. Using some necessary results of this calculation, we determine the parameters of nonlinear correlation equations. For comparison, the experimental conditional means are determined. When analyzing the results of marginal distributions for two characteristics and parameters of the obtained linear and nonlinear equations, a decision is made on the type of connection and on a further increase in the order of parabolic equations.

As shown by the analysis, the most expedient is linear correlation, since the interaction between characteristics is determined by the normal laws of their distribution. The second approximation in the form of second-order equations improves convergence by only one to two percent, but the extrapolation capabilities of statistical models are reduced.

4 Conclusions

1. Tests of elastic properties of polyurethane foam thermal insulation E_p,, $G_{сдв}$ were carried out according to methods, procedures, state standards, developed by the authors with the number of interaction pairs N = 68.

2. The test results are grouped in the form of a correlation table using the Sturges interval widths.

3. The values of the marginal probability distributions were obtained: mean value $\bar{E}$, $\bar{G}$, variance, variability (variation) c_{Ep}, c_G. They indicate their normality (linearity) in the Gaussian plane.

4. An estimate of the correlation coefficient $r_{EG} = 0{,}659$ is given for a straight-line relationship between E_p and $G_{сдв}$. It is compared with the curvilinear correlation of the parabolic type with a difference of no more than two percent. A comparison table is provided.

5. A regression analytical model Ep - Gsdv is given, which makes it possible to predict the development of *Ep, Gsdv* by linear probabilistic-statistical interaction.

References

1. V.M. Bobryashov, T.D. Bobryashova, S.I. Tsaritsenskaya, *Two-dimensional laws of probability distribution of highly filled foams and their determination on a computer. Investigation of building structures with the use of polymeric materials* (Interuniversity collection of scientific papers, Voronezh, 1987).
2. A.A. Gusak, E.A. Brichkova, *Probability theory* (Tetra-System, Minsk, 1999)
3. Andrei Glasunov, Viktor Bobryashov, Nikolay Bushuev, Building Platforms of Polymer Matrix Based Composite Construction Materials for Transportation Facilities, International Scientific Conference Energy Management of Municipal Transportation Facilities and Transport EMMFT 2017, Advances in Intelligent Systems and Computing **692**, (2017)
4 A.K. Mitropol'skiy, *Technique of statistical calculations* (Nauka, Moscow, 1971)
5. G.I. Peresvetov, *Theory of Probability and Mathematical Statistics* (Alt-Press, , 2009)
6. M.G.Venetskiy, V.I. Venetskaya, *Basic mathematical and statistical concepts and formulas in economic analysis* (Statistics, Moscow, 1979)
7. M.R. Efimova, I.E. Varakina and other, *General theory of statistics* (Infra-M, Moscow, 1997)

8. A.M. Long, *Mathematical statistics in technology* (Soviet Science, Moscow, 1958).
9. B.V. Gnedenko, *Probability theory course* (Nauka, Moscow, 1969).
10. N.V. Smirnov, N.V. Dunin-Barkovsky, *Course in Probability Theory and Mathematical Statistics for Technical Applications* (Nauka, Moscow,1969)
11. V.E. Gmurman, *Theory of Probability and Mathematical Statistics* (Yurayt, M., 2010)
12. B.P. Demidovich and other, *Numerical methods of analysis* (State. ed. phys.-mat. lit., M., 1963)
13. N.K. Hilliard, *Applied Mechanics of Cellular Plastics* Mir, Moscow, 1985)
14. GOST 22695-77. Panels of walls and building coverings are layered with foam insulation. Styrofoam. Strength test method. - Introduction. 01/01/1979.
15. GOST 21562-76. Metal panels with foam insulation. General technical conditions. - Introduction. 01/01/1977.
16. GOST 23486-79. Three-layer metal wall panels with polyurethane foam insulation. Technical conditions. - Introduction. 01/01/1980.
17. Y.R. Vainu, *Correlation of series of dynamics* (Statistics, Moscow, 1977)
18. R. Storm, *Probability theory, mathematical statistics, statistical quality control* (Mir, Moscow, 1970)