ON THE PROCESSING OF EXPERIMENTAL DATA BY METHODS OF MATHEMATICAL STATISTICS

[1]Vahobov Valijon
[1]Associate Professor of the Tashkent Institute of Irrigation and Agriculture Mechanization Engineers, Tashkent city, Uzbekistan,

[2]Khidoyatova Muyassar
[2]Assistant of the Tashkent Institute of Irrigation and Agricultural Mechanization Engineers, Tashkent city, Uzbekistan.

ABSTRACT
The work with specific examples examines the main methods and techniques for processing the results of an experiment, the knowledge of which is necessary for a modern economist in analyzing the mass data on agriculture. The use of mathematical statistics methods for calculating specific indicators is shown.
KEY WORDS: *general totality, sample, variation series*

INTRODUCTION

In finding optimal solutions to vital problems in economics, technology, medicine, agriculture and other branches of science, one has to deal with aggregates of statistical data obtained as a result of an experiment.

The main responsibility of the experimenter is to substitute solid, purposeful experiments. Mathematical statistics helps in technical, economic and agronomic research in choosing the optimal conditions for conducting experiments, gives an objective, quantitative assessment of experimental data.

The mass data obtained as a result of scientific and practical experiments, in most cases, are of a probabilistic-random nature. To process this data, it is necessary to apply the methods of mathematical statistics.

Discrete are such quantitative characteristics that can take only discontinuous (integer) values. For example, the number of family members, the number of machines, cars, the number of cracks in a metal disc.

Quantitative characteristics that can take any numerical values within certain limits are called continuous. For example: age, work experience, yield, production cost. Now we will give some initial concepts of mathematical statistics.

Depending on the completeness of the survey of units of the population, a general and a sample population are distinguished. The totality of phenomena from which a part of the units is selected for sample observation is called the general population. The same part of the units that are selected from the general population for sample observation is called the sample population or simply the sample. The number of elements in the general population and the sample is called their volumes and denoted by N and n, respectively.

Let the sample size be extracted from the general population n, moreover x_1 occurred n_1 once, $x_2 - n_2$ once, $x_k - n_k$ once $\sum_{i=1}^{k} n_i = n$. Observed value x_i they are called variants, and the sequence of variants written in ascending order is called variational series. The number of observations is called frequencies, and their ratio to the sample size $\dfrac{n_i}{n} = W_i$ - relative frequencies or frequencies.

MATERIALS AND METHODS

Statistical (discrete or interval) a sample distribution is a list of options and their corresponding frequencies or relative frequencies. Usually the statistical distribution of the sample is set as a table:

x_i	x_1	x_2		x_k	$\sum_{i=1}^{k} n_i = n$
n_i	n_1	n_2		n_k	
w_i	w_1	w_2		w_k	$\sum_{i=1}^{k} w_i = 1$

If the random variable under study is continuous or discrete and the number of possible values is large, it is convenient to construct an interval (variational) distribution series. An interval variation series is an ordered set of intervals for varying the values of a random variable with corresponding frequencies of hits in each of these values. To determine the number of intervals L of artificial grouping use the Sturges formula

$$L = 1 + 3,322 \lg n \qquad (1)$$

Here is the sequence of the procedure for grouping an unordered sample from the General population: 1.formation of a variation series; 2. selection of the minimum and maximum elements of the sample:

$$X_{\min} = X(1), \quad X_{\max} = X(n);$$

3. determining the width of grouping intervals (for equal-flow grouping) $h = \dfrac{x(n) - x(1)}{L} \qquad (2)$

If the evaluation of h необходимо to round the result, keep in mind that the last grouping interval will be smaller than the width h when rounding up or down h when rounding in the smaller party.

4. Formation of a sequence of borders of intervals of splitting by samples of a variational series the borders of intervals of grouping look like.

$$X(1), X(1) + h, \quad X(1) + 2h, ..., X(1) + (L-1)hX(n)$$

Sometimes in order to $X(1)$ и $X(n)$ got inside, respectively 1- go and L- number of grouping intervals,

borders $X(1)$ и $X(n)$ adjust as follows: $X'(1) = X(1) - \dfrac{h}{2} \qquad X'(n) = X(n) + \dfrac{h}{2}$

Consequently, the number of split intervals increases by 1T. e. $L^1 = L + 1$.

In this case, the sequence of boundaries of the partition intervals will be represented as

$$X^1_{(n)}, X^1_{(n)} + h, X^1_{(n)} + 2h, ..., X^1_{(n)} L * h, X_{(n)}$$

5. Determining the number of selection elements n_i, included in each j interval.

2. Statistical characteristics of the quantitative (variability) feature.

The main statistical characteristics of quantitative variability are the sample arithmetic mean $\overline{x_B}$, sample variance S^2, sample mean square deviation S, error of the arithmetic average $S_{\overline{x}}$, coefficient of variation V and the relative error of the sample average $S_{\overline{x}}\% [3,4]$.

Quantitative attributes include those that can be characterized quantitatively-the yield from the plot, the number of heights and weight of plants, the content of protein and gluten in the grain, and so on.

A sample consisting of 20-30 observation units is called small, and a large sample is called large $(n > 30)$ – big.

Statistical characteristics are calculated using the formulas in table 1.

In this table, separate values of the attribute in small samples and group averages in large samples, converted values of the original dates are indicated by X; A-arbitrary beginning, conditional average, -frequency, number of the group; -sample size; t-theoretical value of the student's criterion

Tab №1

Indicator Small sample (not grouped data) Large sample (grouped dat	Indicator Small sample (not grouped data) Large sample (grouped dat	Indicator Small sample (not grouped data) Large sample (grouped dat
Sample arithmetic mean	$\overline{X}_B = \dfrac{1}{n}\sum x_i = A + \dfrac{1}{n}\sum x_i'$	$\overline{X}_B = \dfrac{1}{n}\sum x_i n_i = A + \dfrac{1}{n}\sum x_i' n_i$
Sample variance	$S^2 = \dfrac{\sum\left(x_i - \overline{x}_B\right)^2}{n-1}$	$S^2 = \dfrac{\sum\left(x_i - \overline{x}_B\right)^2 \cdot n_i}{n-1}$
Sample standard deviation	$S = \sqrt{S^2}$	
Coefficient of variation	$V = \dfrac{S}{\overline{X}_B}100\%$	
Average error	$S_{\overline{x}} = \dfrac{S}{\sqrt{n}} = \sqrt{\dfrac{S^2}{n}}$	
Relative very much average	$S_{\overline{x}}\% = \dfrac{S_{\overline{x}}}{\overline{X}_B}100$	
Confidence interval for the average value	$\left(\overline{X}_B - tS_{\overline{x}}; \overline{X}_B + tS_{\overline{x}}\right)$	
Degree of freedom	$n-1$	

Here is an example for calculating statistical characteristics in the case of a small sample.

Example 1. When determining the phosphorus content in plant material, the following results were obtained (in g per 100 g of dry matter); 0,56; 0,53; 0,49; 0,57; 0,48. Need to calculate $\overline{X}_B, S_{\overline{x}} = 95\%$ - and 99%- other confidence intervals for the average value of the population.

Decision.

It is convenient to calculate statistical characteristics in the following sequences:

$$\overline{X}_B = \frac{1}{n}\sum_{i=1}^{5} x_i = \frac{2,63}{5} = 0,526 \Gamma; \quad S^2 = \frac{\sum\left(x_i - x_i'\right)^2}{n-1} = \frac{0,00652}{5-1} = 0,0016;$$

$$S = \sqrt{S^2} = \sqrt{0,0016} = 0,04 \Gamma; \quad V = \frac{s}{x_i}\cdot 100\% = \frac{0,04}{0,526}\cdot 100\% = 7,60\%$$

$$S_{\overline{x}} = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{0,0016}{5}} = 0,018 \Gamma; \quad S_{\overline{x}}\% = \frac{S_x}{\overline{X}_B}\cdot 100\% = \frac{0,018}{0,526}\cdot 100\% = 3,42\%$$

$$\overline{X}_B \pm t_{0,5} S_{\overline{x}} = 0,526 \pm 2,8\cdot 0,018 = 0,526 \pm 0,050 \left(0,48 \div 0,58\right)\Gamma$$

$$\overline{X}_B \pm t_{0,1} S_{\overline{x}} = 0,526 \pm 24,6\cdot 0,018 = 0,526 \pm 0,018 \left(0,44 \div 0,61\right)\Gamma$$

DISCUSSIONS AND RESULTS

Now let's look at an example for calculating statistical characteristics in the case of a large sample (grouped data).

Example 2. As a result of measuring the technical length of the stem (cm) in 100 flax plants, the following data were obtained:

90,1	76,2	79,9	45,4	72,4	70,7	79,1	77,0	92,1	89,4
109,9	82,2	81,4	59,1	68,5	67,0	78,0	76,1	91,5	85,4
99,1	80,0	84,0	60,1	80,7	100,4	83,9	88,1	70,7	93,1
100,1	68,4	108,2	63,3	81,2	103,4	92,2	89,7	79,0	90,0
115,3	69,4	83,3	73,2	84,4	69,0	93,2	94,1	73,5	79,0
68,0	74,4	81,7	87,0	77,0	72,4	81,3	82,0	84,4	83,0
70,4	72,2	94,4	94,7	79,8	74,4	82,0	80,1	79,7	91,0
72,3	69,4	98,0	91,5	81,6	66,1	86,4	81,0	84,0	87,2
73,0	80,0	102,4	88,2	84,3	67,3	89,1	77,0	79,6	80,3
70,1	59,2	101,7	90,1	50,2	52,0	93,5	80,0	84,1	54,7

Based on this data, you need to:1) make an interval distribution of the sample;2) Calculate statistical characteristics $\overline{X}_B, S, V, S_{\overline{x}}, \overline{x} \pm t_{0,5}S_{\overline{x}}$

Decision. To solve the problem, the following sequence is suggested.

1.Set the number of groups (classes), the value of the interval, the beginning and end of each group and group intervals, find the midpoints of the intervals.

We determine the number of partition intervals using the Sturges formula (1)

$$L = 1 + 3,32 \lg 100 = 1 + 3,32 \lg 10^2 = 1 + 2 \cdot 3,32 \lg 10 = 1 + 6,64 = 7,74 \approx 8$$

Find the length of the partition interval h by formula (2)

$$h = \frac{x_{\max} - x_{\min}}{L} = \frac{115,3 - 45,4}{8} \approx 8,74$$

The middle of each interval is found by the formula $x_i' = x_{\min} + \left(i - \frac{1}{2}\right)h = 45,4 + \left(i - \frac{1}{2}\right) \cdot 8,74$

Then the interval distribution of the sample has the form:
Tab №2

Number of intervals	Partial intervals	The number of options that fall within this interval	Midpoint of the interval
$i = L$	$[x_1; x_{1+1}]$	n_i	x_i^*
1	$[45,4; 54,14]$	3	49,77
2	$[54,14; 62,88]$	4	58,51
3	$[62,88; 71,62]$	13	67,25
4	$[71,62; 80,36]$	28	75,99
5	$[80,36; 89,10]$	26	84,73
6	$[89,10; 97,84]$	15	93,47
7	$[97,84; 106,58]$	8	102,21
8	$[106,58; 115,32]$	3	110,95
$\sum$		100	8138,88

Thus, after grouping, a short, easily visible variation series is obtained, which allows us to judge the nature of the variability of the studied trait. So, the most common plants are those with a technical stem length 80-89,9 sm.

2. Let's calculate the statistical characteristics of the trait under study. To do this, make the following calculation table number 3

According to table 3, we calculate the statistical characteristics of the variation series and the confidence interval for the average.

1) Sample arithmetic mean (weighted) $\bar{x}_B = \frac{1}{n}\sum_{i=1}^{8} n_i x_i^* = \frac{8160}{100} = 81,6 см;$

2) Sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{8} n_i\left(x_i^* - \bar{x}_B\right)^2 = \frac{1}{99}\cdot 17244 = 174,2;$

3) Standard deviation (individual observation error) $S = \sqrt{S^2} = \sqrt{174,2} = 13,2 см$

4) it coefficient of variation $V = \frac{S}{\bar{x}_B}\cdot 100 = \frac{13,2}{81,6}\cdot 100 = 16,2\%;$

5) Sample mean absolute error $S_{x_B}^* = \frac{s}{\sqrt{n}} = \frac{13,2}{10} = 1,32 см;$

6) Sample mean relative error $S_{x_B}^*\% = \frac{S_{x_B}^*}{\bar{x}_B}\cdot 1000 = \frac{13,2}{81,6}\cdot 1000 = 1,6\%;$

7) The confidence interval of the general average for the 5% significance level at $n-1 = 100-1 = 99$

degrees of freedom variations $\left(t_{0,5} = 1,981\right)$

$$\bar{x}_B \pm t_{0,5}S_{\bar{x}} = 81,6 \pm 1,98\cdot 1,32 = 81,6 \pm 2,6\left(79,0 \div 84,2\right)$$

Grouping data and calculating statistical characteristics with continuous variability

Tab №3

Group (intervals)	Number of options	Middle of interval	Calculating sum of squares				According to the converted data $x_i^* = (x_i - A):(x_i - 85):10$			
			According to the initial data x_i^*							
x_i / x_{i+1}		x_i^*	$n_i x_i^*$	x_i^{*2}	nx_i^{*2}	$x_i^{'}$	$nx_i^{'}$	$x_i^{'2}$	$nx_i^{'2}$
40-49,9	1	45	45	2025	2025	-4	-4	16	16
50-59,9	5	55	275	3025	15125	-3	-15	9	45
60-69,9	11	65	715	4225	46475	-2	-22	4	44
70-79,9	26	75	1950	5625	146250	-1	-26	1	26
80-89,9	33	85	2805	7225	238425	0	0	0	0
90-99,9	16	95	1520	8025	144400	1	16	1	16
100-109,9	7	105	735	11025	77175	2	14	4	28
110-120	1	115	115	13225	13225	3	3	9	9
	100	-	8160	-	683100	-	-34	-	184
$\sum$			$1/n\sum n_i x_i^* = \frac{1}{100}\cdot 8160 = 81,6$				$A + 1/n\cdot\sum n_i x_i^{'} K = 85 + 0,34\cdot 10 = 81,6$			
$\sum n_i x_i K$			$\sum n_i x_i^{*2} - \frac{\left(\sum n_i x_i^*\right)^2}{n} = 683100 - \frac{(8160)^2}{100} = 1722$				$\left[\sum n_i x_i^{'2} - \sum n_i x_i^{'} K^2/n\right]\cdot K = \left[184 - (-34)^2/100\right]\cdot 10^2 = 17244$			

Output. Thus, the average of the entire population with a 95% level of probability is in the range of cm, the absolute error of the sample mean is 1.32 cm, the relative error is 1.6%; the coefficient of variation of the technical length of flax is 16.2%.

CONCLUSION

In conclusion, we note that the following conclusions can be drawn from the material presented above:

1. In the work, the presented method of processing experimental data is used in solving agronomic, economic or technical and other problems.

2. In the article, using specific examples from agriculture, the processing of data by methods of mathematical statistics is described and the corresponding conclusions are drawn.

3. Methods of mathematical statistics, processing of experimental data proposed in the article will be useful to specialists engaged in experimental work.

REFERENCE

1. *Gataulin A.M., Kharitonova L.A., Gavrilov G.V. Economic and mathematical methods in planning agricultural production. –M.: Kolos, 1986*
2. *Boyarsky A.L. Mathematics for Economists. –M.: 1957*
3. *Zaitsev I.A. Higher mathematics. –M .: "High school" 1991*
4. *Marmoza A.T. Workshop on Mathematical Statistics. Kiev "High School" 1990*
5. *Klimenko Yu.I. Higher mathematics for economists. Theory, examples and tasks-M. Exam. 2005 year*
6. *Vedenyapin G.V. General method of experimental research in the processing of experimental data. –M: Kaloe, 1973, 195s.*
7. *Lakin G.F. Biometrics - M: Higher School, 1980, 292 p.*
8. *Gmurman V.E. Probability theory and mathematical statistics - M: Higher school, 1977, 479 p .;*
9. *Gmurman V.E. A guide to solving problems in probability theory and mathematical statistics - Moscow: Vysshaya Shkola, 1979, 400s.*
10. *Vakhobov V., Khidoyatova M. "On the method of correlation analysis of experimental data". Journal "Irrigation and melioration" Tashkent №4 (14) 2018*