PROBABILISTIC APPROACH TO STATISTICAL INDICES

Anarkul URDALETOVA[2], Syrgak KYDYRALİEV[3], Kanaiym KYDYRALİEVA[4]

ABSTRACT

The purpose of this paper is to illustrate the importance of incorporating probabilistic methods into teaching modern statistics courses. Knowledge of statistical methods in today's conditions implies the ability to predict, which unfortunately is not included in most statistics courses. Students are usually taught to work with data within the framework of absolute certainty, but in real life we often have to make decisions in conditions of uncertainty. The market economy requires the skills to not only process the available statistical information, but also to understand what conclusions can be drawn from the information received, and to know how to predict future events. In this regard, in modern statistics it is impossible to do without knowledge of probabilistic methods. This paper discusses an approach to calculating the main statistical indices using the concept of probability.

Keywords: **Statistical Indices, Probabilistic Methods, Uncertainty.**

Jel Codes: **C18.**

4. 1. INTRODUCTION

Unfortunately, a statistics course is quite often perceived as a course which takes a long, and accordingly tedious, time to talk about how to classify the available datasets. Of course, this is a very important skill, but modern statistics are not limited to this. Knowledge of statistical methods in modern conditions implies the ability to predict, and this is not taught in most statistics courses. In other words, students are usually taught to work with data in conditions of complete certainty, but in real life you have to make decisions in conditions of uncertainty.

The market economy requires the skills to not only process the available statistical information, but also to understand what conclusions can be drawn from the information received, and to know how to predict future events. In this regard, in modern statistics it is impossible to do without knowledge of probabilistic methods.

This paper discusses the main statistical indices. At the same time, a close connection is traced between the traditional approach to calculating the mean and standard deviation and the method for calculating these indices using the concept of probability.

Prominent British Prime Minister Benjamin Disraeli (1804-1881) said that as a rule, the one who has the best information achieves the greatest success (Quotes, 2020).

2 Prof. Dr.,Kyrgyz Turkish Manas University, 00996312541942, anarkul.urdaletova@manas.edu.kg

3 Prof. Dr., American University of Central Asia, 00996312915000, kydyraliev_s@auca.kg

4 Asist.Prof., American University of Central Asia, 00996312915000, kanaiym.k@gmail.com

The question arises: "What does the best information mean?" There may be different answers. In particular, we can probably say that it is correctly processed information that allows us to draw correct conclusions. Of course, this is also rather vague. The aim of this work is to form a clearer picture of "better information".

5. 2. EXAMPLES

Let's start with some examples.

5.1. 2.1. Example 1

In some kingdom, some state, a letter came to parliament. This letter contained a complaint against the management of the company Alpha, which, according to the authors of the letter, gave a clear preference to men in the distribution of bonuses, discriminating against women.

A committee was sent to the firm for verification, which after a while submitted a report stating that the complaint had no grounds.

The report indicated that the firm had two divisions. Therefore, the committee split into two parts and checked both divisions. The first subcommittee checked subdivision A and found that the percentage of men with bonuses was 30% of all men, and the percentage of women with bonuses was 33.33% of all women. The second subcommittee checked subdivision B. There the percentage of men awarded was 47.14% of all men, and the percentage of women awarded was 60%.

It seems that there is no reason for further discussion on this topic. But, let's take a look at the numbers on the basis of which the report was compiled: in subdivision A 90 out of 300 men working there were awarded, and 300 out of 900 women were awarded bonuses; in subdivision B, 330 men out of 700 were awarded, and 60 out of 100 women. The figures given fully confirm the conclusions of the commission, but at the same time, if you look at the picture as a whole for the company, it turns out that 420 out of 1000 men were awarded, and only 360 out of 1000 women were awarded. So, the same dataset allows you to draw completely opposite conclusions.

Perhaps one of the best illustrations of this paradox is another saying attributed to Benjamin Disraeli: «According to Mark Twain, he liked to repeat that there are 3 types of lies: lies, blatant lies, statistics».

5.2. 2.2. Example 2

The Minister of Propaganda of the Mumba-Yumba tribe, speaking at one of the ceremonial meetings, announced that under the leadership of the sun-faced leader, the tribe had achieved an unprecedented success over the past year: the average wage increased by 100%.
The leader of the official opposition, Mr. Cook, who dared to declare that everything was not so great and that life improved only by 40%, since the price increase was 60%,

was eaten at a dinner in honor of such a significant event. How right was the leader of the official opposition?

Opposition leader correctly used Fischer's equation

$$r = i - \pi, \tag{1}$$

where r is the real rate of return, i is the nominal rate of return, π is the inflation rate. But, apparently, he studied at some Western university and, since this is not relevant for economically developed countries, he did not know that equation (1) is valid only at low inflation values. It should be noted that formula (1) is discussed in many books on macroeconomics, in particular, in Chapter 12 of the famous textbook by Gregory Mankiw [2].

Fischer's formula that works for all cases:

$$r = \frac{i - \pi}{1 + \pi}. \tag{2}$$

Accordingly, the life of the Mumba-Yumba tribe improved not by 40%, but by 25%.

To illustrate this statement, consider the following example.

2.3. Example 3

John spent 1000 coins a year buying a product at a price of 10 coins and received: 1000/10 = 100 units of goods. After his salary doubled, he decided to allocate 2000 coins to buy this product. In doing so, he discovered that the price of the product had risen to 16 coins. As a result, he can buy: 2000/16 = 125 units.

In order to derive formula (2), let us denote by I the value of nominal income (the amount of money available), by P — the price level (inflation). Then, I/P ratio is called real income. According to our notations, the amount of real income in the next period is equal to $\frac{I(1+i)}{P(1+\pi)}$, and the two values of real income are linked by equality $\frac{I}{P}(1+r) = \frac{I(1+i)}{P(1+\pi)}$. Let's reduce the equality by I/P and get the formula that connects the indices of nominal income, inflation and real income:

$$(1+r) = \frac{(1+i)}{(1+\pi)}. \tag{3}$$

In particular, from formula (3) we can express r: $r = \frac{1+i}{1+\pi} - 1$, further we use a common denominator in the right-hand side and obtain formula (2). Substituting the initial data into formula (2), we get the same result as in example 2:

$$r = \frac{1 - 0.6}{1 + 0.6} = 0.25 = 25\%.$$

Let's pay attention to the fact that at low inflation rates the results of calculations by formulas (1) and (2) practically coincide. For example, if inflation is 3% (which corresponds to inflation rates in many economically developed countries at the current time), the nominal yield is 8%, then according to formula (1) the real yield is 5%, and its true value, according to formula (2) is $\frac{0.08 - 0.03}{1 + 0.03} = 0.04854$. This number is, of course, very close to 5%, and the simplicity of formula (1) outweighs the small error. Therefore, it is worthwhile to treat the mistake made by the leader of the official opposition with understanding. As mentioned above, he studied at one of the Western universities.

6. 3. AVERAGE VALUES

Let's continue our discussion of the events that took place in the Mumba-Yumba tribe. So, the Minister of Propaganda of the Mumba-Yumba tribe, speaking at one of the ceremonial meetings, announced that under the leadership of the sun-faced leader, the tribe had achieved unprecedented success over the past year - the average salary had doubled.

To confirm these words, a payroll was presented.

The payroll shows that in the tribe there are 200 workers, 10 ministers and a president. Their wages are the following:

1st year wages: worker \$10, minister \$100, president \$587;

2nd year wages: worker \$15, minister \$200, president \$2174.

Then the arithmetic mean of wages

in the 1st year: $\frac{200 \cdot \$10 + 10 \cdot \$100 + \$587}{211} = \$17$;

in the 2nd year: $\frac{200 \cdot \$15 + 10 \cdot \$200 + \$2174}{211} = \$34$.

Thus, the mean of wages has indeed doubled. However, at the same time, the mode of the wages of workers, that is, the wages of the majority of employees, as well as the median of this set, increased from \$10 to \$15, that is, only by 50%. Therefore, remembering the inflation rate of 60%, one can hardly speak of life improvement in the tribe.

7. 4. MEASURES OF DISPERSION

For a more in-depth analysis of the situation, we use measures of dispersion (Bailey, 1998, Cozzens & Porter, 1987, Thomas, 1997, Watsham & Parramore, 1996).

7.1. Definition

If you subtract the smallest from the largest number in a set of numbers, you get a number called the range of the population.

In particular, the range of salaries in the Mumba-Yumba tribe in the first year was: \$587 - \$10 = \$577; in the second year: \$2174 - \$15 = \$2159. The calculation results show that the maximum difference in wages has increased significantly.

But this index may be determined by an insignificant part of the data. In order to avoid this situation, an index called the interquartile range is used.

7.2. Definition

If in a set of numbers we discard 25% of the smallest and 25% of the largest numbers and find the range of the resulting set, then the number found is called the interquartile range.

Since there are 211 people in the Mumba-Yumba tribe's payroll, we exclude 25% of the smallest and 25% of the largest salaries to calculate the interquartile range. It is clear that then only workers' salaries will remain: $10 in the first year and $15 in the second. That is, the interquartile range is zero. It becomes clear that this dataset contains data — we know that these are the salaries of the president and ministers — which are very different from the values of the majority of the elements of the set.

At the same time, it should be noted that in scientific research the most popular average value is the arithmetic mean, and the most popular measures of spread are the variance (dispersion) and the square root of variance — the standard deviation.

7.3. Definition

If you subtract the arithmetic mean from each number in the population, square the resulting numbers, add them up and divide by their cardinality, you get a number called the variance.

The square root of the variance is called the standard deviation.

Let calculate the variance of wages in the Mumba-Yumba tribe

in the 1st year: $\frac{200\cdot(10 - 17)^2 + 10\cdot(100 - 17)^2 + (587 - 17)^2}{211} = \frac{403590}{211} \approx 1912.75$;

in the 2nd year: $\frac{200\cdot(15 - 34)^2 + 10\cdot(200 - 34)^2 + (2174 - 34)^2}{211} = \frac{4927360}{211} \approx 23352.42$.

Correspondingly, the standard deviation

in the 1st year: (1912.75)1/2 ≈ 43.735;

in the 2nd year: (23352.42)1/2 ≈ 152.815.

8. 5. STATISTICAL INDICES WITH PROBABILITY

Next, we will demonstrate how the basic statistical coefficients are calculated using the concept of probability.

So let's turn to the payroll of the Mumba-Yumba tribe for the first year. A randomly selected person from this list has a salary of $10 with a probability of 200/211; a salary of $100 with a probability of 10/211; salary of $587 with a probability of 1/211. Let's present this data in the form of a table:

X	$10	$100	$587

P	200/211	10/211	1/211

Statisticians say that this table expresses the law of distribution of a random variable X, where X is the wage, p is the corresponding probability.

Consider the process of calculating the mean of wages in the 1st year: $\frac{200 \cdot \$10 + 10 \cdot \$100 + \$587}{211} = \$17$.

This equality can be rewritten as:

$$\$10 \cdot \frac{200}{211} + \$100 \cdot \frac{10}{211} + \$587 \cdot \frac{1}{211} = \$17 .$$

So, in order to find the mean of a random variable, you need to find the sum of the products of the values of the random variable by the corresponding probabilities. In statistics, this quantity is called the mathematical expectation of a random variable and is denoted by μ.

In the same way, using the corresponding probabilities, we can represent the process of calculating the variance:

X	\$10	\$100	\$587
p	200/211	10/211	1/211
(X - μ)2	(10 - 17)2	(100 - 17)2	(587 - 17)2

Then, the equality expressing the process of calculating the variance of wages in the 1st year: $\frac{200 \cdot (10 - 17)^2 + 10 \cdot (100 - 17)^2 + (587 - 17)^2}{211} = \frac{403590}{211} \approx 1912.75$, can be rewritten as:

$$(10 - 17)^2 \cdot \frac{200}{211} + (100 - 17)^2 \cdot \frac{10}{211} + (587 - 17)^2 \cdot \frac{1}{211} \approx 1912.75 .$$

Let's consider a couple of examples of using the outlined probabilistic approach to business problems.

8.1. 5.1. Example 4

According to American statistical mortality tables, the probability that a 25-year-old person will live next year is 0.992, and the probability that they will die within the next year is 0.008.The insurance company offers them a life insurance for a year in the amount of \$1000. The insurance premium is \$10. What profit does the company expect from each client?

Let's compose a table of distribution of a random variable — profit from one insurance:

X	\$10	\$10 - \$1000 = -\$990
p	1- 0.008 = 0.992	0.008

Therefore, the expected profit — the mathematical expectation — is equal to:

$\$10 \cdot 0.992 + (-\$990) \cdot 0.008 = \$2.$

8.2. 5.2. Example 5

Amina bakes and sells cakes. The cost of the cake is 150 soms and the price of the cake is 200 soms. Each cake unsold during the day is handed over to the nearest cafe at night for 120 soms. According to her observations, she can sell 8 to 10 cakes daily. The probability that the quantity demanded is 8 cakes is 0.2; 9 cakes — 0.5, 10 cakes — 0.3. Make the profit distribution table and determine how many cakes Amina should bake daily in order to maximize her expected daily profit?

It is clear that Amina should bake 8 or 9 or 10 cakes. At the same time, each cake sold results in 50 soms of profit, and each cake not sold results in 30 soms of losses. In order to answer the question, let's find out the value of the expected profit in all three possible cases and choose the maximum value.

a) Let 8 cakes be made. Then, they will all be sold, since there will be a demand for 8 cakes, or 9 cakes, or 10 cakes: $0.2 + 0.5 + 0.3 = 1$.

Therefore, the distribution table of the random variable — the value of the expected profit is:

X — the expected profit	$8 \cdot 50 = 400$
p	$0.2 + 0.5 + 0.3 = 1$

It is clear that in this case $\mu = 400$.

b) Let 9 cakes be made. In this case, with a probability of 0.2 eight cakes will be sold and one cake will not be sold. All of them will be sold if there is a demand for 9 or 10 cakes: $0.5 + 0.3 = 0.8$. Then the distribution table of the random variable is:

X	$8 \cdot 50 - 1 \cdot 30 = 370$	$9 \cdot 50 = 450$
p	0.2	$0.5 + 0.3 = 0.8$

In this case $\mu = 370 \cdot 0.2 + 450 \cdot 0.8 = 434$.

c) Let 10 cakes be made. In this case, with a probability of 0.2 eight cakes will be sold and two cakes will not be sold, with a probability of 0.5, nine cakes will be sold and one cake will not be sold, all ten cakes will be sold with a probability of 0.3. Therefore, the distribution table of the random variable is:

X	8·50 - 2·30 = 340	9·50 - 1·30 = 420	10·50 = 500
p	0.2	0.5	0.3

Then, $\mu = 340 \cdot 0.2 + 420 \cdot 0.5 + 500 \cdot 0.3 = 428$.

It turns out that Amina is better off baking 9 cakes daily if she wants to maximize her daily expected profit.

8.3. 6. CONCLUSION

Of course, knowing how to calculate statistical coefficients is a useful skill. However, this work can usually be entrusted to computers. Therefore, in the process of teaching students these days, the main emphasis should be on understanding the meaning of these coefficients, on the ability to use them to analyze economic situations. We hope that our work will contribute to solving this problem, and the situations considered will help make the classes of our colleagues more interesting and attractive.

9. LITERATURE

Bailey, D. (1998) Mathematics in Economics. Bercshire, England, McGraw-Hill, Inc.

Cozzens, M., Porter, R. (1987) Mathematics with Calculus USA, Heath and Company.

Mankiw, G.N. (2007) Principles of Macroeconomics. Thomson South Western

Quotes about statistics. (2020) http://www.wisdoms.one/tsitati_pro_statistiku.html

Thomas, R. (1997) Quantitative methods for business studies. New York, Prentice Hall.

Watsham, T.J., Parramore, K. (1996) Quantitative methods for finance. New Jersey, Cengage Learning EMEA.