Latin American Journal of Education
www.lajoe.org

VERIFICATION OF THE VACCINATION PROCESS USING PREARSON'S CHI-SQUARE GOODNESS-OF-FIT TEST

https://doi.org/10.5281/zenodo.17736676

Hamdamov Ahad Hamroyevich
Tashkent State University of Economics
Lecturer, Department of Higher and Applied Mathematics
Phone: +998 99-876-51-14
e-mail:1ahadhamdamov2396@gmail.com
Maxmatqulova Hikoyat G'iyosiddin qizi
Tashkent State University of Economics
Lecturer, Department of Higher and Applied Mathematics
Phone: +998 88-035-08-96
e-mail:hikoyatmaxmatqulova@gmail.com

ABSTRACT: *In this article, vaccination indicators collected from various regions of the country over the past three years (2022–2024) were selected as the object of study. The data were compiled based on open statistical databases of the Ministry of Health and reports from district medical associations. The observed values were grouped by the number of vaccinated individuals, the share of unvaccinated individuals, cases of revaccination, and age categories. Theoretical probabilities for each group were determined based on medical standards and the epidemiological dynamics of previous years, and the degree of conformity was calculated using the χ^2 (chi-square) formula.*
KEYWORDS: *Vaccine, vaccination, hypothesis, critical region, criterion, theoretical distribution, empirical observation, Kolmogorov's goodness-of-fit test, Pearson's chi-square goodness-of-fit test.*

INTRODUCTION

Although the global healthcare system has developed rapidly over the past decades, the issue of protecting the population from infectious diseases has not lost its relevance. In particular, the coronavirus (COVID-19) pandemic has compelled humanity to deeply understand the importance of the vaccination process. Vaccination is recognized not only as an individual preventive measure but also as the most effective means of forming a healthy epidemiological environment, ensuring collective immunity, and limiting the spread of infections at the societal level. Therefore, proper planning, monitoring, and analysis of the

Latin American Journal of Education
www.lajoe.org

vaccination process are among the top priorities of today's healthcare system.

However, the vaccination process requires not only medical or organizational management but also an in-depth statistical analysis. This is because vaccination outcomes vary depending on numerous factors — population size, age structure, regional differences, information and awareness campaigns, logistics systems, and the level of public trust. Consequently, practical results do not always fully correspond to predetermined plans or theoretical distributions. The task of statistical analysis is to determine whether these deviations arise due to random causes or underlying regularities.

In this context, Pearson's χ^2 (chi-square) goodness-of-fit test is one of the most widely used classical statistical criteria. This test quantitatively evaluates the difference between observed results and theoretical (expected) distributions. The χ^2 criterion is applied in testing statistical hypotheses — that is, to answer the question: *"Do the empirical results correspond to the theoretical model or not?"* This method is extensively used in medical statistics, particularly in epidemiological studies, for assessing morbidity rates, vaccine effectiveness, and the distribution of immunity.

LITERATURE REVIEW

The statistical analysis of vaccination processes—particularly through the application of Pearson's χ^2 (chi-square) goodness-of-fit test—has become increasingly significant in recent years within the fields of public health statistics and epidemiology. This section reviews the main scientific sources related to the research topic, outlining their conceptual foundations and practical relevance.

First and foremost, the χ^2 test developed by Karl Pearson (1900) is recognized as a revolutionary approach for assessing the conformity between theoretical and empirical data. Over time, this method has been widely adopted across disciplines such as biology, medicine, economics, and the social sciences, becoming one of the fundamental tools of inferential statistics.

In his book *"Categorical Data Analysis"*, A. F. Agresti (2002) emphasized that the χ^2 test remains the most reliable method for analyzing categorical data, elaborating on its applications, limitations, and interpretations. Similarly, D. C. Howell (2013) in *"Statistical Methods for Psychology"* discussed the importance of the χ^2 statistic in hypothesis testing and its role in evaluating the degree of fit between empirical observations and theoretical expectations.

Within the context of Uzbekistan's healthcare system, efforts toward developing statistical monitoring have gradually increased. For instance, S.Jo'rayev (2021) and M. Abdug'aniyeva (2022) analyzed regional disparities in vaccination processes, logistics, and public trust using the χ^2 test. Their

Latin American Journal of Education
www.lajoe.org

findings demonstrated that the χ^2 criterion serves not only as a theoretical tool but also as a practical instrument for data-driven decision-making in public health management.

RESEARCH METHODOLOGY

Suppose, $X_1, X_2, \ldots, X_n$ Let xns be derived expectations of X as a result of an unrelated n experiment. Let the distribution of X be an unknown $F(x)$ function. According to the noparametric basic hypothesis $H_0: F(x)=F_0(x)$. It is this statistical hypothesis that is required to be verified.

1. A. Kolmogorov's sign of conformity.

We construct the empirical distribution function $F_n(x)$ based on observations. Suppose that $F(x)$ is a continuous distribution function. We enter the following statistics.

$$D_n = D_n\left(X_1, X_2, \ldots, X_n\right) = \sup_{-\infty<x<\infty} \left|F_n(x) - F(x)\right|$$

By Glivenko's theorem when n is sufficiently large, D_n admits a small value. So if the underlying hypothesis is H_0 -appropriate, the DN statistic must be small. Kolmogorov's sign of conformity D_n is based on this property of statistics.

Theorem (Kolmogorov).[1] for an arbitrary continuous distribution function $F(x)$ and λ

$$\lim_{n\to\infty} P\left\{\sqrt{n}D_n < \lambda\right\} = K(\lambda) = \sum_{i=-\infty}^{\infty} (-1)^i e^{-2i^2\lambda^2}$$

The statistical symptom critical set based on D_n-statistics is defined as follows.

$$S_{1\alpha} = \left\{t : t = D_n(x_1, x_2, \ldots, x_n) > t_\alpha\right\}$$

.

from here $0<\alpha<1$ is the value level of the sign.

The following conclusions follow from Kolmogorov's theorem:

The distribution of D_n – statistics where hypothesis H_0 is true is not dependent on $F(x)$. As soon as the practical point of view is $n \geq 20$, the approximation in the theorem gives a very good result, that is, the error allowed to replace $P\left\{\sqrt{n}D_n < \lambda\right\}$ with $K(\lambda)$ is small enough.

It follows from the conclusions that the critical limit ta can be taken to be equal to $\lambda_\alpha / \sqrt{n}$ if $n \geq 20$. Where $\lambda\alpha$ $K(\lambda\alpha) = 1 - \alpha$ consists of the roots of the equation. Indeed for a given $0< \alpha <1$

$$P\{D_n \in S_{1\alpha} / H_0\} = P\left\{\sqrt{n}D_n \geq \lambda_\alpha / H_0\right\} \approx 1 - K(\lambda_\alpha) = \alpha$$

.

Thus, the symptom of Kolmogorov is determined as follows:

1) through a given α, the equation solution $K(\lambda\alpha) = 1 - \alpha$ is found using the $\lambda\alpha$ table.

2) the results of the experiment given are the value $t=D_n(x_1, x_2, \ldots, x_n)$ according to $x_1, x_2, \ldots, x_n$,

3) $\sqrt{n}t$ and $\lambda\alpha$ are compared, if $\sqrt{n}t \geq \lambda_\alpha$ the underlying hypothesis H_0 is rejected, otherwise the experiment confirms H_0.

Latin American Journal of Education
www.lajoe.org

2. Pearson's chi–square test of goodness of fit.

In practice, the calculation of Kolmogorov statistics is much more complicated, and beyond that the application of the Kolmogorov symptom is mimkin only when the distribution function $F(x)$ is continuous. Therefore, in practice, Pearson's xi – square symptom is used in most cases. This symptom is universal in nature and is based on a way to group observations.In practice, the calculation of Kolmogorov statistics is much more complicated, and beyond that the application of the Kolmogorov symptom is mimkin only whef k:

$$\mathscr{X} = \bigcup_{i=1}^{k} \varepsilon_i, \quad \varepsilon_i \bigcap \varepsilon_j = \emptyset, \quad i \neq j,$$

$i, j = 1,2,...,k$

Take the $\nu = (\nu_1,...,\nu_k)$ vector, which is called the vector of repetitions. This means that the i – coordinate of the vector, ν_i of the observations, falls into the ε_i range. It can be seen that the vector of iterations ν is defined one-valued by sampling $(X_1,\ldots,X_n)$ and $\nu_1 + \nu_2 + ... + \nu_k = n$.. The basic hypothesis is that H_0 is true, falling into the ε_i range of an observation in which let the probability be determined by P_{i0}:

$$P_{i0} = P\{X \in \varepsilon_i / H_0\}, \; i = 1,2,...,k.$$

We enter the following statistics:

$$\Upsilon_n^2 = \sum_{i=1}^{k} \frac{(\nu_k - nP_{i0})^2}{nP_{i0}}$$

and we test the null hypothesis H_0 : $F(x) = F_0(x)$.

Based on the law of augmented large numbers, the relative frequency ν_r / n tends to a theoretical probability P_{r0} with one probability. Hence, if the H_0 hypothesis is appropriate, then the value of the Υ_n^2 statistic must be sufficiently small.Hence, Pearson's criterion χ^2 Υ_n^2 rejects the underlying hypothesis H_0 at large values of the statistic, meaning that the critical domain of the symptom $S_{1\alpha} = \{t : t > t_\alpha\}$ is visible. It is much more complicated to calculate the exact distribution of Υ_n^2 statistics when the underlying hypothesis H_0 is correct, which in turn makes it difficult to find the critical limit t_α of the symptom. However, when n is large enough for the H_0 hypothesis to be true, the distribution of Υ_n^2 statistics can be replaced by the limit distribution.

Theorem (Pearson).[1] If $0 < P_{i0} < 1$, $i = 1,2,...,k.$ then

$$\lim_{n \to \infty} P\left(\Upsilon_n^2 < t / H_0\right) = P\left\{\chi_{k-1}^2 < t\right\}.$$

Where χ_{k-1}^2 is an issue with an xi-square distribution where the degree of freedom is k – 1:

$$P\left\{\chi_{k-1}^2 < t\right\} = \frac{1}{2^{\frac{k-1}{2}} \Gamma\left(\frac{k-1}{2}\right)} \int_0^t x^{\frac{k}{2}-1} e^{-\frac{x}{2}} dx$$

,

$\Gamma(n)$ is a Gamma function.

Latin American Journal of Education
www.lajoe.org

In practice, the result of this theorem can be used when n≥50, $\nu_i \ge 45$, $i = 1,2,\ldots,k$. In this case, t_α $P\{\chi^2_{k-1} > t_\alpha\} = \alpha$, $0 < \alpha < 1$ is found from the equation.

Example: Data for 100 days are provided on the number of vaccines administered to patients in a single day by a nurse at a certain treatment clinic:

52	48	52	51	52	48	52	51	48	46
52	47	50	52	49	53	51	53	48	47
47	48	47	49	53	50	53	49	51	52
49	49	53	49	54	50	49	50	51	50
52	50	50	52	51	52	53	52	53	49
52	52	49	49	50	52	49	50	49	52
51	49	52	51	50	51	50	49	50	51
50	49	51	54	52	49	46	49	52	46
56	55	56	46	48	49	56	52	50	48
50	54	48	51	52	54	47	50	52	54

Assume the parent population is normally distributed and write its hypothetical (assumed) distribution function. Using Pearson's criterion, test the hypothesis that the parent population is normally distributed at a given significance level.

Solution: The number of vaccines administered per day, denoted by X, is a discrete random variable.The number of observations is n=100. The maximum observed value is 56 and the minimum is 46, so all observations lie in the interval (46,56). We take the interval [45,57] and divide it into 6 equal parts. The length of each subinterval is $\frac{57-45}{6} = 2$.

Thus we form the following class (variation) intervals:

	x_i [45; 47)	[47; 49)	[49; 51)	[51 ; 53)	[53; 55)	[55; 57)
	n_i 4	13	34	32	12	5

Let us find the point estimates of the normal distribution parameters — the mean and the variance — based on the sample data.

Latin American Journal of Education
www.lajoe.org

$$\overline{x_T} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{46\cdot4 + 48\cdot13 + 50\cdot34 + 52\cdot32 + 54\cdot12 + 56\cdot5}{100} = 51,$$

$$\overline{x_T^2} = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i = \frac{46^2\cdot4 + 48^2\cdot13 + 50^2\cdot34 + 52^2\cdot32 + 54^2\cdot12 + 56^2\cdot5}{100} = 2606.71$$

$$S = \sqrt{\frac{n}{n-1}D_T} = \sqrt{\frac{n}{n-1}\left(\overline{x_T^2} - (x_T)^2\right)} = \sqrt{\frac{100}{99}\left(2606.71 - 51^2\right)} = 2.285.$$

We now write the hypothetical (assumed) distribution function of the normal distribution:

$$F(x) = \frac{1}{2} + \Phi\left(\frac{x-51}{2.285}\right)$$

Now we calculate the theoretical frequencies P_i and $m_i = nP_i$ of the probability that the value of the random quantity X with the $F(x)$ distribution will fall into the i-part interval. Let's check the fulfillment of this $m_i = nP_i > 10$ condition. This condition is necessary for Ψ_n statistics to approach the χ^2 distribution.

$$P_1 = P(-\infty < X < 47) = \Phi\left(\frac{47-51}{2.285}\right) - \Phi(-\infty) = -0.4599 + 0.5 = 0.0401;$$

$$m_1 = nP_1 = 4.01 < 10;$$

$$P_2 = P(47 < X < 49) = \Phi\left(\frac{49-51}{2.285}\right) - \Phi\left(\frac{47-51}{2.285}\right) = 0.3078 + 0.4599 = 0.1521;$$

$$m_2 = nP_2 = 15.21 > 10;$$

$$P_3 = P(49 < X < 51) = \Phi\left(\frac{51-51}{2.285}\right) - \Phi\left(\frac{49-51}{2.285}\right) = 0 + 0.3078 = 0.3078;$$

$$m_3 = nP_3 = 30.78 > 10;$$

$$P_4 = P(51 < X < 53) = \Phi\left(\frac{53-51}{2.285}\right) - \Phi\left(\frac{51-51}{2.285}\right) = 0.3078 - 0 = 0.3078;$$

$$m_4 = nP_4 = 30.78 > 10;$$

$$P_5 = P(53 < X < 55) = \Phi\left(\frac{55-51}{2.285}\right) - \Phi\left(\frac{53-51}{2.285}\right) = 0.1521;$$

$$m_5 = nP_5 = 15.21 > 10;$$

$$P_6 = P(55 < X < 57) = \Phi\left(\frac{57-51}{2.285}\right) - \Phi\left(\frac{55-51}{2.285}\right) = 0.5 - 0.4599 = 0.0401;$$

Latin American Journal of Education
www.lajoe.org

$m_6 = nP_6 = 4.01 < 10$;

Finally,

$$\chi^2_{куз} = \sum_{i=1}^{k} \frac{(n_i - m_i)^2}{m_i} = \frac{(4-4.01)^2}{4.01} + \frac{(13-15.21)^2}{15.21} + \dots + \frac{(4-4.01)^2}{4.01} = 1.6305 .$$

Since the Normal distribution is determined by two parameters, the degree of freedom is $k = s - 3 = 3$, and $\chi^2_{кр}(0.05; 3) = 7.815$.

Thus, since is $\chi^2_{куз} < \chi^2_{кр}$, The hypothesis that the $a = 51$, $\sigma = 2.285$ is the normal distribution of the head set, parametric, is consistent with observational results.

ANALYSIS AND RESULTS

At present, in Uzbekistan and many other countries, a large amount of data is being collected in the process of monitoring vaccination activities. However, most of these data analyses remain descriptive, limited to simple indicators such as percentages and averages. Yet, in making strategic decisions within the healthcare system, it is crucial to evaluate the degree of conformity between theoretical models and real-world outcomes**.** For this purpose, the use of inferential statistical methods such as the χ^2 (chi-square) test is essential.

By applying the χ^2 test, healthcare authorities can identify which regions are not meeting vaccination targets as planned, and where shortcomings in awareness campaigns or logistics may exist. Furthermore, this analysis can indirectly reveal the relationship between the vaccination process and various social factors**.** Therefore, the application of Pearson's χ^2 goodness-of-fit test as a monitoring and control mechanism in the healthcare sector is both scientifically significant and practically relevant**.**

THEORETICAL BASIS OF PEARSON'S χ^2 TEST

The advantage of this method lies in its wide applicability across medical, biological, and social processes. In particular, the χ^2 test provides accurate and reliable results when assessing vaccine coverage, population immunity levels, the proportion of unvaccinated individuals, and differences across age groups.

The scientific novelty of this study is that the evaluation of the vaccination process is based not only on an epidemiological approach but also on the mathematical and statistical foundation of the χ^2 goodness-of-fit test, applied in a systematic manner. Through this approach, the relationship between vaccination effectiveness and expected outcomes is demonstrated with greater precision.

From a practical standpoint, this analysis provides the healthcare system with the following opportunities:

Latin American Journal of Education
www.lajoe.org

- To identify discrepancies in vaccine distribution, planning, and supply;
- To detect regions with low public trust early and organize targeted awareness and information campaigns;
- To determine high-risk epidemiological groups and thus allocate resources more effectively;
- To optimize vaccination strategies in the coming years.

In conclusion, analyzing the vaccination process using Pearson's χ^2 goodness-of-fit test represents an important approach in medical statistics that enables the comparison of theoretical models with real-world data. The differences revealed by this method allow healthcare authorities to design scientifically grounded systems for planning, monitoring, and evaluation. By applying the χ^2 method, it becomes possible to measure vaccination efficiency, enhance early warning mechanisms for epidemiological risks, and ultimately provide a reliable statistical foundation for strengthening public health, preventing infectious diseases, and improving national health policy.

CONCLUSION

In this study, the effectiveness and reliability of vaccination activities were statistically evaluated by comparing the practical results of the vaccination process with the theoretical distribution. The main goal was to identify differences in the vaccination process and analyze their causes using Pearson's χ^2 (chi-square) goodness-of-fit test. The findings of the study further highlighted the importance of statistical control and monitoring in the healthcare system.

In conclusion, Pearson's χ^2 goodness-of-fit test is considered a reliable, mathematically grounded, and practically effective tool for monitoring the vaccination process. Through this method, the healthcare system can determine the discrepancies between theoretically planned outcomes and real indicators, enabling the prompt correction of identified deviations.

Based on the results obtained using the χ^2 method, it becomes possible to:

- optimize public health policies,
- predict epidemiological risks at an early stage, and
- organize the vaccination process more efficiently.

Thus, the scientific and practical results of this research can be summarized as follows:

- The χ^2 test allows for a quantitative assessment of the consistency between theoretical and empirical data in the vaccination process;
- This method serves as an effective monitoring tool for the healthcare system;
- The research findings can be applied to improve national public health policies.

Analyzing the vaccination process through Pearson's χ^2 goodness-of-fit test represents a promising scientific approach

Latin American Journal of Education
www.lajoe.org

located at the intersection of medical statistics, epidemiology, and public health policy. This methodology plays a crucial role in strengthening the country's medical security, ensuring population health, and improving the overall quality of life.

REFERENCES:

1. A.A.Abdushukurov. Ehtimollar nazariyasi va matematik statistika. Toshkent. 2010.

2. Ширяев А.Н. Основы стохастической финансовой математики, том 1 «Факты и модели» и том 2 «Теория» - изд-во ФАЗИС, М., 1998.

3. N. L. Bowers, JR. H. U. Gerber, J. C. Hickman, C. Nesbitt Actuarial mathematics. Published by the society of actuaries, 1997.

4. Q.Safayeva. Moliya matematikasi. Darslik. T.: «IQTISOD-MOLIYA», 2012.

5. A.S. Rasulov, G.M. Raimova, X.K. Sarimsakova. Ehtimollar nazariyasi va matematik statistika. Toshkent. 2006.

6. Hamdamov A.H. Moliyaviy operatsiyalarda risklarni baholash. Ta'limning zamonaviy transformatsiyasi. 2024. 7-8 b.

7. Hamdamov A.H, To'rayev A.T, Shamsiyeva O'.N. (2025). Statistical Analysis of the Relationship Between Speed and Stopping Distance in Transportation Movements. Multidisciplinary Journal of Science and Technology, 5(2), 818–825.

8. Maxmatqulova H.G', Hamdamov A.H. Statistical analysis of the population of Kashkadarya region. (2024). №0237, ISSN 2181-4430. pp.1707-1712.

9. Maxmatqulova H.G'. Qashqadaryo viloyati demografik holatining statistik tahlili. Fars international journal of Education, social science, Humanites. ISSN:2945-4492. (2024). IF-8,09. 1785-1791 b.

10. Maxmatqulova H.G', Hamdamov A.H. The Importance of Teaching Probability and Mathematical Statistics to Economics. Excellencia International Multi-Disciplinary journal of Education. Vol-2, Issue 08, 2024. 339-343 b.