УДК 519.28: 37: 33 **DOI 10.5281/zenodo.1485528**

Джунь И. В., д. ф.- м. н., профессор (Международный экономико-гуманитарный университет имени академика Степана Демьянчука, г. Ривнэ)

РАЗЛИЧИЕ МЕЖДУ КЛАССИЧЕСКИМ И НЕКЛАССИЧЕСКИМ МАТЕМАТИЧЕСКИМ МОДЕЛИРОВАНИЕМ И ВАЖНОСТЬ ЕГО ПОНИМАНИЯ ПРИ ПРОВЕДЕНИИ ЭКОНОМИЧЕСКИХ И ПЕДАГОГИЧЕСКИХ ИССЛЕДОВАНИЙ

Анотація*. В статті досліджено принципові відмінності класичного і некласичного методів математичного моделювання. Розкрито, що аналіз зазначених методів є необхідним для більш точного розуміння алгоритмів некласичних процедур математичної обробки вимірюваної інформації в сучасних наукових експериментах. Показано, що класичний нормальний закон похибок, часто не підтверджується при проведенні наукових досліджень через різке збільшення чисельності вибірок. Обґрунтовано, що при числі спостережень більше 500, розподіли похибок можуть бути представлені законом Пірсона-Джеффріса. Визначено, що нові уявлення про залишкові похибки моделі є найбільш важливим еволюційним проривом в сучасній теорії математичного моделювання, які складають предмет некласичної теорії похибок.*

Ключові слова: *математичне моделювання, розподіл Пірсона-Джеффріса, некласична теорія похибок.*

Аннотация. *В статье исследованы принципиальные отличия классического и неклассического методов математического моделирования. Раскрыто, что анализ указанных методов является необходимым для более точного понимания алгоритмов неклассических процедур математической обработки измерительной информации в современных научных экспериментах. Показано, что классический нормальный закон ошибок часто не подтверждается при проведении научных исследований из-за резкого увеличения объемов выборок. Обосновано, что при числе наблюдений более 500, распределения ошибки могут быть представлены законом Пирсона-Джеффриса. Определено, что новое представление об остаточных ошибках модели есть наибольшим эволюционным прорывом в современной теории математического моделирования, составляющее предмет неклассической теории ошибок.*

Ключевые слова: *математическое моделирование, распределение Пирсона-Джеффриса, неклассическая теория ошибок.*

Annotation. *The article is devoted to the objective analysis of principal differences between classical and non-classical methods of mathematical modelling. Such an analysis is necessary for a more accurate understanding of the algorithms of non-classical procedures for mathematical processing of measurement information in modern scientific experiments. Non-classical methods of data analysis are intended for mathematical processing of samples of n > 500. Exactly these volumes have multiple observations in the era of automation and computerization of scientific and technical research. C. F. Gauss created the first scientifically-based classical instrument of mathematical modelling – the least square method (LSM). In 1809 he published in his famous memoir the main fundamental condition that gives the right to use the LSM. This condition is formed as follows: the weighting function of measurement errors should be constant. In the case when the normal law is not adequate to the real error distribution of measurements, then it is necessary to apply the non-classical methods of processing of the measurement information. They are used when the weight distribution function of the errors is not constant. Unfortunately, in recent times, the normal law is more and more often not confirmed in the practice of scientific research, due to a sharp increase in the number of samples. In reality, many researchers have often seen in the past that real error distributions follow the Gauss' law in experiments. The main question is to say: when the Gauss error law is often observed? This question was answered by the famous Cambridge Professor H. Jeffreys: the hypothesis of normality is practical and, in theory, insolvent with samples' volume n > 500. It is shown that the error distributions for n > 500 are strikingly typical and can be enough and quite satisfactorily represented by the Pearson-Jeffreys law (PJVII-distribution). Mathematical models are perfect, where distribution of residual errors of model has m within 3 < m < 5, where m is the key parameter of the PJVII distribution. Exactly these ideas about residual model errors are the most important evolutionary breakthrough in the modern theory of mathematical modelling and which form the subject of the neoclassical error theory.*
Key words: *Mathematical modelling, Pearson-Jeffreys distribution, non- classical error theory.*

Успешное применение методов математического моделирования требует в наше время точного понимания принципиальных отличий классической и неклассической процедур обработки результатов научных экспериментов, поскольку непонимание этих различий, может привести к ошибкам и получению исследователями неверных результатов. Современной наукой, в частности, автором этой статьи, такое отличие научно обосновано. Поэтому, предложенные автором рекомендации касательно использования на практике неклассических процедур обработки информации при большом массиве данных являются **актуальными**, так как они **не всегда используются** в необходимых случаях.

Проблему необходимости получения достоверных данных при проведении эмпирических исследований развивали в своих публикациях всемирно известные ученые, среди которых следует выделить работы К. Ф. Гаусса, Г. Джеффриса, П. В. Новицкого, Р. Фишера.

Однако, постоянно возрастающий объем информации, сложности при ее обработке и ошибки, допускаемые учеными при проведении экономических и педагогических исследований, **определили актуальность** нашей публикации.

Целью нашего исследования есть раскрытие различий между классическим и неклассическим математическим моделированием и важности их понимания при проведении экономических и педагогических исследований.

Великий немецкий математик К. Ф. Гаусс, создавая первый, научно-обоснованный инструмент математического моделирования, известный как метод наименьших квадратов (МНК), в своем знаменитом мемуаре [1] так определил фундаментальное условие, которое дает право на применение этого метода:

$$\frac{f^{'}\left(x_i\right)}{x_i * f\left(x_i\right)} = const, \tag{1}$$

где $f\left(x_i\right)$ плотность вероятности ошибок наблюдений x_i в данном эксперименте.

Покажем, что равенство (1) возможно только тогда, когда плотность вероятности ошибок измерений $f\left(x\right)$ следует закону нормального распределения. Для этого воспользуемся общим дифференциальным представлением семейств распределений Пирсона [2]:

$$\frac{f^{'}\left(x\right)}{f\left(x\right)} = -\frac{x + c_1}{c_0 + c_1 + c_2 x^2}, \tag{2}$$

где началом отсчета является среднее, а коэффициенты:

$$c_0 = \frac{\sigma^2\left(4\beta_2 - 3\beta_1\right)}{b}; c_1 = \frac{\sigma\sqrt{\beta_1\left(\beta_2 + 3\right)}}{b}; c_2 = \frac{2\beta_2 - 3\beta_1 - 6;}{b} \tag{3}$$

$$\beta_1 = A^2 = \mu_3^2\mu_2^{-3}; \beta_2 = \mu_4\mu_2^{-2}; b = 2\left(5\beta_2 - 6\beta_1 - 9\right); \tag{4}$$

A– асимметрия распределения; эксцесс $\varepsilon = \beta_2 - 3$;

$$\mu_r = \int_{l_1}^{l_2} x^r : f(x)dx; r = 2,3,4; \mu_1 = 0; \mu_2 = \sigma^2; \tag{5}$$

l_1 и l_2 – нижняя и верхняя границы области определения плотности вероятности $f(x)$.

Для нормального распределения $\beta_1 = 0$, $\beta_2 = 3$. Подставляя эти значения в формулы (3) и (4) имеем:

$$c_0 = \sigma^2; c_1 = 0; c_2 = 0 \tag{6}$$

Значения (6) подставляем в формулу (2):

$$\frac{f'(x)}{f(x)} = -\frac{x^2}{\sigma^2} \tag{7}$$

При условии (6) формула (1) приобретает вид:

$$-\frac{f'(x)}{x * f(x)} = \left(-\frac{1}{x}\right) * \left(-\frac{x}{\sigma^2}\right) = \frac{1}{\sigma^2} = const, \tag{8}$$

где σ^2 – дисперсия нормального распределения.

Формула (2) дает также возможность убедиться в том, что главное условие МНК (1) не выполняется для любого другого, а только для нормального распределения.

Классические представления об ошибках означают, прежде всего, нормальный характер их распределения. При таких представлениях об ошибках эквивалентны следующие три важнейшие гипотезы:

1) нормальный закон есть адекватным действительному распределению ошибок наблюдений;

2) среднее арифметическое является эффективной оценкой наблюдаемой величины;

3) средняя квадратическая погрешность (СКП) есть эффективной оценкой точности измерений.

Любая из этих трёх гипотез предусматривает немедленное выполнение двух остальных.

Если же нормальный закон является неадекватным реальному распределению ошибок наблюдений, то это означает **невозможность**:

1) использования арифметического среднего в научном исследовании;

2) применения СКП, как оценки точности наблюдений или измерений;

3) использования стандарта среднего арифметического:

$$\sigma_{\bar{x}} = \sigma / \sqrt{n}, \tag{9}$$

или построения для него доверительных интервалов.

В последнее время все чаще и чаще нормальный закон не получает своего подтверждения в практике научных исследований [3–11]. Это обусловлено резким возростанием количества измерительной информации. Статистический опыт многих исследований в прошлом действительно подтверждает, что часто реальные распределения ошибок наблюдений есть приблизительно нормальными. Главный вопрос, на который мы сегодня должны ответить – когда наблюдается часто нормальное распределение ошибок? На этот вопрос дал ответ знаменитый кембриджский профессор Г. Джеффрис в работах [12–14]. В разделе 5.7 своего фундаментального труда [14], который выдержал в Великобритании уже 9 переизданий, *Джеффрис показал, что при объёмах многократных измерений $n < 500$,* закон Гаусса как правило, сохраняет адекватность. В этом случае наблюдений не так уж и много и с помощью критериальных процедур трудно доказать, что распределение ошибок значительно отклоняется от закона Гаусса. В работе [15] нами установлена нижняя граница минимального объема выборки, которая обеспечивает статистическую значимость выводов МНК, т. е. в итоге имеем:

$$30 < n < 500. \tag{10}$$

Однако при $n > 500$, как показал Джеффрис в работах [12–14], гипотеза нормальности и практически, и теоретически является несостоятельной. Правильность этого вывода подтверждена в работе [7]. Но даже в настоящее время некоторые исследователи склонны считать, что отклонениями от Гауссовой модели ошибок можно игнорировать. Они считают, что статистические процедуры, оптимальные при нормальной модели, будут приблизительно такими же и в случае отклонений от модели. К большому сожалению, как показал еще Р. Фишер [16], а потом и другие исследователи [17–20], такие надежды не имеют какого-либо основания; даже небольшие отклонения от нормальности ошибок рождают эффекты во много раз более сильные, чем это предвидели большинство математиков [20].

Оказалось, что в выборках объема n > 500 отклонения распределений ошибок от закона Гаусса имеют типичный характер.

При n > 500 распределения ошибок наблюдений удовлетворительно может быть представлено законом Пирсона-Джеффриса:

$$f(x) = c\left[1 + \frac{0,5}{M}\left(\frac{x-\lambda}{\sigma}\right)^2\right]^{-m}, \qquad (11)$$

где

$$c = \left[\sqrt{2}(m-0,5)\sigma * B\left(m+\frac{1}{2}\right), \frac{1}{2}\right]^{-1}; \qquad (12)$$

B (z, w) – бэта функция;

$M = (m-0,5)^3 - m^{-2}$; λ, σ, m – параметры распределения (11).

Фактически формула (11) есть обобщением распределений Гаусса и Стьюдента: при $m = \infty$ (11) есть распределением Гаусса, при $m < \infty$ (11) является распределением Стьюдента для дискретных значений степеней свободы $v = 2m - 1$.

Джеффрис в работе [13] пришел к выводу, что если погрешности измерений является идеально случайными (влиянием переменных систематических ошибок в результатах наблюдений можно пренебречь), то параметр m распределения (10) должен быть в таких пределах:

$$3 < m < 5, \qquad (13)$$

где m – ключевой параметр распределения Пирсона-Джеффриса.

Как видим, пределы (13) есть очень и очень далекими от $m = \infty$, которое отвечает в (11) закону Гаусса.

П. Хьюбер [19, п. 1.26] отодвигает левый предел в (13) до $m = 2$, что вполне допустимо для модели случайных ошибок с флюктуациями их дисперсии. Главное, чтобы распределение случайных ошибок не имело $m > 5$, так как такая ситуация свидетельствует об искажении эксперимента неисключенными систематическими ошибками. Таким образом, если при n > 500 остаточные погрешности математической модели имеют $m = \infty$, которое не входит в пределы (13), то это очень серьезный результат, свидетельствующий о её неадекватности.

Именно такие представления об ошибках многократных наблюдений с объемами n > 500, впервые обоснованные Джеффрисом в [13–14], есть наиболее важным эволюционным прорывом в современной теории математического моделирования. Они составляют суть современных представлений о распределении остаточных погрешностей в эру больших выборок, которой по существу и является для исследователей нынешняя эпоха.

Резюмируя изложенное выше следует отметить, что принципиальное различие классических и неклассических методов математического моделирования состоит в следующем. Основное принципиальное условие классических методов моделирования определяется формулой (1). В отличие от этого условия, главным принципиальным постулатом неклассических методов моделирования является следующий:

$$\frac{f'(x)}{xf'(x)} \neq const, \tag{14}$$

т. е. неклассические процедуры позволяют получать эффективные оценки параметров математических моделей даже при условии (14).

Значение неклассических методов математического моделирования обусловлено тем, что левая часть неравенства (14) является ничем иным как весовой функцией распределения ошибок, т. е.:

$$\frac{f'(x_i)}{x_i f(x_i)} = P(x_i), \tag{15}$$

где $P(x_i)$ есть ни что иное как обратная дисперсия погрешности x_i.

Используя в (15) выражение (11) мы получаем весовую функции распределения Пирсона-Джеффриса:

$$P(x) = \left[\left(\frac{m-0,5}{m}\right)^3 \sigma^2 + \frac{x^2}{2m}\right]^{-1}, \tag{16}$$

где σ и m – параметры распределения (11).

Весовая функция (14) позволяет преобразовать Джеффрисовы погрешности x_i *в* гауссовые $x_{Гi}$ с помощью простого оператора:

$$x_{Гi} = x_i\sqrt{P(x_i)}; \tag{17}$$

этот оператор позволяет осуществить главное исходное требования классического МНК (1), изложенное в знаменитом трактате Гаусса [1]:

$$P\left(x_{\Gamma i}\right)=\frac{f\left(x_{\Gamma i}\right)}{x_{\Gamma i} f^{'}\left(x_{\Gamma i}\right)}=const, \tag{18}$$

а значит и осуществлять моделирование привычными и хорошо известными классическими методами.

Выводы:

1. Принципиальное различие классических и неклассических процедур моделирования определяется различием формул (1) и (14).

2. Неклассические процедуры моделирования предназначены для обработки многократных наблюдений объема n > 500, при которых их ошибки, как правило следуют закону Пирсона–Джеффриса.

3. Индивидуальные веса наблюдений, которые подчиняются закону Пирсона-Джеффриса, характеризует их весовая функция (16), которая с помощью простого оператора (17) позволяет нормализировать наблюдения и после этого вести обработку данных привычными классическими методами.

4. Влиянием слабых, неисключенных зависимых систематических ошибок можно пренебречь только тогда, когда показатель m закона Пирсона-Джеффриса попадает в пределы (11).

5. *Неклассические процедуры моделирования всегда проводятся в два этапа:* на первом этапе обработка ведётся классическими методами. На втором этапе методом максимального правдоподобия находят параметры σ и m распределения (11), находят оператор (17) и заканчивают моделирование с учётом найденных весов наблюдений.

1. Gauss C. F. Theoria motus corporum coelestium in sectionibus conicis Solem ambientium. Hamburg, 1809. **2.** Большев Л. Н. Таблицы математической статистики / Л. Н. Большев, Н. В. Смирнов. – М. : Наука, 1983. – 416 с. **3.** Джунь И. В. Распределение Пирсона УП типа в ошибках наблюдений над колебаниями широт / И. В. Джунь // Астрометрия и астрофизика. – 1969. – Вып. 2. – С. 10 –115. **4.** Джунь И. В. Анализ параллельных широтных наблюдений, выполненных по общей программе: автореф. дис. на соискание уч. степени канд. физ.-мат. наук: спец. 01.03.01 «Астрометрия и небесная механика» / И. В. Джунь. – К. : Институт математики АН УССР, 1974. – 19 с. **5.** Джунь И. В. Закон распределения остаточных погрешностей определений времени и широты на астролябии Данжона / И. В. Джунь, А. А. Славинская. – Вращение и приливные деформации Земли, – К. : Наукова думка, 1984. – Вып. 16. – С. 69–74. **6.** Джунь И. В. Особенность закона распределения результатов баллистических измерений ускорения силы тяжести / И. В. Джунь, Г. П. Арнаутов, Ю. Ф. Стусь, С. Н. Щеглов // Повторные гравиметрические наблюдения. – Изд. МГК при Президиуме АН СССР и НПО «Нефтегеофизика». – М., 1984. – С. 87–100. **7.** Джунь И. В. Математическая обработка астрономической и космической информации при негауссовых ошибках наблюдений: автореферат дис. на соиск. уч. степени докт. физ.-мат. наук: спец. 01.03.01 «Астрометрия и небесная механика» / И. В. Джунь. – Киев, ГАО НАН Украины, 1992. – 46 с. **8.** Харин А.С. изучение ошибок наблюдений

Голосеевского каталога звезд широтных программ / А. С. Харин, Я. С. Яцкив // Астрометрия и астрофизика. – 1970. – Вып. 10. – С. 34–43. **9.** Новицкий П. В. Оценка погрешностей результатов измерений. / П. В. Новицкий, И. А. Зограф. – Л. : Энергоатомиздат, 1991. – 304 с. **10.** Орлов А. И. часто ли распределение результатов наблюдений является нормальным? / А. И. Орлов. - Заводская лаборатория. – 1991, № 7. – 64–66. **11.** Gazda V. About the Distribution of random Oscilations of Stock Index RMS – 100 / V. Gazda, I. V. Dzhun // Economika Firiem. Kosice: 9 – 10.09.1999. – 1999. **12.** Jeffreys H. The Law of Errors and the Combinations of Observations / H. Jeffreys // Philos. Trans. Roy. Soc. London, 1937, ser. A., No 237, pp. 231–271. **13.** Jeffreys H. The Law of Errors in the Greenwich Variation of Latitude observations Mon. Not. of the RAS. – 1937, vol. 99, № 9, pp. 703–709. **14.** Jeffreys H. Theory of Probability. Sec. Edition. – Oxford, 1940. – 468 p. **15.** Джунь И. В. Неклассическая теория погрешностей измерений / И. В. Джунь. – Ровно : Естеро, 2015. – 168 с. **16.** Fischer R. A. On the mathematical Foundations of theoretical Statistics. Philos. Trans. Roy. Soc. London, ser. A, 222, 1921, pp. 309–368. **17.** Tukey J. W. Survey of Sampling from Contaminaded Distributions. Paper 39. In «Contributions to Probability and Statist», (ed. Olkin I. et ai.) Stanford Univ. Press, 1960, pp. 448–485. **18.** Tukey J. W. The future of Data Analysis / Ann. Math. Stat., 1962, vol. 33, № 1, pp. 1–67. **19.** Tukey J. W. Data Analysis and the Frontieres of Geophysics. Science, 1965, vol. 148, Numb. 3675, pp. 1283–1289. **20.** Hampel F. R., Rousseeuw P. J., Stahel W. A. Robust statistics. The Approach Based on Influence. John Wiley & Sons. New York, 1986.

Рецензент: д.е.н., професор Борейко В. І.