

ИНФОРМАЦИОННЫЙ КРИТЕРИЙ ДИСКРИМИНАЦИИ ЭМПИРИЧЕСКИХ РАСПРЕДЕЛЕНИЙ

Юдин С.В., д-р техн. наук, профессор, профессор Тульского филиала РЭУ им. Г.В. Плеханова, г. Тула
Этова Е.В., аспирант Тульского филиала РЭУ им. Г.В. Плеханова, г. Тула

Рассмотрена задача дискриминации эмпирических распределений, возникающих при анализе экономических и социологических процессов. Предложено использовать информационно-статистический критерий, основанный на статистике «энтропия». На основании анализа литературы сделан вывод об универсальности этой статистики. Описаны основные понятия математической теории информации, на основе которой были получены расчетные формулы нового критерия и алгоритм процедуры проведения расчетов. На основании данных, полученных при исследовании уровня компетентности сотрудников систем менеджмента качества 46 предприятий, из которых 34 – предприятия АО «Концерн ВКО «Алмаз-Антей» (Концерн), а 16 – предприятия Тульской области, построены рейтинги специалистов. Распределение рейтингов показало неудовлетворительную квалификацию сотрудников. Проведен сравнительный анализ рейтингов по трем эмпирическим распределениям, относящимся к данным 2019 и 2024 года по дочерним предприятиям Концерна и данным 2024 года по предприятиям Тульской области. Сравнение было проведено при помощи критерия Колмогорова-Смирнова и разработанного информационного критерия. Показаны идентичность результатов и большее удобство информационного критерия. Отмечена необходимость организации курсов повышения квалификации и профессиональной переподготовки специалистов систем менеджмента качества предприятий.

Ключевые слова: информационный критерий дискриминации, энтропия, система менеджмента качества, квалификация сотрудников, повышение квалификации.

Цитирование: Юдин С.В., Этова Е.В. Информационный критерий дискриминации эмпирических распределений // Информационно-экономические аспекты стандартизации и технического регулирования. 2025. № 5 (86). С. 124–129.

ВВЕДЕНИЕ

В экономике, социологии, технических науках часто используется дискриминантный анализ, который можно определить как раздел многомерного статистического анализа, позволяющий разделить генеральную совокупность на две или более группы по некоторым признакам (критериям) или оценить значимость изменения характеристик групп, проверить однородность выборок. В частности, авторы занимаются анализом систем менеджмента качества предприятий Тульской области и динамикой изменения соответствующих показателей. В работе [1], например, были проведены сравнения эмпирических распределений количества несоответствий по пунктам стандарта ГОСТ Р ИСО 9001–2015, полученных по результатам инспекции двух предприятий (машиностроительного и радиоэлектронного профилей соответственно) АО «Концерн ВКО «Алмаз-Антей» (Концерн), для оценки деятельности систем менеджмента качества (СМК). Было выявлено множество несоответствий, для которых были построены эмпирические распределения, упорядоченные по пунктам стандарта. Срав-

нение распределений при помощи критерия Колмогорова-Смирнова [2] показало, что отличие их друг от друга незначимо. Это дает основание утверждать, что выявленные проблемы носят фундаментальный характер.

Это же утверждается в интервью руководителя Росстандарта А.П. Шалаева и главного редактора журнала «Стандарты и качество» Г.П. Воронина [3].

Учитывая важность затрагиваемых вопросов, необходим простой математический аппарат, позволяющий оперативно рассчитать и оценить значимость изменения распределения. В этой связи авторы поставили задачу разработки соответствующего критерия, пригодного для использования специалистами, не имеющих специального математического образования.

Существует большое количество критериев, применяемых при решении задач дискриминации, однако они имеют ряд недостатков, которые снижают их эффективность при наличии нелинейных зависимостей, наличии распределе-

ний, отличных от нормального, и других, что показано в монографии [4]. Там же предложено использовать такую статистику, как энтропия распределения, имеющую минимум ограничений.

Цель исследования: 1) разработать информационный критерий дискриминации эмпирических распределений; 2) провести исследования рейтинга сотрудников СМК Тульских предприятий; 3) провести сравнение распределений рейтинга сотрудников СМК Концерна и предприятий Тульской области при помощи информационного критерия.

КРИТЕРИЙ ДИСКРИМИНАЦИИ

С. Кульбак (Solomon Kullback) в своей фундаментальной монографии «Information Theory and Statistics» [5], переведенной и опубликованной в СССР в 1967 году, показал универсальность математического аппарата теории информации для решения задач математической статистики, проверки статистических оценок и других.

Впервые понятие информации в том смысле, который вкладывается в него в настоящее время, было определено Р.А. Фишером (R.A. Fisher) в его работе 1925 года “Theory of statistical estimation” [6].

Дальнейшее развитие теории информации связано с именами таких ученых, как Клод Шеннон [7], Норберт Винер [8], Натаниэль Мартин и Джеймс Ингланд [9], Лиам Панински [10], Роберт Грей [11], Амос Голан [12], и других.

Монография В.Г. Григоровича и др. [4] показала возможность использования методов теории информации для решения вопросов управления качеством продукции в машиностроении и электронной промышленности.

Рассмотрим основные характеристики математической энтропии.

Энтропия распределения равна математическому ожиданию информации по всем состояниям, т.е.

$$H = -\sum_{i=1}^n p_i \ln p_i. \quad (1)$$

Пусть для множества событий вероятности вычислены на основе наблюдений по формуле

$$\hat{p}_i = \frac{k_i}{k}, i = 1 \dots n, \quad (2)$$

где k_i – количество наблюдений события E_i , k – общее количество наблюдений ($k = \sum_{i=1}^n k_i$).

Тогда эмпирическая энтропия будет равна

$$\hat{H} = -\sum_{i=1}^n \hat{p}_i \ln \hat{p}_i. \quad (3)$$

Эмпирическая энтропия имеет нормальное распределение (распределение Гаусса) со следующими характеристиками [4, 13]:

$$M\hat{H} = H; D\hat{H} = \frac{a^2 - H^2}{k}. \quad (4)$$

Здесь $M\hat{H}$ – математическое ожидание (Expected Value (i.e. Mean)) эмпирической энтропии; $D\hat{H}$ – ее дисперсия (Variance).

Параметр a^2 определяется по формуле

$$a^2 = \sum_{i=1}^n p_i \ln^2 p_i. \quad (4)$$

На основе вышеизложенного С.В. Юдин [4, 13] разработал и широко применял при исследовании случайных процессов в машиностроении, электронной промышленности и эконометрике информационный критерий идентификации вида закона распределения.

Он показал, что величина

$$J_c = \frac{h - \hat{H}}{\sqrt{\frac{a^2 - h^2}{k}}}, \quad (6)$$

где

$$\begin{cases} h = -\int_{-\infty}^{\infty} w(x) \ln w(x) dx \\ a^2 = \int_{-\infty}^{\infty} w(x) \ln^2 w(x) dx \end{cases}. \quad (7)$$

Здесь $w(x)$ – нормированная функция плотности вероятностей данного класса распределений.

В нашем случае мы имеем дело не с задачей идентификации закона распределения, а со сравнением эмпирических распределений, возникающих при опросах для определения имеющихся тенденций.

В этой связи нам необходимо использовать не теоретические, предполагаемые параметры распределений, а эмпирические энтропии.

На базе описанного критерия J_c можно составить критерий наличия тенденций

$$T_Y = \frac{\hat{H}_1 - \hat{H}_2}{\sqrt{\frac{\hat{a}_1^2 - \hat{H}_1^2}{k_1} + \frac{\hat{a}_2^2 - \hat{H}_2^2}{k_2}}} \quad (8)$$

Здесь \hat{H}_1, \hat{H}_2 – эмпирические энтропии, вычисленные по формуле (3) на первом и втором этапах исследования соответственно;

$$\hat{a}_m^2 = \sum_{i=1}^{k_m} \hat{p}_{mi} \ln^2 \hat{p}_{mi} \quad (9)$$

В (12) приняты следующие обозначения: m – номер опроса ($m=1...2$); k_m – общее количество интервалов; \hat{p}_{mi} – оценка вероятности наблюдения i -го варианта в наборе данных с номером m .

Данная величина имеет t -распределение Стьюдента с числом степеней свободы $r = n_1 + n_2 - 2$, где n_1 и n_2 – количество интервалов разбиения интервалов значений выборок.

Функция плотности вероятностей распределения Стьюдента определяется следующим образом:

$$f_k(x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{\pi k} \Gamma(\frac{k}{2})} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}} \quad (10)$$

Гамма-функция $\Gamma(z)$ определяется через несобственный интеграл

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \quad (11)$$

Пусть задана доверительная вероятность α , тогда критическое значение критерия T_Y можно определить как решение уравнения

$$T_Y(r, \alpha): \int_{-T_Y(r, \alpha)}^{T_Y(r, \alpha)} f_r(x) dx = \alpha \quad (12)$$

Решение уравнения (12) можно доверить функции СТЬЮДЕНТ.ОБР.2Х (Т.INV.2Т) в MS Excel®.

Если расчетное значение критерия T_Y (11) превосходит табличное (12), то гипотеза о неоднородности подтверждается, в противном случае отвергается.

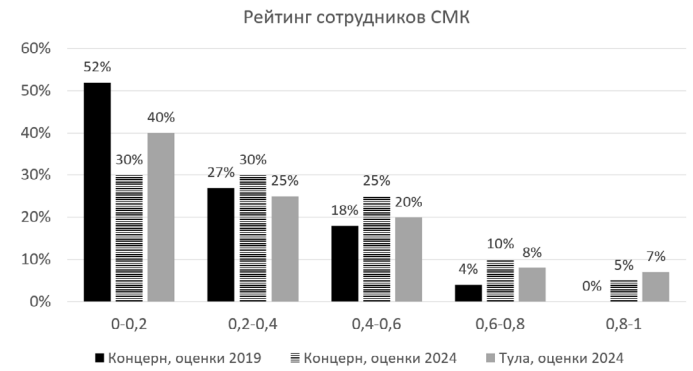
Адекватность, точность и надежность этого критерия обуславливается тем же для информационного критерия J_c , подробно исследованного в работах [4, 13], где было доказано, что он точнее стандартных критериев типа критерия Пирсона χ^2 , требует меньшего объема выборки, имеет более узкие доверительные границы.

АНАЛИЗ КАДРОВОГО СОСТАВА СИСТЕМ МЕНЕДЖМЕНТА КАЧЕСТВА

В 2017–2019 гг. в Тульском филиале РЭУ им. Г.В. Плеханова была выполнена НИР «Разработка механизма и методологии инспекционного контроля деятельности ДО Концерна в области обеспечения качества и надежности продукции и эффективности системы управления качеством Концерна», заключенная с АО «Концерн ВКО «Алмаз-Антей» [14]. Одной из задач, поставленных заказчиком, была проверка знаний и умений сотрудников СМК. Система оценок была согласована с Департаментом управления качеством Концерна. Оценка «удовлетворительно» ставится при рейтинге 0,4–0,6; «хорошо» – 0,6–0,8; «отлично» – 0,8–1,0.

В октябре 2024 г. Департамент управления качеством того же концерна, предоставил данные за 2024 год. В обоих случаях количество сотрудников составило 200 человек, количество предприятий – 34.

В декабре 2024 г. было проведено исследование рейтинга сотрудников СМК предприятий Тульской области. Анкетирование проводили студенты 4 курса направления «Менеджмент» Тульского филиала РЭУ им. Г.В. Плеханова в рамках выполнения курсовой работы по предмету «Система



Рейтинг сотрудников СМК на предприятиях АО «Концерн ВКО «Алмаз-Антей» и предприятиях Тульской области (0 – сотрудник ничего не знает и не умеет; 1 – сотрудник все знает и умеет) (составлен авторами)

менеджмента качества». Общее количество анкет – 56, количество предприятий – 12.

Информация о распределении доли анкетированных с указанным значением рейтинга приведена на рисунке.

Как правило, сравнение двух эмпирических распределений осуществляется на основе критерия Колмогорова-Смирнова λ [2]. Значение этого критерия вычисляется по следующей формуле:

$$\lambda_{ЭМП} = d_{max} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \quad (13)$$

Здесь $\lambda_{эмп}$ – значение критерия Колмогорова-Смирнова; d_{max} – максимальное значение разности вероятностей (частот) по всем интервалам; n_1, n_2 – объемы исследуемых выборок.

Расчеты проводим для уровня значимости $\alpha = 0,05$. Критическое значение статистики Колмогорова-Смирнова $\lambda_{кр} = 1,36$. Исходные данные приведены в табл. 1 (сравнение распределений по данным, полученным в Концерне за 2019 и 2024 годы).

Максимальное значение разности долей $d_{max} = 0,22$. Рассчитанное значение критерия Колмогорова-Смирнова $\lambda = 2,2$. Эта величина превышает критическое значение $\lambda_{кр} = 1,36$. Таким образом, возможно сделать вывод о значимом отличии двух распределений.

Таблица 1

Сравнение распределений рейтинга сотрудников Концерна в 2019 г. и 2024 г.

Диапазон оценок рейтинга	Доля сотрудников с указанным рейтингом		Разность долей
	Концерн, 2019 г.	Концерн, 2024 г.	
0–0,2	0,52	0,30	0,220
0,2–0,4	0,27	0,30	0,030
0,4–0,6	0,18	0,25	0,070
0,6–0,8	0,04	0,10	0,060
0,8–1	0,00	0,05	0,050

В табл. 2 приведены сравнения по данным, полученным в Концерне в 2024 г., и данным тульских предприятий.

Таблица 2

Сравнение распределений рейтинга сотрудников Концерна и тульских предприятий в 2024 г.

Диапазон оценок рейтинга	Доля сотрудников с указанным рейтингом		Разность долей
	Концерн, 2024 г.	Тула, 2024 г.	
0–0,2	0,30	0,40	0,100
0,2–0,4	0,30	0,25	0,050
0,4–0,6	0,25	0,20	0,050
0,6–0,8	0,10	0,08	0,020
0,8–1	0,05	0,07	0,020

Максимальное значение разности долей $d_{max} = 0,10$. Рассчитанное значение критерия Колмогорова-Смирнова $\lambda = 0,66$. Эта величина меньше, чем критическое значение $\lambda_{кр} = 1,36$. Таким образом, возможно сделать вывод о незначимом различии двух распределений.

Рассмотрим решение этой же задачи при помощи информационного критерия T_y . Расчеты были проведены по формулам (6), (8), (11). Результаты сведены в табл. 3.

Таблица 3

Расчеты энтропийных параметров

	Концерн, оценки 2019	Концерн, оценки 2024	Тула, оценки 2024
Энтропия H	1,13	1,30	1,24
Параметр a^2	1,63	1,88	1,84
Объем выборки n	200,00	200,00	56,00

Значение критерия T_y для первых двух распределений равен $T_y = -5,91$.

То же для второго и третьего распределений: $T_y = 0,31$.

Критическое значение двустороннего распределения Стьюдента для уровня значимости $\alpha = 0,05$ и числа степеней свободы $k=8$ равно $t_{крит} = 2,31$.

Таким образом, оба метода дают идентичные результаты, т.е. отличие эмпирических распределений данных по Концерну 2019 и 2024 годов является значимым, в то время как отличие эмпирических распределений по данным 2024 года Концерна и Тулы является статистически незначимым.

Значимость изменения эмпирического распределения для Концерна можно объяснить кампанией повышения квалификации и профессиональной переподготовки сотрудников СМК предприятий. Тем не менее, общий уровень квалификации остался неудовлетворительным.

ЗАКЛЮЧЕНИЕ

На основе проведенных исследований можно сделать вывод о том, что информационно-статистический подход, основанный на статистике «энтропия», является универсальным. На его основе был предложен информационный критерий дискриминации T_y .

Проверка критерия T_y была проведена на реальных данных, полученных при исследовании предприятий АО «Концерн ВКО «Алмаз-Антей» и Тульской области. На основе этих данных можно сделать вывод о том, что реальная квалификация сотрудников СМК является недопустимо низкой.

Проведено сравнение результатов, полученных при помощи критерия Колмогорова-Смирнова λ и информационного критерия ТУ. Показано, что они дали идентичные результаты, что дает возможность использовать критерий T_u для решения задач дискриминации, при этом информационный критерий, имеющий распределение Стюдента, более удобен для использования.

Можно отметить, что кампания повышения квалификации и профессиональной переподготовки сотрудников СМК предприятий Концерна дала некоторый результат. Есть ос-

нование полагать, что положение со временем улучшится, поскольку эти курсы возобновлены в учебном центре Концерна АНО ДПО «НОЦ ВКО «Алмаз – Антей».

**Рецензент: Будкин Юрий Валерьевич, доктор технических наук, доцент, советник генерального директора, директор научного центра ФГБУ «Институт стандартизации», г. Москва, Российская Федерация.
E-mail: y.v.budkin@gostinfo.ru**

Список использованных источников и литературы

1. Юдин С.В., Кузнецов Г.В. Оценка компетентности сотрудников систем менеджмента качества и программы повышения квалификации // Информационно-экономические аспекты стандартизации и технического регулирования. 2024. № 6 (81). С. 176–180.
2. Смирнов Н.В. Оценка расхождения между эмпирическими кривыми распределения в двух независимых выборках // Бюллетень МГУ. 1939 Т. 2, № 2. С. 3–14.
3. Шалаев А.П., Воронин Г.П., Киселева Т.В. Стандарты – для качества жизни и устойчивого роста экономики // Стандарты и качество. 2025. № 3. С. 12–21.
4. Информационные методы в управлении качеством / В.Г. Григорович, С.В. Юдин, Н.О. Козлова, В.В. Шильдин; под общ. ред. В.Г. Григоровича. – М.: РИА «Стандарты и качество», 2001. – 208 с.
5. Кульбак С. Теория информации и статистика. – М.: Наука, 1967. – 408 с.
6. Фишер Р.А. Статистические методы для исследователей. – М.: Госстатиздат, 1958. – 267 с.
7. Шеннон К. Э. Математическая теория связи // Работы по теории информации и кибернетике / Пер. С. Карпова. – М.: ИИЛ, 1963. – 830 с.
8. Винер Н. Кибернетика, или управление и связь в животном и машине. 1948–1961. – 2-е издание. – М.: Наука; Главная редакция изданий для зарубежных стран, 1983. – 344 с.
9. Математическая теория энтропии / Н. Мартин, Дж. Ингленд; Пер. с англ. В. А. Каймановича; Под ред. [и с предисл.] А.М. Вершика. – Москва : Мир, 1988. – 350 с.
10. Paninski L., Estimation of entropy and mutual information // Neural computation. 2003. Т.15, № 6. С.1191–1253.
11. Robert M. Gray, Entropy and Information Theory. Springer, New York. 2011. DOI: 10.1007/978-1-4419-7970-4
12. Golan A. Information and Entropy Econometrics – A Review and Synthesis, Foundations and Trends® in Econometrics. 2008. Vol 2, no. 1–2, pp. 1–145.
13. Yudin S. The informational criterion of identification of the distribution law // Modern European Researches. 2015. № 4. С. 133–137.
14. Разработка (актуализация) и верификация нормативных и методических документов инспекционной деятельности; Выработка рекомендаций по повышению эффективности системы управления качеством продукции и надежностью технологических систем (применительно к уровням управления качеством в интегрированной системе): Разработка механизма и методологии инспекционного контроля деятельности ДО Концерна в области обеспечения качества и надежности продукции и эффективности системы управления качеством Концерна : отчет о НИР (окончательный) / Тульский филиал РЭУ им. Г.В. Плеханова; рук. С.В. Юдин. – Тула: 2019. – 285 с. – № ГР АААА-А17-117110970026-5).

INFORMATION CRITERION FOR DISCRIMINATION OF EMPIRICAL DISTRIBUTIONS

Yudin S.V., Doctor of Technical Sciences, Professor, Professor of the Tula Branch of Plekhanov Russian University of Economics, Tula

Etova E.V., postgraduate student at the Tula Branch of Plekhanov Russian University of Economics, Tula

The study addresses the challenge of discriminating between empirical distributions encountered in the analysis of economic and sociological processes. An information-statistical criterion based on entropy statistics is proposed for this purpose. An analysis of existing literature confirms the universality of these statistics. The paper outlines the fundamental concepts of mathematical information theory, which serve as the basis for the derivation of computational formulas for the new criterion and the development of a corresponding calculation algorithm. The method is applied to assess the competency levels of quality management system (QMS) personnel across 46 enterprises: 34 enterprises belonging to JSC Concern VKO Almaz-Antey and 16 enterprises from the Tula region. Rating distributions revealed insufficient qualification levels among employees. A comparative analysis was conducted using three empirical datasets: 2019 data for subsidiaries of the Concern, 2024 data for the same subsidiaries, 2024 data for Tula region enterprises. The analysis employed the Kolmogorov-Smirnov test and the newly developed information criterion. Both criteria produced consistent results but the information criterion demonstrated greater practicality. The study highlights the need for: advanced training programs and professional retraining courses for QMS specialists.

Keywords: information criterion of discrimination, entropy, quality management system, employee qualifications, professional development.

For citation: Yudin S.V., Etova E.V. Information criterion for discrimination of empirical distributions. Information and Economic aspects of standardization and technical regulation. 2025. 5 (86): 124–129. (In Russ.).

References

1. Yudin S.V., Kuznecov G.V. Ocenka kompetentnosti sotrudnikov sistem menedzhmenta kachestva i programmy` povy`sheniya kvalifikacii. Informacionno-e`konomicheskie aspekty` standartizacii i texnicheskogo regulirovaniya. 2024. No. 6 (81), pp. 176–180.
2. Smirnov N.V. Ocenka rasxozhdeniya mezhdru e`mpiricheskimi krivy`mi raspredeleniya v dvux nezavisimy`x vy`borkax. Byulleten` MGU, 1939. Vol. 2. No. 2, pp. 3–14.
3. Shalayev A.P., Voronin G.P., Kiseleva T.V. Standarty – dlya kachestva zhizni i ustoychivogo rosta ekonomiki. Standarty i kachestvo. 2025. No. 3, pp. 12–21.
4. Informacionny`e metody` v upravlenii kachestvom: V.G. Grigorovich, S.V. Yudin, N.O. Kozlova, V.V. Shildin. Moscow: RIA “Standarty` i kachestvo”, 2001, 208 p.
5. Kullback S. Information Theory and Statistics. Glouchester, Mass. Peter Smith, 1978, 399 p.
6. Fisher R.A. Statisticheskie metody` dlya issledovatelej. Moscow: Gosstatizdat, 1958, 267 p.
7. Shannon K. E. Matematicheskaya teoriya svyazi. Raboty` po teorii informacii i kibernetike. – M.: IIL Publ., 1963, 830 p.
8. Viner N. Kibernetika, ili upravlenie i svyaz` v zhivotnom i mashine. 1948–1961: 2-nd ed. Moscow: Nauka; Glavnaya redakciya izdanij dlya zarubezhny`x stran, 1983, 344 p.
9. Martin N., England J.W. Mathematical Theory of Entropy (Encyclopedia of Mathematics and its Applications Book 12). Cambridge University Press, 1984, 286 p.
10. Paninski L., Estimation of entropy and mutual information. Neural computation. 2003. Vol.15. No. 6, pp. 1191–1253.
11. Robert M. Gray, Entropy and Information Theory. Springer, New York, 2011. DOI: 10.1007/978-1-4419-7970-4
12. Golan A. Information and Entropy Econometrics – A Review and Synthesis, Foundations and Trends® in Econometrics, 2008. Vol 2. No. 1–2, pp. 1–145.
13. Yudin S. The informational criterion of identification of the distribution law. Modern European Researches, 2015. No. 4, pp. 133–137.
14. Razrabotka (aktualizaciya) i verifikaciya normativny`x i metodicheskix dokumentov inspekcionnoj deyatel`nosti; Vy`rabotka rekomendacij po povy`sheniyu e`ffektivnosti sistemy` upravleniya kachestvom produkcii i nadezhnost`yu texnologicheskix sistem (primenitel`no k urovnjam upravleniya kachestvom v integrirovannoj sisteme) : Razrabotka mexanizma i metodologii inspekcionnogo kontrolya deyatel`nosti DO Koncerna v oblasti obespecheniya kachestva i nadezhnosti produkcii i e`ffektivnosti sistemy` upravleniya kachestvom Koncerna : otchet o NIR (okonchatel`ny`j): Tul`skij filial RE`U im. G.V. Plexanova; ruk. S.V. Yudin. Tula, 2019, 285 p. – No. GR AAAA-A17-117110970026-5).