

УДК 519.226

<https://doi.org/10.26160/2541-9579-2019-6-3-6>

ПРОВЕРКА ВХОДНЫХ ПАРАМЕТРОВ ЭКСПЕРТНОЙ СИСТЕМЫ НА СООТВЕТСТВИЕ НОРМАЛЬНОМУ ЗАКОНУ РАСПРЕДЕЛЕНИЯ

Серобабова А.С.

Омский государственный технический университет, г.Омск

Ключевые слова: проверка на соответствие нормальному распределению, функция распределения, тест Д'Агостино и Пирсона, Scipy.

Аннотация. В статье рассмотрен подход проверки экспериментальных данных на их соответствие нормальному распределению.

VERIFICATION THE INPUT PARAMETERS OF THE EXPERT SYSTEM FOR COMPLIANCE WITH THE NORMAL DISTRIBUTION

Serobabova A.S.

Omsk State Technical University, Omsk

Keywords: checking for compliance with normal distribution, function of distribution, test D'agostino and Pearson, Scipy.

Abstract. The article considers the approach of testing experimental data for their compliance with the normal distribution.

При проведении статистического анализа медицинских данных, полученных в результате лабораторных исследований и значений физиологических данных пациентов, возникает необходимость проверки гипотезы H_0 о том, что исследуемые количественные признаки соответствуют нормальному закону распределения. Принадлежность наблюдаемых данных нормальному закону является необходимой предпосылкой, чтобы применить рассчитанные специальные методы математической статистики, рассчитанные на данное распределение. Под нормальным распределением понимается симметричное распределение колоколообразной формы, при котором около 68% данных отличается от среднего арифметического не более чем на одно и примерно 95% – не более чем на два стандартных отклонения в каждую сторону [1].

В статье выполняется проверка данных на соответствие нормальному закону распределения ключевых данных из проводящегося исследования в рамках нахождения взаимосвязи лабораторных, физических данных пациента на развития заболевания неалкогольной жировой болезни печени, что является продолжением работы [2]. Необходимо подтвердить или опровергнуть гипотезу о виде закона распределения. В общем виде задача имеет вид $H_0: F(x)=F(x)$, где $F(x)$ – функция распределения вероятностей, с которой проверяется согласие наблюдаемой выборки. Для решения этой задачи используются несколько различных критериев согласия: критерий Колмогорова-Смирнова [3], χ^2 [4], тест Д'Агостино и Пирсона [5]. В качестве основного критерия выбран тест Д'Агостино и Пирсона.

Введем следующие обозначения: D_{NASH} – наличие стеатогепатита, C_p – обхват талии пациента (см), P_l – размеры печени (мм), P_{Gep} – наличие гепатомегалии, D_{AG} – наличие заболевания артериальная гипертензия, $D_{Эл}$ – степень заболевания неалкогольной жировой болезни печени, L_{lep} – содержание лептина в организме пациента, L_{ObR} – количество рецепторов воспринимающих лептин, L_{T2} – значение тканевой ингибитор матриксных протеиназ 2, L_{MMP9} – значение лабораторного параметра матриксной металлопротеиназы 9, P_{DT} – время динамики заболевания печени, M_M – принимаемый препарат пациентом метформин, L_{GGT} – значение лабораторного параметра гамма-глутамилтранспентидазы, P_{dis} – параметр пациента дискомфорт, D_O – наличие ожирения у пациента, D_{OC} – наличие остеоартроза у пациента, P_{din} – значение динамики массы пациента.

Алгоритм проверки на соответствие выборки нормальному распределению изображен на рисунке 1 и выполнен в программной среде Google Colab. Для вычисления значения асимметрии используется библиотека языка Python Scipy и функция skewtest() удовлетворяющая вычислениям формулы:

$$A = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{\sigma^3}, \quad (1)$$

где x_i – i значение исследуемой выборки; σ – среднеквадратическое отклонение; $\bar{x}$ – среднее значение параметра; n – количество элементов параметра.

Для вычисления значения эксцесса вызывается функция kurtosistest(), удовлетворяющая вычислениям формулы

$$E = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{\sigma^4}. \quad (2)$$

Для вычисления значения критерия необходимо воспользоваться формулой

$$Z = A^2 + E^2. \quad (3)$$

На рисунке 2 изображен фрагмент результатов формируемых программой.

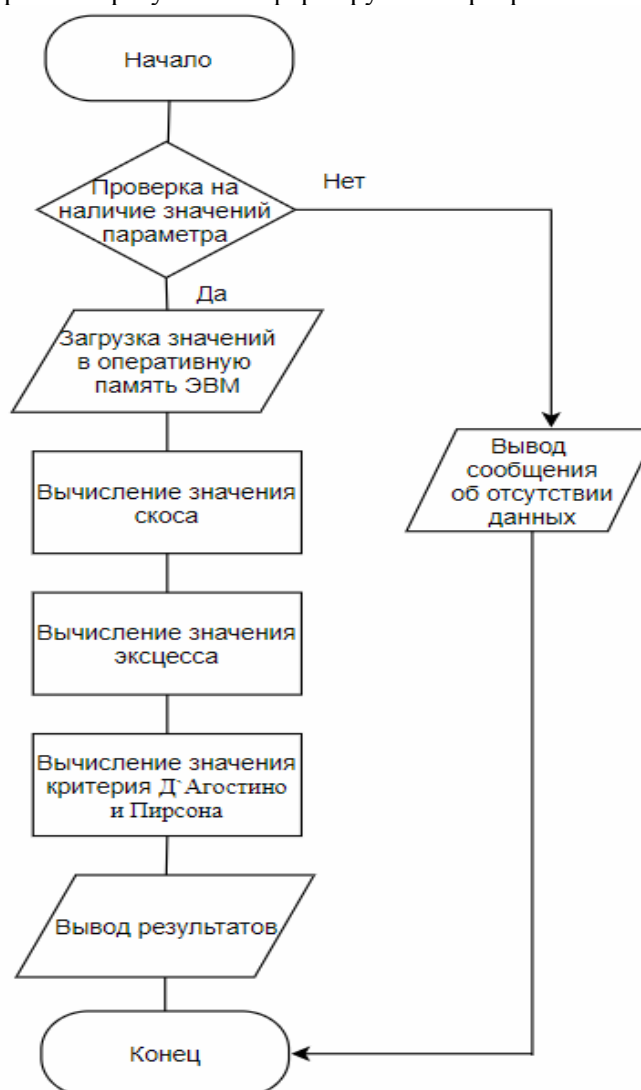


Рис. 1. Алгоритм проверки на соответствие выборки нормальному закону распределению

остеоартроз					
Pirson	p-value	kurtosis	kurtosis p-value	skew	skew p-value
68.87976852110451	1.103947215161818e-15	3.68853223655274	0.00022555145239072947	7.43468213584251	1.0481970937721453e-1
метформин					
Pirson	p-value	kurtosis	kurtosis p-value	skew	skew p-value
76.31079900824962	2.6873271544814788e-17	-7.751521143979509	9.079822287884092e-15	4.027991926840002	5.625527486928966e

Рис. 2. Фрагмент результатов вычисления значений эксцесса, асимметрии и теста Д'Агостино и Пирсона

Все полученные результаты представлены в таблице 1 с вычисленным значением p-value [6], которое позволяет сделать вывод о том принять или отвергнуть гипотезу H_0 .

На основе данных таблицы 1 можно сделать следующие выводы: исследуемые параметры C_p , D_{NASH} , P_I удовлетворяют условию $p\text{-value} > 0,05$; расхождение значения асимметрии у параметров: C_p ,

D_{AG} , P_{dis} , $P_{G\phi}$, D , D_{NAS} , P_{di} , P , D_{OC} удовлетворяют условию $A < |3|$, параметры L_{T2} , L_{lep} , C_P , P_{DT} , D_{NASH} , P_I , D_{OC} также удовлетворяют условию $E < |5|$ [7]. На рисунке 3 изображены две гистограммы, которые изображают распределение исследуемых величин C_P и P_I , которые соответствуют всем 3 условиям, а также графики гистограмм соответствуют плотности распределения нормальному распределению.

Табл. 1. Значения критерия Пирсона, p-value, эксцесса, скоса исследуемых физических и лабораторных параметров

№	Оцениваемый физический или лабораторный параметр	Значение критерия	Значение p-value	Расхождение значения эксцесса параметра и нормального распределения	Расхождение значения асимметрии параметра и нормального распределения
1	$D_{Эл}$	8,88	0,012	минус 1,972	2,23
2	L_{lep}	59,642	0	4,46	6,31
3	C_P	4,504	0,105	1,06	1,839
4	L_{ObR}	68,307	0	5,29	6,35
5	L_{T2}	22,321	0	2,256	4,151
6	L_{MMP9}	68,62	0	5,322	6,348
7	P_{DT}	22,91	0	3,057	0,33
8	M_M	76,31	0	минус 7,752	4,02
9	D_{AG}	8,609	0,014	0	минус 2,934
10	D_{NASH}	0,122	0,941	0	минус 0,11
11	P_{dis}	45,6	0	2,668	6,204
12	$P_{G\phi}$	1884,73	0	43,308	минус 2,954
13	D_O	205,962	0	8,071	минус 11,867
14	D_{OC}	68,89	0	3,689	7,435
15	P_I	1,744	0,418	0	1,32
16	P_{din}	14,44	0	2,216	3,685

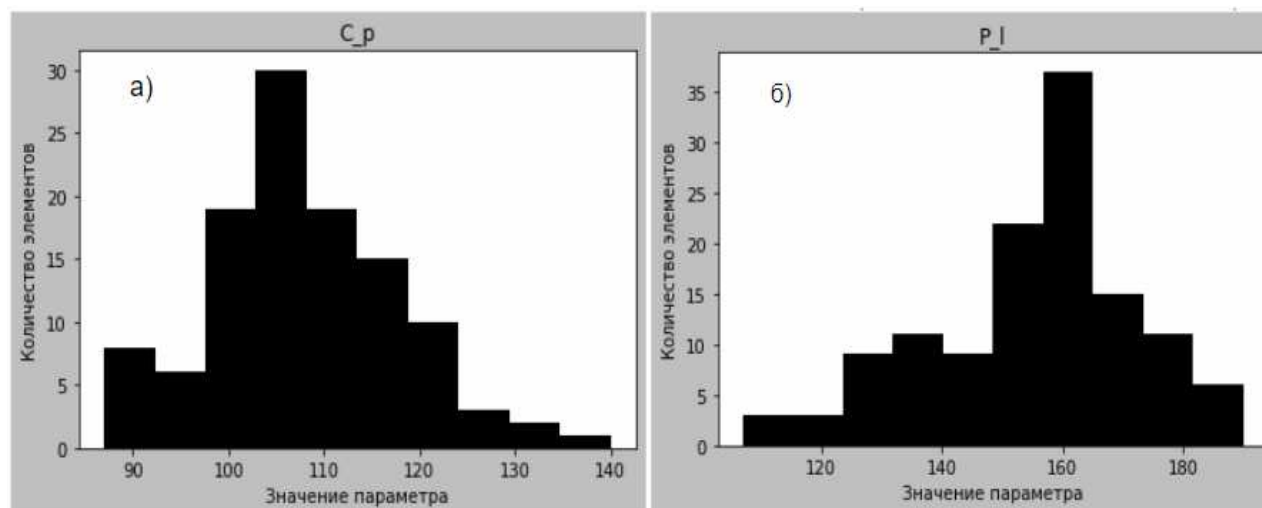


Рис. 3. Гистограмма распределения исследуемых величин: а) C_P – обхват талии пациента в см; б) P_I – размеры печени мм

По результатам проведенного анализа входной выборки данных выявлены 3 параметра: C_P , P_I , D_{NASH} , которые удовлетворяют допустимым расхождениям эксцесса, асимметрии, а так же $p\text{-value} > 0,05$. Таким образом, получены параметры, позволяющие в дальнейшем исследовании применять в отношении себя математический аппарат, ориентированный на наличие выборки соответствующе нормальному распределению.

Список литературы

- Гржибовский А.М. Типы данных, проверка распределения и описательная статистика // Экология человека. – 2008. – №1. – С. 52-58.
- Серобабова А.С., Чебаненко Е.В., Денисова Л.А., Кролевец Т.С. Разработка экспертной системы ранней диагностики заболеваний: программные средства первичной обработки и выявления зависимостей. // Омский научный вестник. – 2018. – № 4 (160). – С. 179-184. – DOI: 10.25206/1813-8225-2018-160-179- 184.

3. Lilliefors H.W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown // J. Am. Statist. Assoc. – 1967. – Vol. 62. – P. 399-402.
4. Джеймс Г. и др. Введение в статистическое обучение с примерами на языке R / Пер. с англ. С.Э. Мостицкого. – М.: ДМК Пресс. – 2016.
5. D'Agostino R.B. An omnibus test of normality for moderate and large sample size // Biometrika. – 1971. – №58. P. 341-348.
6. Бослаф С. Статистика для всех / Пер. с англ. П.А. Волкова, И.М. Флямер, М.В. Либерман, А.А. Галицына. – М.: ДМК Пресс, 2017. – 586 с.
7. Процедура Экспресс-оценки [Электронный ресурс] <https://domstor.ru/userfiles/express.pdf>

Сведения об авторах:

Серобабов Александр Сергеевич – аспирант по направлению «Информатика и вычислительная техника», ОмГТУ, г.Омск.

УДК 621.3

<https://doi.org/10.26160/2541-9579-2019-6-6-7>

СТРУКТУРА ПРОГРАММЫ КОМПЛЕКСНЫХ ИСПЫТАНИЙ ЭЛЕКТРОТЕХНИЧЕСКИХ СИСТЕМ КОСМИЧЕСКИХ АППАРАТОВ

Брусков А.А.

Государственное бюджетное образовательное учреждение высшего образования Московской области «Технологический университет», г.Королев

Ключевые слова: спутник, испытания, космический аппарат, комплексные электрические испытания, структура программы испытаний.

Аннотация. Автором описана оптимальная структура программы комплексных испытаний электротехнических систем космических аппаратов.

STRUCTURE OF THE PROGRAM OF COMPLEX TESTS OF ELECTRICAL SYSTEMS OF SPACECRAFT

Bruskov A.A.

State budgetary educational institution of higher education of Moscow region "Technological University", Korolev

Keywords: satellite, tests, spacecraft, complex electrical tests, structure of the test program.

Abstract. The author describes the optimal structure of the program of complex tests of electrical systems of spacecraft.

Введение. В процессе производства космических аппаратов проводится большое количество испытаний, начиная от самых простейших – контрольных и заканчивая одними из самых сложных – комплексных.

В результате планирования испытаний необходимо определить структуру программы комплексных электрических испытаний космических аппаратов [1-4].

Планирование структуры программы испытаний позволяет оптимизировать и повысить эффективность комплексных испытаний космических аппаратов.

Структура программы комплексных испытаний электротехнических систем космических аппаратов. Определение объема и содержания программы комплексных испытаний электротехнических систем космических аппаратов необходимо проводить с целью оптимизации всего процесса испытаний, т.е. обеспечить отработку изделия наилучшим качеством при минимальных затратах средств и времени.

В результате планирования объема и содержания программы испытаний необходимо оптимально разбить общую программу испытаний на частные программы, определить цели и объем частных испытаний, установить критерии их законченности и готовности изделия перехода к высшим уровням испытаний. Эта задача является трудной, так как процесс испытаний электротехнических систем космических аппаратов является сложным техническим процессом.

Сложность систем космического аппарата является одной из главных причин возникновения отказов и нештатных ситуаций при комплексных испытаниях электротехнических систем космических аппаратов, программного обеспечения. В общем случае, сложность изделия определяется сложностью взаимодействия его компонентов.

Перечисленные особенности процесса комплексных испытаний электротехнических систем космических аппаратов являются определяющими для сложных технических систем и предопределяют использование для комплексного планирования процесса испытаний иерархического метода.

Иерархический метод позволяет разбить систему на элементарные уровни. Концепция уровней позволяет анализировать систему, скрывая несущественные для данного уровня детали реализации