

АВТОМАТИЗИРОВАННАЯ СИСТЕМА РАСЧЕТА КРИТЕРИЯ ЭППСА-ПАЛИ

Журавский А.А.¹⁾, Соснова Е.Б.¹⁾, Кутовая О.В.²⁾, Синяева О.В.¹⁾

¹⁾ГП «УХИН», ХНСХУ им. П. Василенка

²⁾Кафедра технологій фармацевтичних препаратів

Національний фармацевтичний університет, м. Харків

tfp@nuph.edu.ua

Принадлежность наблюдаемых данных нормальному закону является необходимой предпосылкой для корректного применения большинства классических методов математической статистики, используемых в задачах обработки измерений, стандартизации и контроля качества. Поэтому проверка на отклонение от нормального закона является частой процедурой в ходе проведения измерений, контроля и испытаний.

Стандарт ГОСТ Р ИСО 5479-2002 “Статистические методы. Проверка отклонения распределения вероятностей от нормального распределения” [1], введенный в действие в 2002 г., представляет собой аутентичный текст международного стандарта ISO 5479-97. В стандарте рассматривается графический метод проверки на нормальность с использованием вероятностной оценки, критерии проверки на симметричность и на значение эксцесса, статистики которых представляют собой функции от оценок моментов закона распределения, критерии Шапиро-Уилка, основанные на регрессионном анализе порядковых статистик, критерий Эппса-Палли, статистика которого измеряет некоторое расстояние между выборочной характеристической функцией и характеристической функцией нормального закона.

Автоматизация процесса определения позволит сократить время проведения расчёта до уровня, необходимого для введения исходных данных и исключить влияние т.н. «человеческого фактора». Решению этих задач и была посвящена разработанная программа.

Настоящая программа устанавливает критерии, с помощью которых можно проверить, подчиняется ли генеральная совокупность данных нормальному закону распределения. Данный критерий [2-5] базируется на сравнении эмпирической и теоретической характеристических функций. В стандарте предусмотрено его применение при $8 \leq n \leq 200$.

Работу с программой начинают с заполнения таблицы исходных данных (рисунок 1).

1|Критерий Эппса - Палли

Настоящая программа устанавливает критерии, с помощью которых можно проверить, подчиняется ли генеральная совокупность данных нормальному закону распределения.

Этот критерий применим при количестве переменных больше 8. Многосторонний критерий с высокой мощностью при многих альтернативных гипотезах использует сумму квадратов модулей разности между характеристическими функциями на основе выборочных данных и нормального распределения с весовыми коэффициентами.

Таблица 1.1 - Исходные данные для расчёта критерия Эппса - Палли

X1	X2	X3	X4	X5	X6	X7	X8
3	4,9	6	8	8,9	10,9	12	12,9
X9	X10	X11	X12	X13	X14	X15	X16
13,9	24	0	0	0	0	0	0
X17	X18	X19	X20	X21	X22	X23	X24
0	0	0	0	0	0	0	0
X25							
0							

По 10 наблюдениям (таблица 1.1) вычисляются следующие параметры:

Статистику критерия Эппса - Палли определяют по формуле:

Таблица ввода исходных данных

X1	X2	X3	X4	X5	X6	X7	X8
4,9	6	24	8	8,9	12	10,9	3

Введите в ячейку, выделенную голубым цветом, значение K

10

Введите в ячейку, выделенную зеленым цветом, значение квантиля статистики критерия Эппса - Палли "P", выбранного из ряда 0,90; 0,95; 0,975 и 0,99: P = 0,95

Таблица 1.1a - Исходные данные и расчёт основных моментов

X1	X2	X3	X4	X5	X6	X7	X8
3	4,9	6	8	8,9	10,9	12	12,9
1	1	1	1	1	1	1	1
55,503	30,803	19,803	6,0025	2,4025	0,2025	2,4025	6,0025
0,647	0,7853	0,8561	0,954	0,9813	0,9984	0,9813	0,954
	0,9449	0,8683	0,6755	0,5791	0,3755	0,2805	0,2148
		0,9812	0,86	0,778	0,5684	0,4534	0,3663
			0,9392	0,8764	0,6861	0,5684	0,4737
				0,9874	0,8764	0,778	0,6861
					0,9392	0,86	0,778
						0,9812	0,9392
							0,9874

Рисунок 1. Таблица ввода исходных данных

Данная таблица находится на полях справа и не имеет нумерации. При этом абсолютно безразлично, в какой последовательности заполняется таблица исходных данных, поскольку компьютер все данные выстраивает по ранжиру и переносят эти данные в таблицу 1.1, находящуюся в начале расчёта. Следует подчеркнуть, что первая цифра, находящаяся в заголовке таблиц, обозначает порядковый номер раздела, который может меняться. Дело в том, что при разработке программы авторы предполагали, что результаты расчётов могут войти на правах отдельного раздела в некий общий расчёт. Поэтому в программе предусмотрено возможность автоматической перенумеровки не только глав, но

нумерацию рисунков и таблиц. Далее требуется ввести в ячейку, выделенную голубым цветом (рисунок 2) значение коэффициента K , который может меняться в пределах $2 \leq K \leq n$. Также следует ввести в ячейку, выделенную зеленым цветом квантили статистики p .

Введите в ячейку, выделенную голубым цветом, значение K

10

Введите в ячейку, выделенную зеленым цветом, значение квантиля статистики критерия Эппас - Палли " P ", выбранного из ряда 0,90; 0,95; 0,975 и 0,99: $P =$ 0,99

Рисунок 2. Ввод значений коэффициента K и квантилей статистики p

В случае ввода ошибочных значений, на экране монитора появляется соответствующее сообщение, а все остальные расчёты закрашиваются черной заливкой (рисунок 3).

Внимание! Значение коэффициента K не может меньше 2!

Введите в ячейку, выделенную голубым цветом, значение K

1

Введите в ячейку, выделенную зеленым цветом, значение квантиля статистики критерия Эппас - Палли " P ", выбранного из ряда 0,90; 0,95; 0,975 и 0,99: $P =$ 0,99

Рисунок 3. Сообщение о том, что значение коэффициента K выбрано неправильно

Кроме того, программа воспроизводит пошагово все расчёты на случай, если потребуется аргументированный вывод, соответствует ли данная выборка требованиям нормальности распределения.

Данная программа выполнена при помощи электронных таблиц Microsoft Excel, что позволяет вносить исправления с целью улучшения программы. В некоторых случаях приходилось вносить собственные оригинальные решения, например, перенос данных из таблицы исходных данных в таблицу 1.1 в строгой последовательности по возрастанию. Конечно, можно было бы воспользоваться стандартной функцией сортировки, но это может вызвать проведения дополнительных операций, что затруднит работу оператора.

Далее программа самостоятельно делает вывод о том, соответствует ли исследуемый массив нормальному распределению. Поэтому при выполнении программы мы использовали девиз: «Вы вводите только исходные данные, все остальное выполняем мы!».

При разработке программы была предусмотрена защита от случайного её повреждения. Все ячейки, где происходят расчёты, защищены. При попытке несанкционированного доступа к вычислительным ячейкам, работа программы приостанавливается, а экран выводится соответствующее сообщение.

Список использованных литературных источников:

1. Стандарт ГОСТ Р ИСО 5479-2002. Статистические методы. Проверка отклонения распределения вероятностей от нормального распределения. – М.: Издательство стандартов, 2002. – 30 с.
2. Baringhaus L., Danschke R., Henze N. Recent and classical tests for normality the empirical characteristic function // *Metrika*. 35, 1988. – P.339-348.
3. Baringhaus L., Henze N. A consistent test for multivariate normality based on the empirical characteristic function // *Metrika*. 35, 1988. – P.339-348.
4. Epps T.W., Pulley L.B. A test for normality based on the empirical characteristic function // *Biometrika*. 70, 1983. – P. 723-726.
5. Henze N. An approximation to the limit distribution of the Epps-Pulley test statistic for normality // *Metrika*. 37, 1990. – P.7-18.