Сахибназарова Виктория Бахтиёровна
магистрант
ФГАОУ ВО «Самарский государственный
аэрокосмический университет
им. академика С.П. Королёва (НИУ)»
г. Самара, Самарская область

МОДЕЛИРОВАНИЕ СЛУЧАЙНОЙ ВЕЛИЧИНЫ И ПРОВЕРКА ГИПОТЕЗЫ О ВИДЕ ЕЕ РАСПРЕДЕЛЕНИЯ

Аннотация: *в статье с помощью встроенной функции Excel моделируется случайная величина, имеющая нормальный закон распределения с параметрами, а затем проверяется гипотеза о виде ее распределения.*

Ключевые слова: *случайная величина, нормальный закон распределения.*

Чтобы построить гистограмму распределения необходимо по сгенерированной выборке $x_1, x_2 ... x_n$ построить вариационный ряд $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$ [1]. Далее весь промежуток $\lfloor x_{(1)}, x_{(n)} \rfloor$ разбивается точками $x_{(1)} = u_0 < u_1 < ... < u_N = x_{(n)}$ на $N = [1 + 3.32 \lg n] + 1$ непересекающихся интервалов $J_k = [u_{k-1}, u_k)$, $k = \overline{1, N}$. Высоты столбцов гистограммы ищутся по формуле $h_k = \frac{\nu_k}{n \cdot \Delta u_k}$, $k = \overline{1, N}$, где $\Delta u_k = u_k - u_{k-1}$, ν_k- частоты попадания выборочных значений в k-ый интервал J_k, и $\tilde{p}_k^* = \frac{\nu_k}{n}$, $k = \overline{1, N}$ – относительные частоты. Для оценки правильности построенной гистограммы, используется теоретическое значение функции нормального распределения с заданными параметрами в середине интервала, которое находится по формуле:

$$f(a, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}}, \text{ где } a = 3.9,\ \sigma^2 = 4.3$$

Таблица 1

Таблица данных для гистограммы

Интервалы J_k		Частоты ν_k	Относительные частоты $\tilde{p}_k^*$	Высота столбца гистограммы h_k	Теоретическое значение в середине интервала $\tilde{h}_k$
Левая граница	Правая граница				
− 1,55316435	− 0,42164369	14	0,028	0,02474546	0,01196498
− 0,42164369	0,709876977	19	0,038	0,03358313	0,03730771
0,709876977	1,841397642	41	0,082	0,07246885	0,08637193
1,841397642	2,972918306	86	0,172	0,15200783	0,14846843
2,972918306	4,10443897	102	0,204	0,18028836	0,18948856
4,10443897	5,235959634	109	0,218	0,19266109	0,179564
5,235959634	6,367480299	73	0,146	0,1290299	0,1263406
6,367480299	7,499000963	34	0,068	0,06009612	0,06600153
7,499000963	8,630521627	15	0,03	0,02651299	0,02560074
8,630521627	9,762042291	7	0,014	0,01237273	0,0073729
Весь промежуток		500	1		

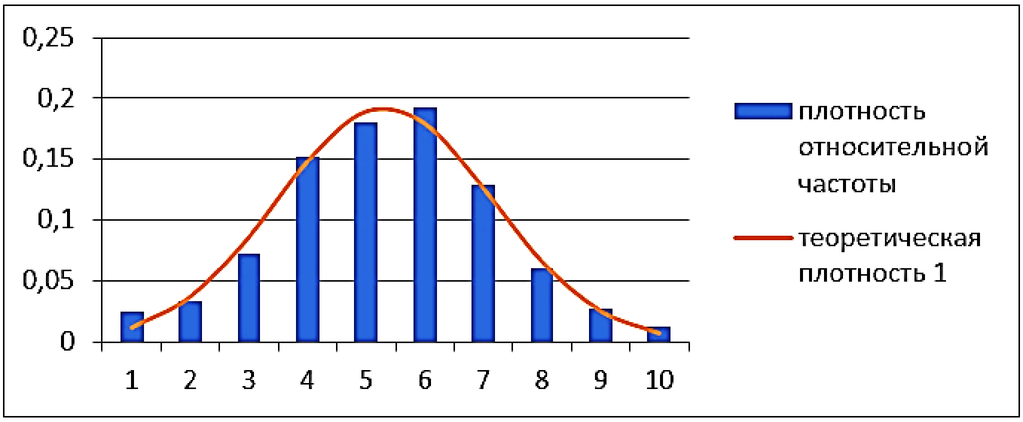

Рис. 1. Гистограмма распределения

Для проверки гипотезы о нормальном распределении заданной функции с помощью критерия χ^2 Пирсона, необходимо сгруппировать и представить выборочные данные $(x_1, x_2, ..., x_n)$ в виде интервального статистического ряда, где $J_k = [u_{k-1}, u_k)$, $k = \overline{1, N}$ – интервалы группировки, $\nu_1, \nu_2, ..., \nu_N$ $(\nu_1 + \nu_2 + ... + \nu_N = 1)$- частоты попадания выборочных значений в интервалы $J_1, J_2, ..., J_N$.

Теоретическая вероятность $\bar{p}_k$ попадания случайной величины X в J_k при неизвестных параметрах распределения находится по формуле

$\overline{p}_k = \Phi(\overline{z}_k) - \Phi(\overline{z}_{k-1}),\ k = \overline{1, N}$, где $\overline{z}_k = \dfrac{u_k - \overline{x}}{\overline{\sigma}}$, $\overline{z}_{k-1} = \dfrac{u_{k-1} - \overline{x}}{\overline{\sigma}}$; $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$ – функция Лапласа, значение которой находится по таблице; среднее выборочное по сгруппированным данным $\tilde{x} = 3{,}926$; дисперсия по сгруппированным данным $\tilde{\sigma}^2 = 4{,}375$.

Теоретическая вероятность p_k попадания случайной величины X в J_k при известных параметрах распределения находится по формуле $p_k = \Phi(z_k) - \Phi(z_{k-1}),\ k = \overline{1, N}$, где $z_k = \dfrac{u_k - a}{\sigma}$, $\overline{z}_{k-1} = \dfrac{u_{k-1} - a}{\sigma}$; $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$ – функция Лапласа, значение которой находится по таблице; математическое ожидание $a = 3{,}9$; дисперсия $\sigma^2 = 4{,}3$.

Таблица 2

Интервальный статистический ряд

Интервалы J_k		Частоты ν_k	Относительные частоты $\tilde{p}_k^*$	Теоретическая вероятность, найденная при условии, что а и сигма неизвестны $\overline{p_k}$	Теоретическая вероятность p_k при условии, что а и сигма известны
Левая граница	Правая граница				
−1,55316435	−0,42164369	14	0,028	0,014289	0,014304
−0,42164369	0,709876977	19	0,038	0,043024	0,043397
0,709876977	1,841397642	41	0,082	0,097224	0,098443
1,841397642	2,972918306	86	0,172	0,164923	0,166992
2,972918306	4,10443897	102	0,204	0,210028	0,211859
4,10443897	5,235959634	109	0,218	0,200809	0,201028
5,235959634	6,367480299	73	0,146	0,144144	0,142666
6,367480299	7,499000963	34	0,068	0,077675	0,07572
7,499000963	8,630521627	15	0,03	0,031419	0,030051
8,630521627	9,762042291	7	0,014	0,009538	0,008917
Весь промежуток		500	1	0,993073	0,993378

Объем выборки $n = 500 > 50$, частоты попадания значений в выбранные интервалы $\nu_k \geq 5,\ k = \overline{1, N}$, следовательно, статистика критерия χ^2 находится по формуле: $\overline{\chi}_n^2 = \sum_{k=1}^{N} \dfrac{(\nu_k - n \cdot \overline{p}_k)^2}{n \cdot \overline{p}_k}$, $\chi_n^2 = \sum_{k=1}^{N} \dfrac{(\nu_k - n \cdot p_k)^2}{n \cdot p_k}$

При неизвестных значениях параметров распределения $\overline{\chi}_n^2 = 10{,}72759$. Для заданного уровня значимости $\alpha = 0{,}025$ порог $\chi_{1-\alpha;\,N-1}^2 = 16{,}013$, где $r = 2$ – число неизвестных параметров. При известных значениях параметров распределения $\chi_n^2 = 11{,}08465$. Для заданного уровня значимости $\alpha = 0{,}025$ порог $\chi_{1-\alpha;\,N-1-2}^2 = 19{,}023$. Так как $\overline{\chi}_n^2 < \overline{\chi}_{1-\alpha;\,N-1-r}^2$ и $\chi_n^2 < \chi_{1-\alpha;\,N-1}^2$, то гипотеза о нормальном распределении заданной функции принимается.

Список литературы

1. Коломиец Э.И. Моделирование и статистический анализ случайных данных. – Самара: СГАУ, 2011. – 72 с.