Шашков В.Б.
Оренбургский государственный университет

АВТОМАТИЗИРОВАННЫЙ РАСЧЕТ КРИТЕРИЯ ПИРСОНА (математическая статистика без статистических таблиц)

Разработана новая методика расчета критерия согласия Пирсона, которая позволяет легко программировать эти расчеты. Построена таблица для значений интегральной функции распределения как для отрицательных, так и для положительных значений стандарта случайной величины. Таблица интерпретирована полиномом регрессии, который введен в программу расчета критерия Пирсона.

При проверке гипотезы о нормальном распределении массива значений данной случайной величины используется критерий согласия Пирсона в виде уравнения

$$\chi^2_{\text{раб}} = \sum_{i=1}^{к} \frac{(m_i - p_i \times W)^2}{p_i \times W}, \qquad (1)$$

где к – количество групп, на которые разбит массив с шагом Δx:

m_i – экспериментальное количество значений в i-й группе;

p_i- теоретическая вероятность попадания значений в i-ю группу;

W – объем массива;

$p_i \times W$ – теоретическое количество значений в i-й группе, рассчитанное для нормального распределения.

Методика расчета критерия приведена в ряде учебных пособий [1, 2].

Эти методики не свободны от ряда недостатков.

Во-первых, это необходимость использования статистических таблиц (функции Лапласа, критических точек χ^2-распределения и пр). Это рутинная и ручная работа, которая препятствует компьютеризации расчетов.

Во-вторых, при использовании стандартной функции

$$U_i = \frac{x_i - Mx}{\sigma}$$

приходится использовать таблицы только для положительных значений этой функции. В силу этого при расчете интегральной функции распределения приходится определять не функцию F(±u), а функцию F(|u|), что вызывает определенные трудности. Поясним ситуацию на конкретном примере.

Имеем массив значений размеров обуви, зафиксированных случайным образом у мужчин старше двадцати лет. Объем массива W равен 500. Согласно формуле $k = 1+3{,}2 \times \lg W$ (где «к» есть количество групп) массив разбит на 10 групп с шагом $\Delta x = 2$. Данные по расчету функции F(|u|) по существующей методике и по расчету функции F(±u) по предлагаемой методике представлены в таблице 1.

По F(±u) и F(|u|) рассчитывают теоретическую вероятность попадания значений случайной величины в i-ю группу.

По существующим методикам [1] схема расчетов теоретической вероятности P_i по F(|u|) на протяжении столбца 4 для значений F(|u|) меняется три раза. Для первой строки P_i вычисляется как разность $P_1 = 1 - 0{,}9924 = 0{,}0076$.

Для строк №№2-5 (область отрицательных значений U) P_i рассчитывают по разности

$$F(|u|)_i - F(|u|)_{i+1},$$

а для строк №№5-10 (область положительных значений U) P_i рассчитывают, наоборот, по разности

$$F(|u|)_{i+1} - F(|u|)_i.$$

Все это затрудняет как восприятие учебного материала студентами, так и программирование расчетов. Предлагаемая в данной работе методика расчета критерия согласия Пирсона включает три оригинальных момента.

Во-первых, составлена таблица для функции F(±u) для значений U от -4 до +4, т. е. и для отрицательных и для положительных значений U. Эти результаты приведены в таблице 2. Насколько нам известно, такие данные в печати отсутствуют.

Таблица 1. Расчет интегральной функции для i-й группы

№ пп	Границы групп	U_i для правой границы	F(|u|) по табл.	F(±u) по урав.	P_i по F(|u|)	P_i по F(±u)
1	31-33	-2,43	0,9924	0,0074	0,0076	0,0074
2	33-35	-1,83	0,9664	0,0340	0,0260	0,0266
3	35-37	-1,22	0,8888	0,1116	0,0776	0,0776
4	37-39	-0,61	0,7291	0,2713	0.1597	0,1597
5	30-41	0	0,5000	0,5000	0,2291	0,2287
6	41-43	0,61	0,7291	0,7283	0.2291	0,2283
7	43-45	1,22	0,8888	0,8881	0,1597	0,1599
8	45-47	1,83	0.9664	0,9658	0,0776	0,0777
9	47-49	2,43	0,9924	0,9926	0,0260	0,0266
10	49-51	3,04	0,9988	0,9988	0,0064	0,0062

Таблица 2. Значения интегральной функции нормированного нормального распределения

U	0	2	4	6	8
1	2	3	4	5	6
-4,0	**0,0000**831				
-3,9	3453	3380	1081	2667	1921
-3,8	2499	2773	3028	3244	3396
-3,7	1343	1514	1723	1962	2225
-3,6	1126	1082	1081	1124	1212
-3,5	1929	1699	1501	1338	1212
-3,4	3466	3116	2784	2475	2189
-3,3	5454	5029	4617	4218	3834
-3,2	7286	7780	6807	6343	5892
-3,1	**0,000**10548	9947	9373	8822	8292
-3,0	14053	13274	12537	11839	11178
-2,9	18742	17685	16691	15758	14880
-2,8	25165	23713	22349	21070	19869
-2,7	33943	31967	30107	28357	26711
-2,6	45757	43117	40621	38264	36040
-2,5	61347	57889	54607	51496	48548
-2,4	81529	77083	72849	68820	64988
-2,3	107224	**0,0**101596	96221	91091	86196
-2,2	139478	132447	125716	119275	113114
-2,1	179483	170799	162468	154477	146817
-2,0	228597	217975	207765	197955	188531
-1,9	288334	275459	263061	251127	239644
-1,8	360349	344880	329958	315569	301699
-1,7	446410	427987	410184	392987	376380
-1,6	548343	526598	505547	485176	465468
-1,5	667971	642544	617882	593970	570795
-1,4	807045	777595	748975	721173	694176
-1,3	967159	933384	900496	868484	837337
-1,2	1149673	1111323	1073905	1037411	**0,**1001832
-1,1	1355626	1312523	1270381	1229195	1188961
-1,0	1585671	1537718	1490738	1444731	1399694
-0,9	1840005	1787202	1735365	1684495	1634597
-0,8	2118326	2060776	2004161	1948490	1893770
-0,7	2419798	2357710	2296507	2236202	2176805
-0,6	2743039	2676733	2611241	2546577	2482758
-0,5	3086132	3076035	2946657	2878019	2810140
-0,4	3446655	3373289	3300529	3228402	3156929
-0,3	3821725	3745700	3670153	3595113	3520605
-0,2	4208071	4130068	4052403	3975105	3898203
-0,1	4602112	4522868	4443811	4364973	4286382

Продолжение табл. 2

1	2	3	4	5	6
0	5000056	5079770	5159454	5239074	5318601
0,1	5398002	5477247	5556305	5635144	5713736
0,2	5792049	5870053	5947720	6025019	6101923
0,3	6178403	6254430	6329979	6405021	6479531
0,4	6553483	6626851	6699612	6771742	6843217
0,5	6914014	6984114	7053493	7122132	7190012
0,6	7257114	7323420	7388913	7453577	7517397
0,7	7580357	7642444	7703646	7763950	7823346
0,8	7881823	7939372	7995984	8051652	8106370
0,9	8160131	8212931	8264765	8315630	8365524
1,0	8414445	8462392	8509367	8555368	8600399
1,1	8644461	8687558	8729693	8770872	8811099
1,2	8850380	8888722	8926133	8962619	8998190
1,3	9032855	9066622	9099503	9131507	9162645
1,4	9192930	9222372	9250984	9278779	9305768
1,5	9331967	9357387	9382043	9405949	9429118
1,6	9451565	9473304	9494351	9514718	9534422
1,7	9553477	9571897	9589698	9606894	9623500
1,8	9639530	9654999	9669922	9684312	9698184
1,9	9711552	9724430	9736831	9748769	9760257
2,0	9771309	9781937	9792153	9801970	9811401
2,1	9820456	9829149	9837489	9845488	9853156
2,2	9860505	9867545	9874285	9880736	9886906
2,3	9892806	9898444	9903828	9908968	9913871
2,4	9918546	9923001	9927243	9931280	9935118
2,5	9938766	9942230	9945516	9948632	9951583
2,6	9954377	9957019	9959516	9961873	9964097
2,7	9966192	9968166	9970023	9971769	9973410
2,8	9974951	9976397	9977753	9979025	9980218
2,9	9981336	9982385	9983369	9984293	9985162
3,0	9985980	9986751	9987480	9988170	9988825
3,1	9989449	9990045	9990617	9991167	9991698
3,2	9992212	9992711	9993198	9993672	9994137
3,3	9994591	9995037	9995472	9995898	9996313
3,4	9996716	9997106	9997481	9997838	9998175
3,5	9998490	9998779	9999040	9999269	9999463
3,6	9999620	9999737	9999811	9999842	9999826
3,7	9999766	9999661	9999514	9999330	9999114
3,8	9998875	9998623	9998374	9998144	9997956
3,9	9997837	9997817	9997935	9998234	9998767
4,0	9999593				

Во-вторых, данные таблицы 2 были использованы для составления задачи регрессии с целью получения аналитической зависимости в виде алгебраического полинома регрессии

$$F(u) = \eta(\bar{b}, \bar{u}), \qquad (2)$$

где $\boldsymbol{b}$ – эмпирические коэффициенты регрессии.

С использованием метода регрессионного анализа, включающего метод наименьших квадратов, получена зависимость («) в виде уравнения тринадцатой степени по аргументу U, которое с высокой степенью точности воспроизводит табличную функцию F(±u).

Значения коэффициентов регрессии B(i) для уравнения зависимости F(u) от статистики U приводим ниже (14 коэффициентов)

5.00005592187335E-0001
3.98598844443313E-0001
1.03515179165209E-0005
-6.54749455349020E-0002
-1.45789153812304E-0005
9.14752711739197E-0003
5.10973768210604E-0006
-8.85709923119649E-0004
-7.40906522448195E-0007
5.48773595507313E-0005
4.76628001280766E-0008
-1.92680234034626E-0006
-1.11873972911260E-0009
2.88511374453072E-0008.

По этому уравнению получили столбец 5 таблицы 1 для значений F(±u), который дает возможность по всем строкам таблицы 1 производить расчет теоретической вероятности P_i по единой схеме

$$F(\pm u)_{i+1} - F(\pm u)_i.$$

В-третьих, таблицы χ^2-распределения Пирсона [1, 2] интерпретированы аналитической зависимостью в виде уравнения регрессии вида

$$\chi_{табл}{}^2 = f(\bar{b}, \nu), \qquad (3)$$

где $\chi^2{}_{табл}$ – граница критической области; ν – число степеней свободы при принятой вероятности ***p***.

Уравнение 8-й степени с 9-ю коэффициентами регрессии с высокой степенью точности воспроизводит табличные значения функции χ^2. Ниже для примера воспроизводим значения коэффициентов регрессии для вероятности р = 0,95.

Коэффициенты регрессии B(i) для уравнения критических границ «хи» - квадрат-распределения при Р = 0,95

3,06433322098383E+0000
1,73011778826185E+0000
-2,86218706529553E-0002
9,99199810586049E-0004
-2,32445651853208E-0005
3,45634338701911E-0007
-3,13382273001714E-0009
1,57614536933970E-0011
-3,36110681066872E-0014

Итак, имея в программе значения критериев $\chi^2{}_{раб}$ и $\chi^2{}_{таб}$, рассчитанные по уравнениям (1) и (3), и сопоставляя их, можем выносить заключение о гипотезе нормального распределения рассматриваемой случайной величины.

В заключение приведем результаты программного анализа гипотезы о нормальном распределении случайной величины «Размер обуви взрослых мужчин». Количество значений случайной величины – 500. Расчетное значение Мх – 41,002; σ^2 = 10,806.

При выполнении программы получены значения

$$\chi^2{}_{раб} = 11{,}88 \text{ и } \chi^2{}_{таб} = \mathbf{!6{,}92},$$

т. е. гипотезу о нормальном распределении массива значений данной случайной величины отклонять нет оснований.

Список использованной литературы:
1. Бородюк В.П., Вощинин А.П., Иванов А.З. и др. Статистические методы в инженерных исследованиях. – М.: Высшая школа, 1983. – 216 с., ил.
2. Гмурман В.Е. Теория вероятностей и математическая статистика. Учебное пособие. – М.: Высшая школа, 2001. – 479 с.: ил.