

Statystyka i teoria obsługi masowej

Laboratorium 2

Damian Kwaśniok

1. Na poziomie istotności  $\alpha = 0.07$  zweryfikuj hipotezę mówiącą o tym, że rozkład zmiennej, z której próba znajduje się w arkuszu 02\_Testy\_zgodnosci\_zad1.sta jest rozkładem wykładniczym. Parametr  $\mu$  rozkładu oszacuj na podstawie wyników z próby. Pamiętaj o uwzględnieniu wag przypadków.

H<sub>0</sub> = Rozkład zmiennej v4 (liczebność) jest rozkładem wykładniczym  
H<sub>1</sub> = Rozkład zmiennej v4 (liczebność) nie jest rozkładem wykładniczym

Zmienna v5 (Liczebność oczekiwana) jest dystrybuantą rozkładu wykładniczego Exp( $\lambda=0,37$ ) obliczoną ze wzoru:

$$np_i = 100 * (Exp(a_{i+1}; 0,37) - Exp(a_i; 0,37))$$

gdzie a<sub>i</sub> i a<sub>i+1</sub> sa lewa i prawa granica *i*-tej klasy.

Dla tak wyznaczonej zmiennej **v5** został wykonany test Chi<sup>2</sup> dla liczności obserwowanych wz. oczekiwanych z uwzględnieniem wag przypadków. Poniższa tabela pokazuje wyniki otrzymanego testu:

Przypadek	Liczności obserwowane i oczekiwane (02_Testy_zgodnosci_zad1) Chi kwadrat= 71,47844 df = 99 p = ,983202 UWAGA: Różne sumy oczekiwanych i obserwowanych			
	obserw. liczebność	oczekiw. Liczebność oczekiwana	obs-ocz	(ob-oc) <sup>2</sup> /ocz
C: 1	24,000	30,927	-6,927	1,55133
C: 2	24,000	30,927	-6,927	1,55133
C: 3	24,000	30,927	-6,927	1,55133
C: 4	24,000	30,927	-6,927	1,55133
C: 5	24,000	30,927	-6,927	1,55133
C: 6	24,000	30,927	-6,927	1,55133
C: 7	24,000	30,927	-6,927	1,55133
C: 8	24,000	30,927	-6,927	1,55133
C: 9	24,000	30,927	-6,927	1,55133
C: 10	24,000	30,927	-6,927	1,55133
C: 11	24,000	30,927	-6,927	1,55133
C: 12	24,000	30,927	-6,927	1,55133
C: 13	24,000	30,927	-6,927	1,55133
C: 14	24,000	30,927	-6,927	1,55133
C: 15	24,000	30,927	-6,927	1,55133
C: 16	24,000	30,927	-6,927	1,55133
C: 17	24,000	30,927	-6,927	1,55133
C: 18	24,000	30,927	-6,927	1,55133
C: 19	24,000	30,927	-6,927	1,55133
C: 20	24,000	30,927	-6,927	1,55133
C: 21	24,000	30,927	-6,927	1,55133
C: 22	24,000	30,927	-6,927	1,55133
C: 23	24,000	30,927	-6,927	1,55133
C: 24	24,000	30,927	-6,927	1,55133
C: 25	21,000	21,362	-0,362	0,00614
C: 26	21,000	21,362	-0,362	0,00614
C: 27	21,000	21,362	-0,362	0,00614
C: 28	21,000	21,362	-0,362	0,00614
C: 29	21,000	21,362	-0,362	0,00614
C: 30	21,000	21,362	-0,362	0,00614
C: 31	21,000	21,362	-0,362	0,00614
C: 32	21,000	21,362	-0,362	0,00614
C: 33	21,000	21,362	-0,362	0,00614
C: 34	21,000	21,362	-0,362	0,00614
C: 35	21,000	21,362	-0,362	0,00614
C: 36	21,000	21,362	-0,362	0,00614
C: 37	21,000	21,362	-0,362	0,00614
C: 38	21,000	21,362	-0,362	0,00614
C: 39	21,000	21,362	-0,362	0,00614
C: 40	21,000	21,362	-0,362	0,00614
C: 41	21,000	21,362	-0,362	0,00614
C: 42	21,000	21,362	-0,362	0,00614
C: 43	21,000	21,362	-0,362	0,00614
C: 44	21,000	21,362	-0,362	0,00614
C: 45	21,000	21,362	-0,362	0,00614
C: 46	15,000	14,755	0,245	0,00405
C: 47	15,000	14,755	0,245	0,00405
C: 48	15,000	14,755	0,245	0,00405
C: 49	15,000	14,755	0,245	0,00405
C: 50	15,000	14,755	0,245	0,00405
C: 51	15,000	14,755	0,245	0,00405
C: 52	15,000	14,755	0,245	0,00405
C: 53	15,000	14,755	0,245	0,00405
C: 54	15,000	14,755	0,245	0,00405
C: 55	15,000	14,755	0,245	0,00405
C: 56	15,000	14,755	0,245	0,00405
C: 57	15,000	14,755	0,245	0,00405
C: 58	15,000	14,755	0,245	0,00405
C: 59	15,000	14,755	0,245	0,00405
C: 60	15,000	14,755	0,245	0,00405
C: 61	12,000	10,192	1,808	0,32068
C: 62	12,000	10,192	1,808	0,32068
C: 63	12,000	10,192	1,808	0,32068
C: 64	12,000	10,192	1,808	0,32068
C: 65	12,000	10,192	1,808	0,32068
C: 66	12,000	10,192	1,808	0,32068
C: 67	12,000	10,192	1,808	0,32068
C: 68	12,000	10,192	1,808	0,32068
C: 69	12,000	10,192	1,808	0,32068
C: 70	12,000	10,192	1,808	0,32068
C: 71	12,000	10,192	1,808	0,32068
C: 72	12,000	10,192	1,808	0,32068
C: 73	9,000	7,040	1,960	0,54565
C: 74	9,000	7,040	1,960	0,54565
C: 75	9,000	7,040	1,960	0,54565
C: 76	9,000	7,040	1,960	0,54565
C: 77	9,000	7,040	1,960	0,54565
C: 78	9,000	7,040	1,960	0,54565
C: 79	9,000	7,040	1,960	0,54565
C: 80	9,000	7,040	1,960	0,54565
C: 81	9,000	7,040	1,960	0,54565
C: 82	7,000	4,863	2,137	0,93929
C: 83	7,000	4,863	2,137	0,93929
C: 84	7,000	4,863	2,137	0,93929
C: 85	7,000	4,863	2,137	0,93929
C: 86	7,000	4,863	2,137	0,93929
C: 87	7,000	4,863	2,137	0,93929
C: 88	7,000	4,863	2,137	0,93929
C: 89	4,000	3,359	0,641	0,12236
C: 90	4,000	3,359	0,641	0,12236
C: 91	4,000	3,359	0,641	0,12236
C: 92	4,000	3,359	0,641	0,12236
C: 93	5,000	2,320	2,680	3,09545
C: 94	5,000	2,320	2,680	3,09545
C: 95	5,000	2,320	2,680	3,09545
C: 96	5,000	2,320	2,680	3,09545
C: 97	5,000	2,320	2,680	3,09545
C: 98	3,000	5,182	-2,182	0,91871
C: 99	3,000	5,182	-2,182	0,91871
C: 100	3,000	5,182	-2,182	0,91871
Sum	1566,000	1672,460	-106,460	71,47844

Przyjmując poziom istotności  $\alpha = 0,07$  nie mamy podstaw do odrzucenia hipotezy H<sub>0</sub>, gdyż wyliczony przez program komputerowy poziom istotności p = 0,983202 jest większy niż  $\alpha$ .

2. Na podstawie arkusza 02\_Testy\_zgodnosci\_zad2.sta<sup>1</sup> sprawdź, czy pierwsza zmienna może pochodzić z rozkładu jednostajnego a druga z rozkładu geometrycznego. Parametry należy oszacować na podstawie dostępnych prób. Za poziom istotności przyjmij  $\alpha = 0.1$ . Dla obu zmiennych utwórz histogram z naniesioną gęstością dopasowywanego rozkładu oraz wykres dystrybuanty empirycznej z teoretyczną.

W pierwszym kroku zostały przeliczone dane z 3 zmiennych i otrzymano następujące wyniki:

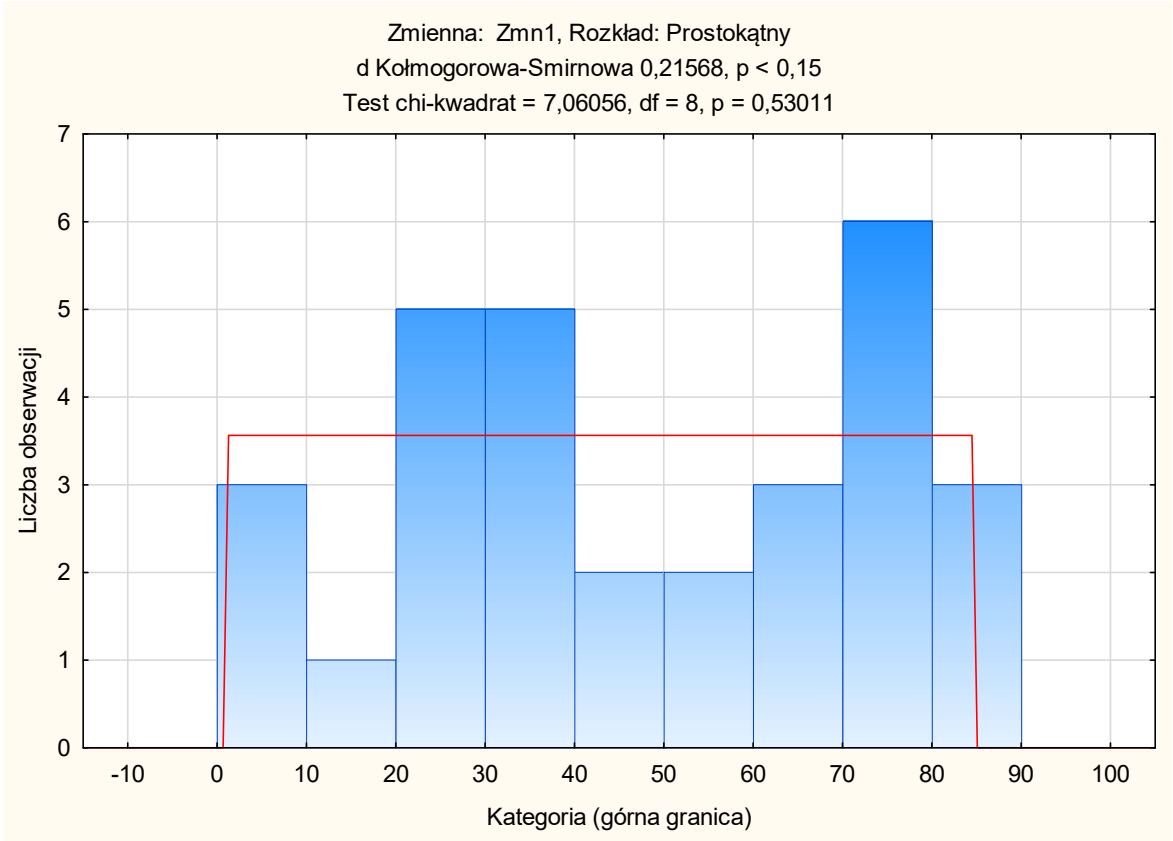
	1	2	3
	Zmn1	Zmn2	Zmn3
1	33,8	81	-0,59
2	60,6	69	0,75
3	79,9	37	-0,79
4	77,8	1	0,4
5	89,7	78	-0,88
6	80	3	-0,52
7	7,8	45	-1,15
8	21,2	3	1,05
9	37,3	145	-0,02
10	67,4	64	-1,16
11	14,3	65	-1,16
12	80,2	46	-1,62
13	79,3	24	0,48
14	29,2	34	4
15	3,3	91	3,09
16	32,2	2	-0,16
17	81,1	1	1,02
18	29,5	110	0,18
19	49	28	-1,42
20	40,8	22	1,28
21	79,8	15	1,11
22	26,7	82	-0,59
23	67,7	95	1,95
24	79,9	12	0,64
25	34,7	22	2,05
26	37,8	13	0,11
27	50,4	36	1,5
28	51,9	138	0,19
29	7	127	-1,93
30	22,2	23	0,92

Następnie wykonano dopasowanie rozkładu prostokątnego (jednostajnego), którego wynik prezentuje się następująco:

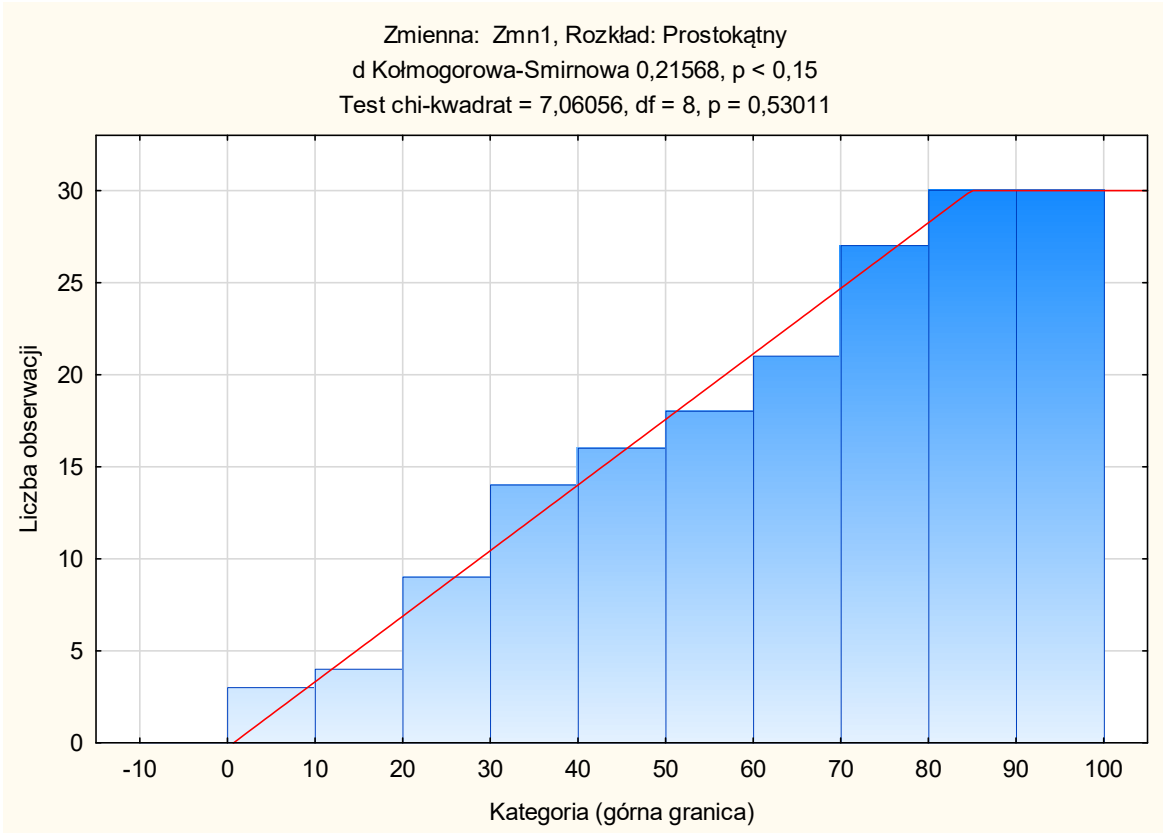
Zmienna: Zmn1, Rozkład: Prostokątny (02_Testy_zgodnosci_zad2)									
d Kołmogorowa-Smirnowa 0,21568, p < 0,15									
Chi-kwadrat = 7,06056, df = 8, p = 0,53011									
Górna Granica	Obserw. Liczność	Skumulow. Obserw.	Procent Obserw.	Skumul. % Obserw.	Oczekiwana Liczność	Skumulow. Oczekiwana	Procent Oczekiwana	Skumul. % Oczekiwana	Obserw. - Oczekiwana
<= 0,00000	0	0	0,00000	0,0000	0,000000	0,00000	0,00000	0,0000	0,00000
10,00000	3	3	10,00000	10,0000	3,313539	3,31354	11,04513	11,0451	-0,31354
20,00000	1	4	3,33333	13,3333	3,562945	6,87648	11,87648	22,9216	-2,56295
30,00000	5	9	16,66667	30,0000	3,562945	10,43943	11,87648	34,7981	1,43705
40,00000	5	14	16,66667	46,6667	3,562945	14,00238	11,87648	46,6746	1,43705
50,00000	2	16	6,66667	53,3333	3,562945	17,56532	11,87648	58,5511	-1,56295
60,00000	2	18	6,66667	60,0000	3,562945	21,12827	11,87648	70,4276	-1,56295
70,00000	3	21	10,00000	70,0000	3,562945	24,69121	11,87648	82,3040	-0,56295
80,00000	6	27	20,00000	90,0000	3,562945	28,25416	11,87648	94,1805	2,43705
90,00000	3	30	10,00000	100,0000	1,745843	30,00000	5,81948	100,0000	1,25416
<nieskończoność	0	30	0,00000	100,0000	0,000000	30,00000	0,00000	100,0000	0,00000

Z otrzymamanych wyników wynika wniosek, że przyjmując poziom istotności  $\alpha = 0,1$  nie ma podstaw do odrzucenia hipotezy, że zmienna **Zmn1** pochodzi z rozkładu jednostajnego (prostokątnego), gdyż wartość  $\chi^2(1-\alpha)$  dla 8 stopni swobody wynosi 13,361566, a więc otrzymany wynik testu 7,06056 nie należy do przedziału krytycznego  $<13,361566, +\infty$ ).

Histogram z gęstością dopasowanego rozkładu prezentuje się następująco:



Natomiast wykres dystrybuanty empirycznej i teoretycznej wygląda następująco:

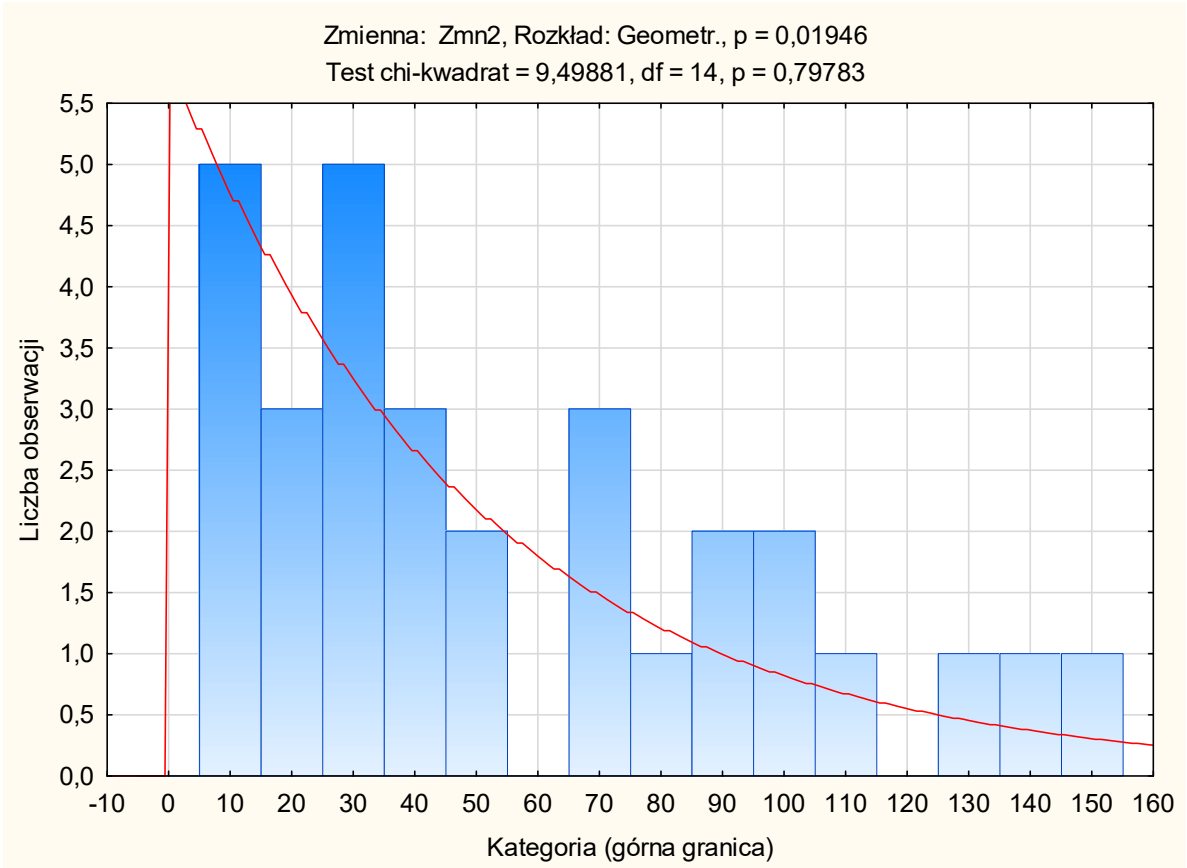


Analogiczne czynności wykonałem dla drugiej zmiennej **Zmn2** dopasowując ją do rozkładu geometrycznego. Wynik tego dopasowanie prezentuje się następująco:

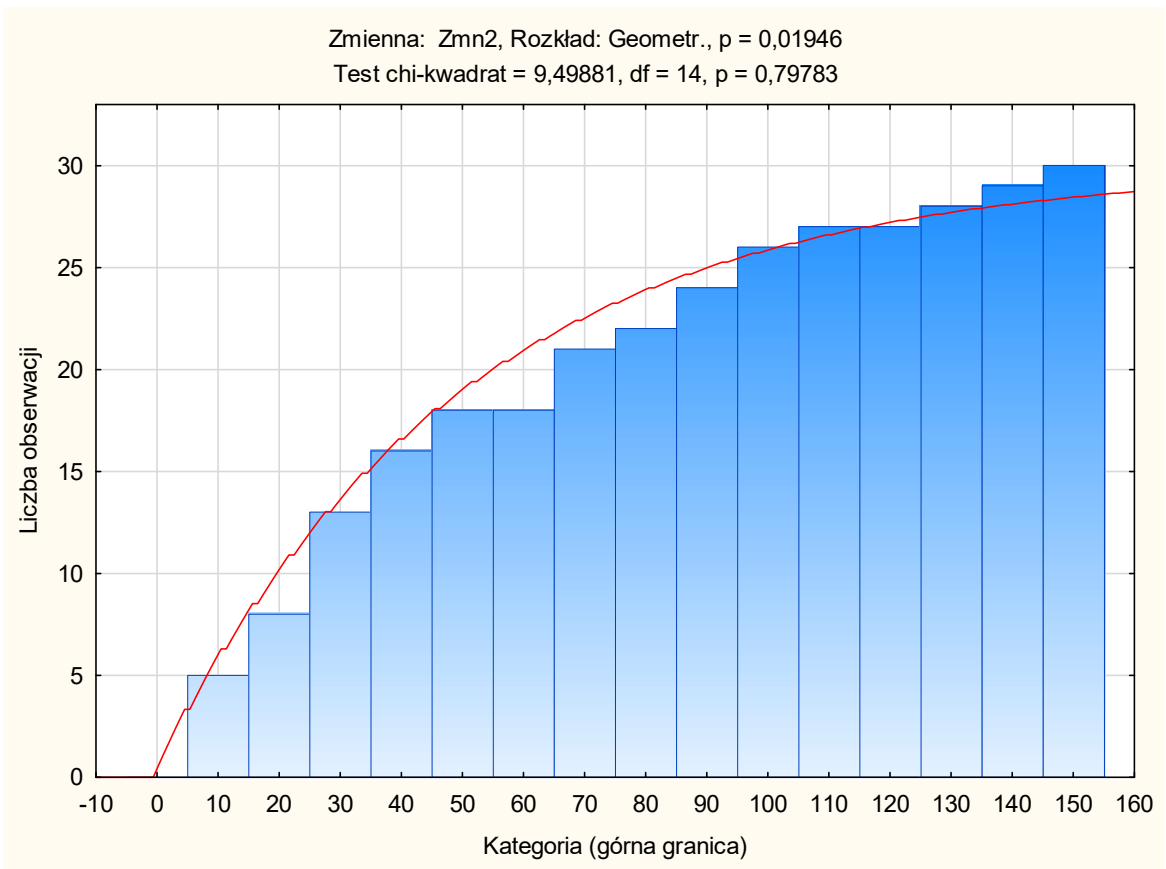
Zmienna: Zmn2, Rozkład: Geometr., p = 0,01946 (02_Testy_zgodnosci_zad2)									
Chi-kwadrat = 9,49881, df = 14, p = 0,79783									
Klasa	Obserw. Liczność	Skumulow. Obserw.	Procent Obserw.	Skumul. % Obserw.	Oczekiwana Liczność	Skumulow. Oczekiwana	Procent Oczekiwana	Skumul. % Oczekiwana	Obserw. - Oczekiwana
<= 0,00000	0	0	0,00000	0,0000	0,583658	0,58366	1,94553	1,9455	-0,58366
10,00000	5	5	16,66667	16,6667	5,247111	5,83077	17,49037	19,4358	-0,24711
20,00000	3	8	10,00000	26,6667	4,311163	10,14193	14,37054	33,8064	-1,31116
30,00000	5	13	16,66667	43,3333	3,542163	13,68405	11,80721	45,6136	1,45784
40,00000	3	16	10,00000	53,3333	2,910333	16,59443	9,70111	55,3148	0,08967
50,00000	2	18	6,66667	60,0000	2,391206	18,98563	7,97069	63,2854	-0,39121
60,00000	0	18	0,00000	60,0000	1,964677	20,95031	6,54892	69,8344	-1,96468
70,00000	3	21	10,00000	70,0000	1,614225	22,56454	5,38076	75,2151	1,38577
80,00000	1	22	3,33333	73,3333	1,326293	23,89083	4,42098	79,6361	-0,32629
90,00000	2	24	6,66667	80,0000	1,089717	24,98055	3,63239	83,2685	0,91028
100,00000	2	26	6,66667	86,6667	0,895340	25,87589	2,98447	86,2530	1,10466
110,00000	1	27	3,33333	90,0000	0,735634	26,61152	2,45211	88,7051	0,26437
120,00000	0	27	0,00000	90,0000	0,604416	27,21594	2,01472	90,7198	-0,60442
130,00000	1	28	3,33333	93,3333	0,496604	27,71254	1,65535	92,3751	0,50340
140,00000	1	29	3,33333	96,6667	0,408023	28,12057	1,36008	93,7352	0,59198
<nieskończoność	1	30	3,33333	100,0000	1,879434	30,00000	6,26478	100,0000	-0,87943

Z otrzymamanych wyników wynika wniosek, że przyjmując poziom istotności  $\alpha = 0,1$  nie ma podstaw do odrzucenia hipotezy, że zmienna **Zmn2** pochodzi z rozkładu geometrycznego, gdyż wartość  $\chi^2(1-\alpha)$  dla 14 stopni swobody wynosi 21,064144, a więc otrzymany wynik testu 9,49881 nie należy do przedziału krytycznego  $<21,064144, +\infty$ ).

Histogram z gęstością dopasowanego rozkładu prezentuje się następująco:

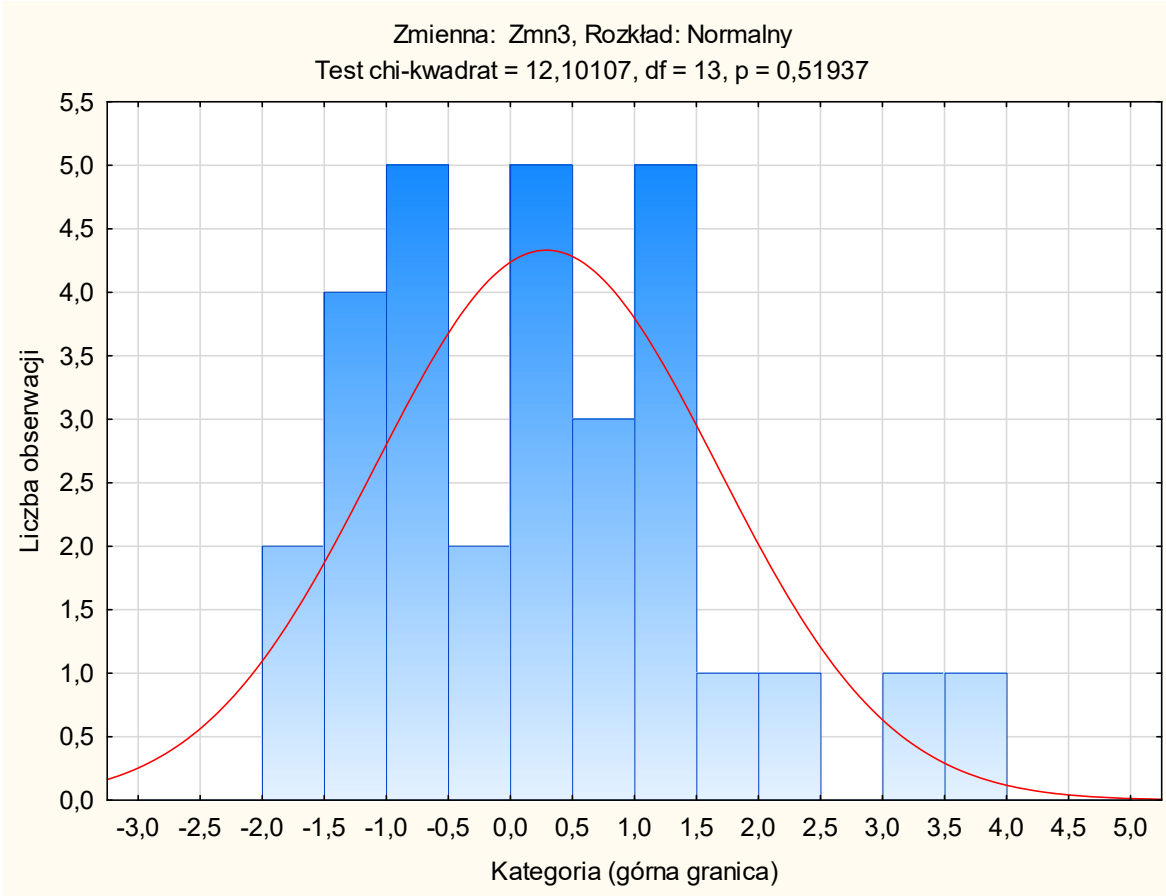
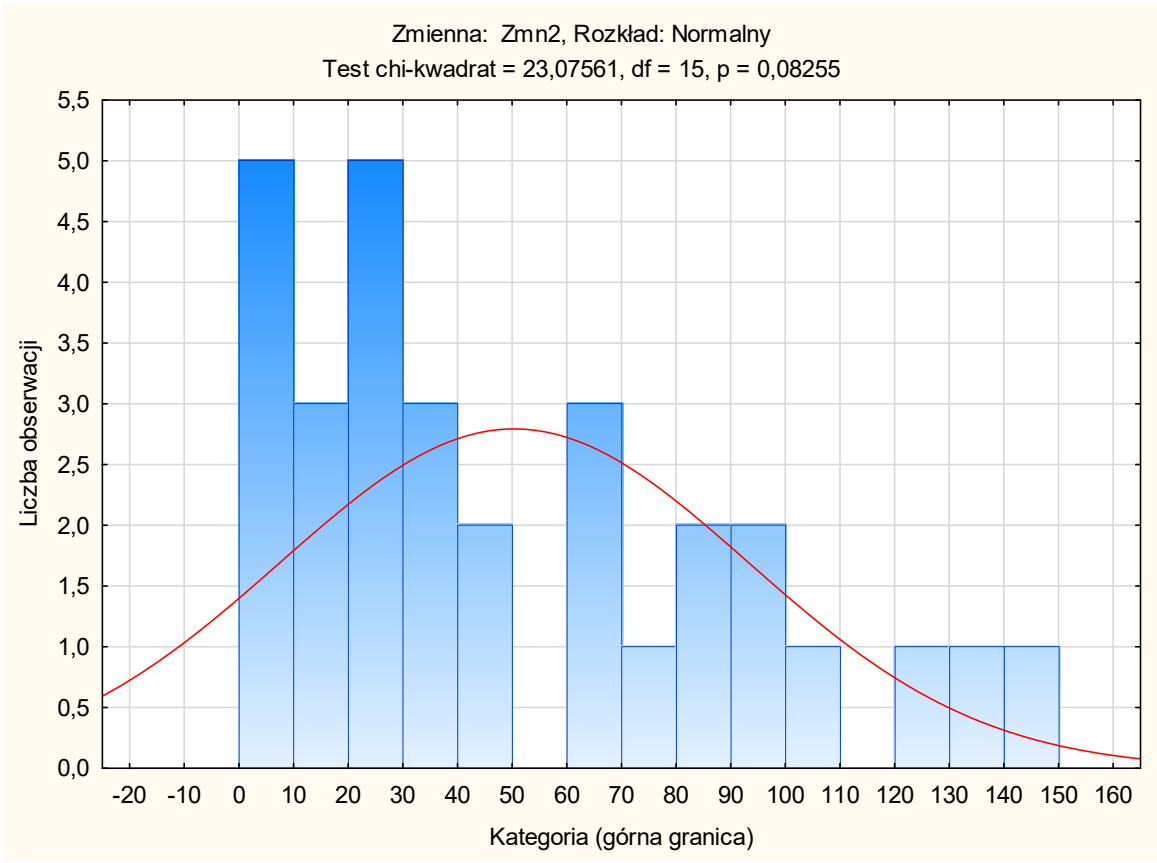
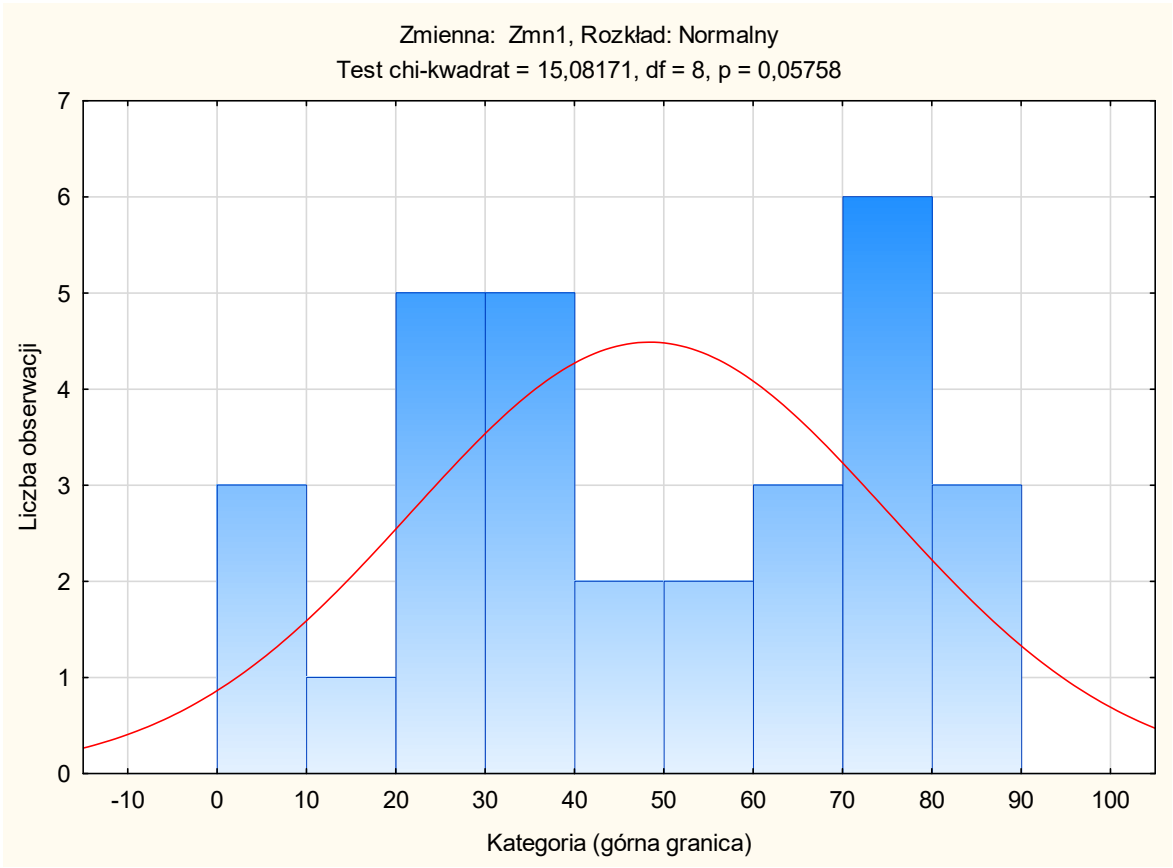


Natomiast wykres dystrybuanty empirycznej i teoretycznej wygląda następująco:



3. Sprawdź, które ze zmiennych w arkuszu 02\_Testy\_zgodnosci\_zad2.sta mogą pochodzić z rozkładu normalnego. Odpowiedź uzasadnij i zwizualizuj przy pomocy odpowiednich wykresów.

Dopasowania trzech zmiennych do rozkładu normalnego prezentują się następująco:



Na podstawie powyższych wykresów można stwierdzić, że tylko trzecia zmienna **Zmn3** może pochodzić z rozkładu normalnego. Jest to widoczne "gołym okiem" widząc jak wykres dystrybuanty teoretycznej w pewnym stopniu pokrywa się z dystrybuantą empiryczną.

Wniosek ten potwierdza również wyliczony przez program poziom istotności **p**, który tylko dla dopasowanie trzeciej zmiennej do rozkładu normalnego jest stosunkowo duży i wynosi około **0,51**.