

Zadania z Analizy danych

zestaw 6

Zadanie 1 (Test F -Fishera hipotezy o równości wariancji)

Za pomocą dwóch przyrządów pomiarowych wykonano pomiary pewnej wielkości fizycznej. Wyniki pomiarów są następujące

Przyrząd 1: 18; 18; 15; 27; 26; 22; 21; 19; 21

Przyrząd 2: 16; 23; 22; 19; 21; 22; 20

Stosując test obustronny F -Fishera-Snedecora przy wartości poziomu istotności 10% (0,1) określ, czy możemy uważać, że dokładności obu przyrządów (czyli wariancje obu prób) są takie same.

Rozwiązanie

Dla jednej i drugiej próby możemy zdefiniować zmienne

$$Q_1^2 = \frac{(n_1 - 1)S^2(X_1)}{\sigma^2(X_1)},$$

$$Q_2^2 = \frac{(n_2 - 1)S^2(X_2)}{\sigma^2(X_2)}$$

podlegające rozkładowi χ^2 odpowiednio dla $n_1 - 1$ i $n_2 - 1$ stopni swobody. Zakładając równość wariancji $\sigma^2(X_1) = \sigma^2(X_2)$ statystykę testową F możemy wyrazić przez

$$F = \frac{n_2 - 1}{n_1 - 1} \frac{Q_1^2}{Q_2^2} = \frac{S^2(X_1)}{S^2(X_2)}.$$

Statystyka ta przyjmuje tylko dodatnie wartości i podlega rozkładowi Fischera-Snedecora $F(n_1, n_2)$. Jako hipotezę alternatywną przyjmujemy hipotezę o braku równości wariancji: $\sigma^2(X_1) \neq \sigma^2(X_2)$ lub hipotezę, że wariancja w liczniku jest większa od wariancji w mianowniku: $\sigma^2(X_1) > \sigma^2(X_2)$. Do określania obszarów krytycznych korzystamy z kwantyli rozkładu Fischera-Snedecora.

$$F \approx 2,66$$

- $H_1: \sigma^2(X_1) \neq \sigma^2(X_2)$, obszar krytyczny składa się z dwóch przedziałów $F < F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)$ i $F > F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)$.

F0,05(8,6)		0,279284		
F0,95(8,6)		4,146804		
F0,05(8,6)	<=	F	<=	F0,95(8,6)
0,279284	<=	2,659605	<=	4,146804

Wartość statystyki leży poza obszarem krytycznym, a zatem na poziomie istotności 0,1 nie możemy odrzucić hipotezy o równości wariancji.

- $H_1: \sigma^2(X_1) > \sigma^2(X_2)$, $F > F_{1-\alpha}(n_1 - 1, n_2 - 1)$.

poziom istotności α		0,1			
$1-\alpha=$		0,9			
Kwantyl rozkładu F-Fishera l. st. swobody 8, 6 i prawdopodob. 0,9					
$F_{0,9(8,6)}$		2,983036			
F	\leq	$F_{0,9(8,6)}$			
2,659605	\leq	2,983036			

Wartość statystyki leży poza obszarem krytycznym, a zatem na poziomie istotności 0,1 nie możemy odrzucić hipotezy o równości wariancji.

Zadanie 2 (Test Studenta hipotezy o średniej równej danej liczbie)

Z populacji o rozkładzie normalnym wylosowano 30 elementów. Przyjmując poziom istotności 10% (0,1) zweryfikuj hipotezę stwierdzającą, że wylosowane elementy należą do populacji o wartości średniej równej 25,5. Wartości wylosowanych elementów są następujące

25,02; 26,12; 24,78; 26,19; 25,57; 25,33; 23,45; 23,5; 25,42; 22,45
 27,56; 24,67; 22,51; 28,01; 24,77; 25,66; 25,02; 27,91; 21,98; 24,55
 24,51; 22,55; 25,7; 25,03; 26,72; 24,32; 26,23; 28,11; 23,79; 25,05

Rozwiązanie:

Statystyka:

$$t = \frac{\bar{X} - E_0}{S(\bar{X})} \approx -1,424,$$

dla określenia obszarów granicznych wykorzystujemy kwantyle rozkładu t -Studenta:

- $H_1: \bar{X} \neq E_0$, $|t| > t_{n-1, 1-\alpha/2}$, czyli $-t_{n-1, 1-\alpha/2} > t > t_{n-1, 1-\alpha/2}$.
- $H_1: \bar{X} > E_0$, $t > t_{n-1, 1-\alpha}$.
- $H_1: \bar{X} < E_0$, $t < t_{n-1, \alpha}$.

Zastosujemy test dwustronny.

poziom istotności α		0,1			
$1-\alpha/2=$		0,95			
Kwantyl rozkładu T-Studenta dla prawd. 0,9 i 29 st. swobody					
$T(0,95,29)=$	1,699127				
$ t =$	1,424126				
1,424126	<	1,699127			

Wartość statystyki leży poza obszarem krytycznym, a zatem na poziomie istotności 0,1 nie możemy odrzucić hipotezy mówiącej, że wylosowane elementy należą do populacji o wartości średniej równej 25,5.

Zadanie 3 (Test Studenta hipotezy o równości wartości średnich dwóch serii pomiarów)

Wykonano pomiar stężenia kwasu neuraminowego zawartego w czerwonych ciałkach krwi u pacjentów zmarłych na skutek pewnej choroby krwi (grupa x) oraz u osób zdrowych z grupy

kontrolnej (grupa y). Przyjmując poziom istotności 10% (0,1), określ czy uzyskany materiał doświadczalny jest wystarczający do wykazania związku pomiędzy stężeniem kwasu neuraminowego a śmiercią pacjenta. Wyniki obu serii pomiarów (w jednostkach umownych) są następujące:

Grupa x: 21; 24; 18; 19; 25; 17; 18; 22; 21; 23; 18; 13; 16; 23; 22; 24

Grupa y: 16; 20; 22; 19; 18; 19; 19

Rozwiązanie

Nie znamy wariancji obu populacji, więc przed przystąpieniem do testu musimy sprawdzić, czy wariancje są sobie równe. W tym celu przeprowadzamy test F -Fishera-Snedecora jak w zadaniu 1.

$$F = \frac{S^2(X_1)}{S^2(X_2)} \approx 3,549.$$

Założmy hipotezę alternatywną $H_1: \sigma^2(X_1) \neq \sigma^2(X_2)$. Wówczas obszar krytyczny składa się z dwóch przedziałów $F < F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)$ i $F > F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)$.

poziom istotności α		0,1		
$\alpha/2=$	0,05		$1-\alpha/2=$	0,95
Kwantyle rozkładu F-Fishera l. st. swobody 14, 6 i prawdopodob. 0,05 i 0,95				
F0,05(14,6)		0,351157		
F0,95(14,6)		3,955934		
F0,05(8,6)	<=	F	<=	F0,95(8,6)
0,351157	<=	3,548571	<=	3,955934

Statystyka leży poza obszarem krytycznym, więc na poziomie istotności 0,1 możemy przyjąć, że wariancje obu populacji są jednakowe.

Teraz przystępujemy do testu hipotezy H_0 : średnie obu populacji są jednakowe. Skoro wariancje obu populacji są jednakowe, to statystyka przyjmuje postać

$$t = \frac{\bar{X} - \bar{Y}}{S(X, Y) \sqrt{\frac{n_x + n_y}{n_x n_y}}},$$

gdzie

$$S(X, Y) = \sqrt{\frac{(n_x - 1)S^2(X) + (n_y - 1)S^2(Y)}{n_x + n_y - 2}}.$$

Statystyka podlega rozkładowi t -Studenta dla $\nu = n_x + n_y - 2$ stopni swobody.

liczba stopni swobody=		20				
Xśr.	Yśr.	S(X)	S^2(X)	S(Y)	S^2(Y)	S(X,Y)
20,4	19	3,439269	11,82857	1,825742	3,333333	2,877499
t=	1,062909					

poziom istotności α		0,1				
$\alpha/2=$	0,05		$1-\alpha/2=$	0,95		
Kwantyle rozkładu t-Studenta dla l. st. swobody 20 i prawdopodob. 0,05 i 0,95						
$t_{0,05}(20)$		-1,72472				
$t_{0,95}(20)$		1,72472				
$F_{0,05}(8,6)$	\leq	t	\leq	$F_{0,95}(8,6)$		
-1,72472	\leq	1,062909	\leq	1,724718		

Statystyka leży poza obszarem krytycznym, więc na poziomie istotności 0,1 nie możemy odrzucić hipotezy o równości wartości średnich, a co za tym idzie nie możemy stwierdzić, że istnieje związek pomiędzy stężeniem kwasu neuraminowego a śmiercią pacjenta.

Zadanie 4 (Jednoczynnikowy test ANOVA)

W dwóch grupach wykonano pomiar pewnej zmiennej losowej. Grupa A poddana była pewnemu badanemu czynnikowi, zaś grupa B stanowiła grupę kontrolną niepoddaną temu czynnikowi. Stosując jednoczynnikowy test ANOVA na poziomie istotności $\alpha = 0,05$, odpowiedz na pytanie, czy istnieje wpływ badanego czynnika na mierzoną zmienną. Wyniki pomiarów podaje poniższa tabela

Grupa A	Grupa B
5,97	4,15
4,44	11,63
3,82	4,60
7,32	5,83
6,19	6,78
6,62	8,55
5,21	6,04
8,11	
7,14	
9,29	

Rozwiązanie

Przed wykonaniem testu sprawdź hipotezę o równości wariancji tych dwóch populacji. W tym celu wykonaj test Fischera-Snedecora.

Następnie wyliczamy statystykę

$$F \equiv \frac{s_b^2}{s_w^2} = \frac{N - k}{k - 1} \cdot \frac{Q_b}{Q_w},$$

gdzie $k = 3$ (liczba grup), $N = 30$ (liczba wszystkich pomiarów),

$$Q_b = \sum_{i=1}^k n_i \bar{x}_i^2 - N \bar{x}^2, \quad Q_w = \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i=1}^k n_i \bar{x}_i^2$$

$n_1 = 10$, $n_2 = 7$ (liczba pomiarów w poszczególnych grupach).

Zmienna F podlega rozkładowi Fischera-Snedecora dla $k - 1 = 1$ i $N - k = 17 - 2 = 15$ stopni swobody.

Hipotezę zerową musimy odrzucić jeśli przy zadanym poziomie istotności α zachodzi nierówność

$$F > F_{1-\alpha}(k - 1, N - k).$$

Wyniki

<x1>	<x2>	<x>	
6,41	6,80	6,57	
Qb	0,613967		
Qw	64,63323		
Q	65,2472		
Sprawdzenie			
Qb+Qw	65,2472	OK	
Sb^2=Qb/(k-1)		0,613967	
Sw^2=Qw/(N-k)		4,308882	
F=Sb^2/Sw^2		0,142489	
poziom istotności α		0,05	
prawdopodobieństwo		0,95	
Obszar krytyczny (tylko prawostronny)			
$F > F_{1-\alpha}(k - 1, N - k).$			
Kwantyl rozkładu Fishera Snedecora dla l. st. swobody 2-1=1; 17-2=15 i prawdopodobieństwa 0,95			
F0,95(2,27)=		4,543077	
0,142489	>	4,543077	
Hipotezę zerową musimy odrzucić			

Ten sam test możemy wykonać korzystając z Dodatku do Excela „Analiza danych” wybierając funkcję „Analiza wariancji: jednoczynnikowa”. Wyniki są następujące

Analiza wariancji: jednoczynnikowa

PODSUMOWANIE

<i>Grupy</i>	<i>Licznik</i>	<i>Suma</i>	<i>Średnia</i>	<i>Wariancja</i>
Grupa A	10	64,11	6,411	2,762277
Grupa B	7	47,58	6,797143	6,62879

ANALIZA WARIANCJI

<i>Źródło wariancji</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>Wartość- p</i>	<i>Test F</i>
Pomiędzy grupami	0,613967	1	0,613967	0,142489	0,711107	4,543077
W obrębie grup	64,63323	15	4,308882			
Razem	65,2472	16				

Wartość-p oznacza poziom istotności od którego wartość statystyki F będzie na granicy obszaru krytycznego. Liczymy go z wyrażenia

$$\text{Wartość} - p = 1 - \text{Rozkł. } F(F, k - 1, N - k, 1).$$

Zadanie 5

W trzech grupach, każda poddana innemu czynnikowi, wykonano pomiar pewnej zmiennej losowej. Wyniki pomiarów zebrano w poniższej tabeli:

Grupa A	Grupa B	Grupa C
6,04	4,29	5,48
3,03	3,05	5,60
4,61	2,60	8,32
4,93	4,42	6,18
3,50	4,79	2,73
4,23	3,30	8,88
5,15	4,30	4,45
4,55	3,27	6,93
5,86	2,78	2,89
4,14	3,74	5,55

Sprawdź czy istnieje wpływ czynników, jakim poddano poszczególne grupy, na wyniki pomiarów. W tym celu wykonaj jednoczynnikowy test ANOVA dla poziomu istotności $\alpha = 0,05$. Przed wykonaniem testu sprawdź hipotezę o równości wariancji tych trzech populacji. W tym celu wykonaj test Bartletta. Następnie wyliczamy statystykę

$$F \equiv \frac{s_b^2}{s_w^2} = \frac{N - k}{k - 1} \cdot \frac{Q_b}{Q_w},$$

gdzie $k = 3$ (liczba grup), $N = 30$ (liczba wszystkich pomiarów),

$$Q_b = \sum_{i=1}^k n_i \bar{x}_i^2 - N \bar{x}^2, \quad Q_w = \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i=1}^k n_i \bar{x}_i^2$$

$n_1 = n_2 = n_3 = 10$ (liczba pomiarów w poszczególnych grupach)

Zmienna F podlega rozkładowi Fischera-Snedecora dla $k - 1 = 2$ i $N - k = 30 - 3 = 27$ stopni swobody.

Hipotezę zerową musimy odrzucić jeśli przy zadanym poziomie istotności α zachodzi nierówność

$$F > F_{1-\alpha}(k - 1, N - k).$$

Wyniki

<x1>	<x2>	<x3>	<x>			
4,60	3,65	5,70	4,65			
Qb	20,98706					
Qw	50,39777					
Q	71,38483					
Sprawdzenie						
Qb+Qw	71,38483	OK				
Sb^2=Qb/(k-1)		10,49353				
Sw^2=Qw/(N-k)		1,866584				
F=Sb^2/Sw^2		5,621783				
poziom istotności α		0,05				
prawdopodobieństwo		0,95				
Obszar krytyczny (tylko prawostronny)						
$F > F_{1-\alpha}(k - 1, N - k)$						
Kwantyl rozkładu Fishera Snedecora dla l. st. swobody 3-1=2; 30-3=27 i prawdopodobieństwa 0,95						
F0,95(2,27)=		3,354131				
5,621783	>	3,354131		Hipotezę zerową musimy odrzucić		

Ten sam test możemy wykonać korzystając z Dodatku do Excela „Analiza danych” wybierając funkcję „Analiza wariancji: jednoczynnikowa”. Wyniki są następujące

Analiza wariancji: jednoczynnikowa

PODSUMOWANIE

Grupy	Licznik	Suma	Średnia	Wariancja
Grupa A	10	46,04	4,604	0,899827
Grupa B	10	36,54	3,654	0,579871
Grupa C	10	57,01	5,701	4,120054

ANALIZA WARIANCJI

Źródło wariancji	SS	df	MS	F	Wartość-p	Test F
Pomiędzy grupami	20,98706	2	10,49353	5,621783	0,009096468	3,354131
W obrębie grup	50,39777	27	1,866584			
Razem	71,38483	29				

Wartość-p oznacza poziom istotności od którego wartość statystyki F będzie na granicy obszaru krytycznego. Liczymy go z wyrażenia

$$\text{Wartość} - p = 1 - \text{Rozkł. } F(F, k - 1, N - k, 1).$$

Zadanie 6 (Test χ^2 dobroci dopasowania (test Pearsona) dla rozkładu dyskretnego)

Komora pęcherzykowa naświetlana jest wiązką kwantów γ wykorzystywanych do badań oddziaływań fotonów z protonami. Część fotonów w wyniku zderzeń kreuje pary elektron-pozyton. Ten uboczny efekt naświetlania komory jest wykorzystywany do monitorowania wiązki fotonów. Częstość pojawiania się zdjęć przedstawiających 0, 1, 2 itd. par elektron-pozyton powinna podlegać rozkładowi Poissona ($f(k) = \frac{\lambda^k}{k!} e^{-\lambda}$). Analizując odchylenia rozkładu doświadczalnego od rozkładu Poissona można wnioskować o istnieniu strat, które w konsekwencji będą prowadzić do błędów systematycznych prowadzonych eksperymentów. Przeprowadź test χ^2 hipotezy: *rozkład częstości występowania par elektron-pozyton jest zgodny z rozkładem Poissona*. Przyjmij poziom istotności $\alpha = 0,01$. Dane dla testu znajdują się w poniższej tabelce

Uwaga: Próba składa się z $r = 8$ elementów, na jej podstawie wyznaczmy jeden parametr rozkładu (λ), czyli $p = 1$. A zatem liczba stopni swobody rozkładu χ^2 wynosi $f = r - 1 - p = 6$. Dla poziomu istotności $\alpha = 0,01$ i liczby stopni swobody $f = 6$ kwantyl rozkładu χ^2 jest równy $\chi^2_{1-0,01}(6) = \chi^2_{0,99}(6) = 16,81$.

k – liczba par e-p na zdjęciu	n_k – liczba zdjęć zawierających k par	np_k – wartość oczekiwana liczby k par e-p zgodna z rozkładem Poissona	$\frac{(n_k - np_k)^2}{np_k}$
0	44		
1	75		
2	80		
3	73		
4	52		

5	18		
6	9		
7	1		
8	2		
	$n =$		$\chi^2 =$

Porównaj za pomocą histogramu rozkład doświadczalny z rozkładem teoretycznym.

Rozwiązanie

Znajdźmy metodą momentów wartość parametru λ . Wiemy, że dla rozkładu Poissona wartość oczekiwana jest równa

$$E(k) = \lambda.$$

Wiemy też, że wartość oczekiwana jest równa momentowi pierwszego rzędu względem zera

$$E(k) = m_1(0).$$

Zgodnie z metodą momentów moment pierwszego rzędu względem zera możemy estymować przez

$$T(m_1(0)) = \frac{1}{r} \sum_{k=0}^r k_i.$$

Ostatecznie

$$\lambda = E(k) = \frac{1}{r} \sum_{k=0}^r k_i = \frac{1}{8} (1 + 2 + 3 + 4 + 5 + 6 + 7 + 8) = 4.$$

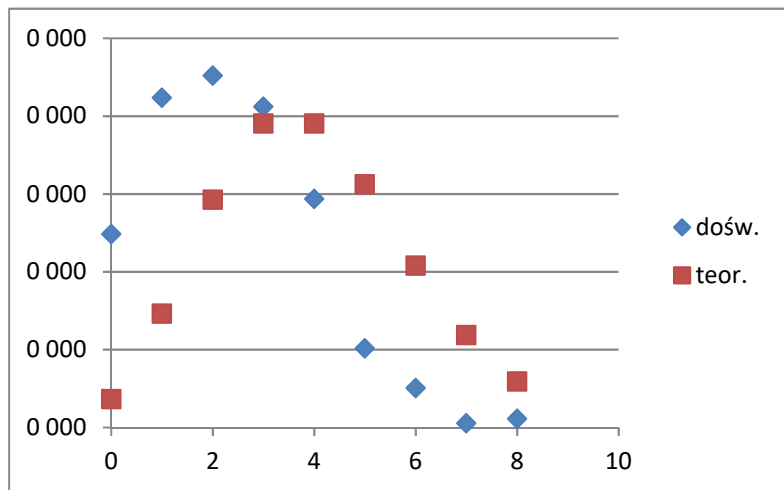
k – liczba par e-p na zdjęciu	n_k – liczba zdjęć zawierających k par	np_k – wartość oczekiwana liczby k par e-p zgodna z rozkładem Poissona	$\frac{(n_k - np_k)^2}{np_k}$
0	44	6,484	217,077
1	75	25,935	92,824
2	80	51,870	15,256
3	73	69,160	0,213
4	52	69,160	4,258
5	18	55,328	25,184
6	9	36,885	21,081
7	1	21,077	19,125
8	2	10,539	6,918
	$n = 354$		$\chi^2 = 401,935$

Wartość statystyki leży w obszarze krytycznym, więc musimy odrzucić hipotezę, mówiącą, że zmienna ma rozkład Poissona.

Porównajmy wykres prawdopodobieństwa zmiennej $\lambda = 4$.

k	n_k	$p_{dośw.} = \frac{n_k}{n}$	$p_{teor.} = \frac{\lambda^k}{k!} e^{-\lambda}$
0	44	0,124	0,018
1	75	0,212	0,073
2	80	0,226	0,147

3	73	0,206	0,195
4	52	0,147	0,195
5	18	0,051	0,156
6	9	0,025	0,104
7	1	0,003	0,060
8	2	0,006	0,030



Zadanie 7 (Test χ^2 dobroci dopasowania (test Pearsona) dla rozkładu ciągłego)

Wykonano pomiary kontrolnej próby oporników. Dane zebrano w poniższej tabeli.

193,199	195,673	195,757	196,051	196,092
196,596	196,679	196,763	196,847	197,267
197,392	197,477	198,189	198,65	198,944
199,070	199,111	199,153	199,237	199,698
199,572	199,614	199,824	199,908	200,118
200,160	200,234	200,285	200,453	200,704
200,746	200,830	200,872	200,914	200,956
200,998	200,998	201,123	201,208	201,333
201,375	201,543	201,543	201,584	201,711
201,878	201,919	202,004	202,004	202,088
202,172	202,172	202,297	202,339	202,281
202,507	202,591	202,633	202,716	202,884
203,051	203,052	203,094	203,094	203,177
203,178	203,219	203,764	203,765	203,848
203,890	203,974	204,184	204,267	204,352
204,352	204,729	205,106	205,148	205,231
205,357	205,400	205,483	206,070	206,112
206,154	206,155	206,616	206,665	206,993
207,243	207,621	208,124	208,375	208,502
208,628	208,670	208,711	210,012	211,394

- Korzystając z testu χ^2 zweryfikuj hipotezę zerową: *zmienna losowa x podlega rozkładowi Gaussa*. Przyjmij poziom istotności $\alpha = 0,1$. Podziel zakres zmienności zmiennej losowej x na przedziały o szerokości 2Ω .
- Porównaj wykres doświadczalny prawdopodobieństwa p_k dla przedziałów o szerokości 2Ω z wykresem teoretycznym zgodnym z rozkładem Gaussa. Wskazówka: dla rozkładu teoretycznego prawdopodobieństwo w przedziale policz w przybliżeniu jako $f(x_k) \cdot \Delta x$ z $\Delta x = 2 \Omega$.

W celu wyliczenia χ^2 w punkcie b) skonstruuj tabelę zawierającą

x_k	n_k	Ψ_+	Ψ_-	np_k	$\frac{(n_k - np_k)^2}{np_k}$
193					
195					
...					

gdzie

$$\Psi_{\pm} \equiv \Psi\left(x_k \pm \frac{1}{2}\Delta x\right).$$

Ψ jest dystrybuantą rozkładu normalnego dla parametrów μ i σ (wartości parametrów zastępujemy wartościami ich estymatorów), a $\Delta x = 2 \Omega$.

Rozwiązanie:

ad a)

Statystyką jest

$$X^2 = \sum_{i=1}^r u_i^2 = \sum_{i=1}^r \frac{(n_i - np_i)^2}{np_i}.$$

Podlega ona rozkładowi chi kwadrat o $r - 1$ stopniach swobody. Obszar krytyczny wyznaczamy za pomocą nierówności

$$X^2 > \chi_{r-1, 1-\alpha}^2.$$

Korzystając z estymatorów parametrów μ i σ dostajemy

$$\mu = \frac{1}{n} \sum_{i=0}^n x_i \approx 202,277,$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \approx 3,5295.$$

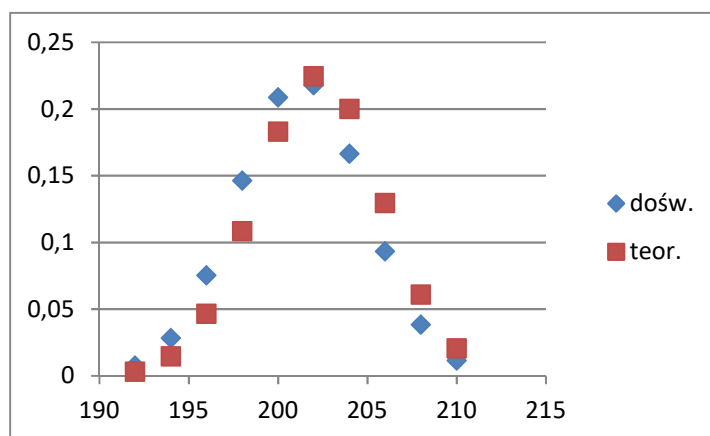
Środki przedziałów	Liczebność w przedziale	brzegi przedziałów		Dystrybunata na brzegach				
x_k	n_k	x_-	x_+	Ψ_+	Ψ_-	$p_k(x_k)$	$n \cdot p_k$	$\frac{(n_k - n \cdot p_k)^2}{n \cdot p_k}$
193	1	192	194	0,00185	0,00968	0,00784	0,78373	0,05968
195	2	194	196	0,00968	0,03808	0,02840	2,84001	0,24846
197	9	196	198	0,03808	0,11346	0,07538	7,53761	0,28372
199	12	198	200	0,11346	0,26002	0,14656	14,65584	0,48128
201	23	200	202	0,26002	0,46882	0,20880	20,87984	0,21528
203	25	202	204	0,46882	0,68680	0,21798	21,79835	0,47025
205	11	204	206	0,68680	0,85356	0,16677	16,67650	1,93222

207	9	206	208	0,85356	0,94705	0,09348	9,34846	0,01299
209	6	208	210	0,94705	0,98544	0,03839	3,83940	1,21586
211	2	210	212	0,98544	0,99699	0,01155	1,15499	0,61823
$n =$		100				$\chi^2 =$		5,53796

Kwantyl rozkładu Chi kwadrat dla $r - 1 = 9$ stopni swobody i prawdopodobieństwa $1 - \alpha = 0,9$ wynosi 14,68. A zatem wartość naszej statystyki leży poza obszarem krytycznym i na poziomie istotności 0,05 możemy przyjąć hipotezę, że zmienna losowa podlega rozkładowi Gaussa.

ad b)

x_i	$p_{dośw.}$	$p_{teor.}$
192	0,007837	0,00333
194	0,028400	0,014641
196	0,075376	0,046783
198	0,146558	0,108628
200	0,208798	0,183287
202	0,217983	0,224731
204	0,166765	0,200233
206	0,093485	0,129642
208	0,038394	0,060996
210	0,01155	0,020854



Zadanie 8 (Test zgodności λ – Kołmogorowa-Smirnowa)

Dla danych z zadania 7 wykonaj test Kołmogorowa-Smirnowa dla poziomu istotności $\alpha = 0,05$ i hipotezy zerowej: *zmienna losowa podlega rozkładowi Gaussa*. Narysuj wykres empirycznej dystrybuanty i nałóż na nią dystrybuantę rozkładu normalnego dla wyliczonych parametrów μ i σ .

Rozwiązanie:

$$D_n^+ = \max_{1 \leq m \leq n} \left(\frac{m}{n} - F_n(x_m^*) \right),$$

$$D_n^- = \max_{1 \leq m \leq n} \left(F_n(x_m^*) - \frac{m-1}{n} \right),$$

$$D_n = \max(D_n^-, D_n^+).$$

Jeśli liczba prób jest co najmniej równa 10 ($n \geq 10$), a poziom istotności $\alpha \geq 0,01$, to kwantyle rozkładu Kołmogorowa $D_{n,1-\alpha}$ możemy wyliczyć z dużą dokładnością z następującego wzoru

$$D_{n,1-\alpha} \approx \sqrt{\frac{1}{2n} \left(y - \frac{2y^2 - 4y - 1}{18n} \right)} - \frac{1}{6n},$$

$$y = -\ln \frac{\alpha}{2}.$$

Przeprowadzając test (prawostronny) Kołmogorowa, hipotezę zerową musimy odrzucić na poziomie istotności α , gdy $D_n > D_{n,1-\alpha}$.

Korzystając z estymatorów parametrów μ i σ dostajemy

$$\mu = \frac{1}{n} \sum_{i=0}^n x_i \approx 202,277,$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \approx 3,5295.$$

Próba losowa	Próba posort.	dystrybuanta		D+	D-
		empir.	teor.		
193,199	193,199	0,01	0,005163	0,004837	0,005163
196,596	195,673	0,02	0,031037	-0,01104	0,021037
197,392	195,757	0,03	0,032735	-0,00273	0,012735
199,070	196,051	0,04	0,039291	0,000709	0,009291
199,572	196,092	0,05	0,040285	0,009715	0,000285
200,160	196,596	0,06	0,054247	0,005753	0,004247
200,746	196,679	0,07	0,056876	0,013124	-0,00312
200,998	196,763	0,08	0,059638	0,020362	-0,01036
201,375	196,847	0,09	0,062504	0,027496	-0,0175
201,878	197,267	0,10	0,078472	0,021528	-0,01153
202,172	197,392	0,11	0,083776	0,026224	-0,01622
202,507	197,477	0,12	0,087534	0,032466	-0,02247
203,051	198,189	0,13	0,124057	0,005943	0,004057
203,178	198,650	0,14	0,152753	-0,01275	0,022753
203,890	198,944	0,15	0,173189	-0,02319	0,033189
204,352	199,070	0,16	0,182458	-0,02246	0,032458
205,357	199,111	0,17	0,185539	-0,01554	0,025539
206,154	199,153	0,18	0,188729	-0,00873	0,018729
207,243	199,237	0,19	0,195210	-0,00521	0,01521
208,628	199,572	0,20	0,222370	-0,02237	0,03237
195,673	199,614	0,21	0,225921	-0,01592	0,025921
196,679	199,698	0,22	0,233119	-0,01312	0,023119
197,477	199,824	0,23	0,244150	-0,01415	0,02415
199,111	199,908	0,24	0,251657	-0,01166	0,021657

199,614	200,118	0,25	0,270946	-0,02095	0,030946
200,234	200,160	0,26	0,274890	-0,01489	0,02489
200,830	200,234	0,27	0,281908	-0,01191	0,021908
200,998	200,285	0,28	0,286795	-0,00679	0,016795
201,543	200,453	0,29	0,303168	-0,01317	0,023168
201,919	200,704	0,30	0,328378	-0,02838	0,038378
202,172	200,746	0,31	0,332678	-0,02268	0,032678
202,591	200,830	0,32	0,341344	-0,02134	0,031344
203,052	200,872	0,33	0,345708	-0,01571	0,025708
203,219	200,914	0,34	0,350094	-0,01009	0,020094
203,974	200,956	0,35	0,354499	-0,0045	0,014499
204,729	200,998	0,36	0,358925	0,001075	0,008925
205,400	200,998	0,37	0,358925	0,011075	-0,00108
206,155	201,123	0,38	0,372205	0,007795	0,002205
207,621	201,208	0,39	0,381324	0,008676	0,001324
208,670	201,333	0,40	0,394854	0,005146	0,004854
195,757	201,375	0,41	0,39943	0,01057	-0,00057
196,763	201,543	0,42	0,417864	0,002136	0,007864
198,189	201,543	0,43	0,417864	0,012136	-0,00214
199,153	201,584	0,44	0,422392	0,017608	-0,00761
199,824	201,711	0,45	0,436481	0,013519	-0,00352
200,285	201,878	0,46	0,455128	0,004872	0,005128
200,872	201,919	0,47	0,459723	0,010277	-0,00028
201,123	202,004	0,48	0,469265	0,010735	-0,00073
201,543	202,004	0,49	0,469265	0,020735	-0,01073
202,004	202,088	0,50	0,478712	0,021288	-0,01129
202,297	202,172	0,51	0,488172	0,021828	-0,01183
202,633	202,172	0,52	0,488172	0,031828	-0,02183
203,094	202,281	0,53	0,500455	0,029545	-0,01954
203,764	202,297	0,54	0,502259	0,037741	-0,02774
204,184	202,339	0,55	0,506992	0,043008	-0,03301
205,106	202,507	0,56	0,525910	0,034090	-0,02409
205,483	202,591	0,57	0,535349	0,034651	-0,02465
206,616	202,633	0,58	0,540062	0,039938	-0,02994
208,124	202,716	0,59	0,549358	0,040642	-0,03064
208,711	202,884	0,60	0,568086	0,031914	-0,02191
196,051	203,051	0,61	0,586552	0,023448	-0,01345
196,847	203,052	0,62	0,586662	0,033338	-0,02334
198,650	203,094	0,63	0,591277	0,038723	-0,02872
199,237	203,094	0,64	0,591277	0,048723	-0,03872
199,908	203,177	0,65	0,600361	0,049639	-0,03964
200,453	203,178	0,66	0,60047	0,05953	-0,04953
200,914	203,219	0,67	0,604937	0,065063	-0,05506
201,208	203,764	0,68	0,662802	0,017198	-0,0072
201,584	203,765	0,69	0,662905	0,027095	-0,01709
202,004	203,848	0,70	0,671426	0,028574	-0,01857

202,339	203,890	0,71	0,675704	0,034296	-0,0243
202,716	203,974	0,72	0,684191	0,035809	-0,02581
203,094	204,184	0,73	0,704981	0,025019	-0,01502
203,765	204,267	0,74	0,713021	0,026979	-0,01698
204,267	204,352	0,75	0,721145	0,028855	-0,01886
205,148	204,352	0,76	0,721145	0,038855	-0,02886
206,070	204,729	0,77	0,755769	0,014231	-0,00423
206,665	205,106	0,78	0,787932	-0,00793	0,017932
208,375	205,148	0,79	0,791355	-0,00136	0,011355
210,012	205,231	0,80	0,798024	0,001976	0,008024
196,092	205,357	0,81	0,807899	0,002101	0,007899
197,267	205,400	0,82	0,811201	0,008799	0,001201
198,944	205,483	0,83	0,817473	0,012527	-0,00253
199,698	206,070	0,84	0,858054	-0,01805	0,028054
200,118	206,112	0,85	0,860703	-0,0107	0,020703
200,704	206,154	0,86	0,863319	-0,00332	0,013319
200,956	206,155	0,87	0,863380	0,00662	0,00338
201,333	206,616	0,88	0,889877	-0,00988	0,019877
201,711	206,665	0,89	0,892461	-0,00246	0,012461
202,088	206,993	0,90	0,908634	-0,00863	0,018634
202,281	207,243	0,91	0,919694	-0,00969	0,019694
202,884	207,621	0,92	0,934454	-0,01445	0,024454
203,177	208,124	0,93	0,950725	-0,02072	0,030725
203,848	208,375	0,94	0,957541	-0,01754	0,027541
204,352	208,502	0,95	0,960687	-0,01069	0,020687
205,231	208,628	0,96	0,963618	-0,00362	0,013618
206,112	208,670	0,97	0,964555	0,005445	0,004555
206,993	208,711	0,98	0,965450	0,01455	-0,00455
208,502	210,012	0,99	0,985567	0,004433	0,005567
211,394	211,394	1,00	0,994999	0,005001	0,004999

$$D_n^+ = \max_{1 \leq m \leq n} \left(\frac{m}{n} - F_n(x_m^*) \right) = 0,065063,$$

$$D_n^- = \max_{1 \leq m \leq n} \left(F_n(x_m^*) - \frac{m-1}{n} \right) = 0,038378,$$

$$D_n = \max(D_n^-, D_n^+) = 0,065063.$$

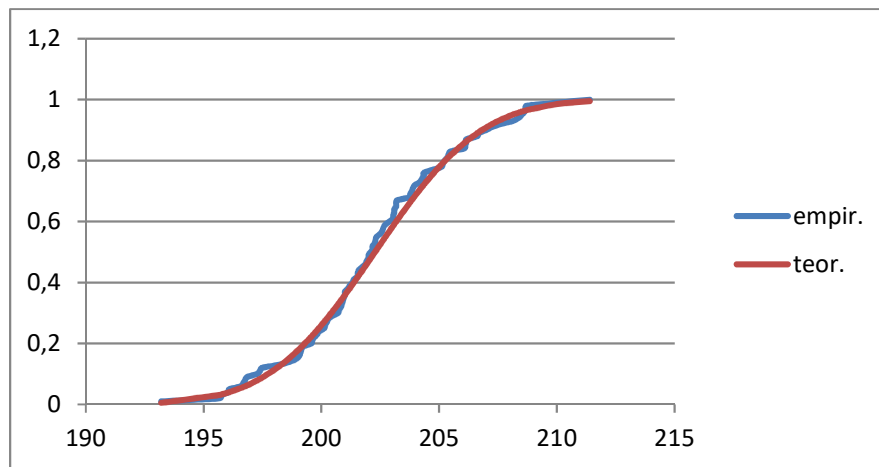
Liczba prób jest większa niż 10 ($n = 100$), a poziom istotności $\alpha = 0,05 \geq 0,01$, więc kwantyl rozkładu Kołmogorowa $D_{n,1-\alpha}$ możemy wyliczyć z dużą dokładnością z następującego wzoru

$$D_{n,1-\alpha} \approx \sqrt{\frac{1}{2n} \left(y - \frac{2y^2 - 4y - 1}{18n} \right)} - \frac{1}{6n},$$

$$y = -\ln \frac{\alpha}{2} \approx 3,6889.$$

$$D_{n,1-\alpha} \approx 0,134.$$

Wartość statystyki leży poza obszarem krytycznym, więc na poziomie istotności 0,05 możemy przyjąć hipotezę, że zmienna podlega rozkładowi Gaussa. Na poniższym wykresie pokazano wykres teoretycznej i empirycznej dystrybuanty tego rozkładu.



Zadanie 9 (Test znaków)

Wykonano 10 niezależnych pomiarów zmiennej X i 10 niezależnych pomiarów zmiennej Y . Wyniki pomiarów w kolejności ich otrzymywania zebrano w poniższej tabeli

X	7,48	7,59	7,81	6,99	9,64	8,84	7,77	6,69	7,16	6,76
Y	5,34	8,21	7,68	8,44	5,22	7,28	6,46	8,44	6,36	6,38

Stosując test znaków na poziomie istotności $\alpha = 0,1$ zweryfikuj hipotezę: *rozkłady obu zmiennych są jednakowe*.

Rozwiązanie

X	7,48	7,59	7,81	6,99	9,64	8,84	7,77	6,69	7,16	6,76
Y	5,34	8,21	7,68	8,44	5,22	7,28	6,46	8,44	6,36	6,38
	+	-	+	-	+	+	+	-	+	+

Liczba znaków + wynosi $k_+ = 7$.

Dla testu lewostronnego szukamy liczby k_l , dla której

$$P(k \leq k_l) = \sum_{i=1}^{k_l} \binom{n}{i} \left(\frac{1}{2}\right)^{n-i} \left(\frac{1}{2}\right)^i = F(k_l) = \alpha.$$

Hipotezę odrzucamy, gdy $k_+ < k_l$.

Dla testu lewostronnego szukamy liczby k_p , dla której

$$P(k \geq k_p) = \sum_{i=k_p}^n \binom{n}{i} \left(\frac{1}{2}\right)^{n-i} \left(\frac{1}{2}\right)^i = 1 - F(k_l) + P(k_p) = \alpha.$$

Hipotezę odrzucamy, gdy $k_+ > k_p$.

W przypadku testu dwustronnego

$$P(k \leq k_l) = \sum_{i=1}^{k_l} \binom{n}{i} \left(\frac{1}{2}\right)^{n-i} \left(\frac{1}{2}\right)^i = F(k_l) = \frac{\alpha}{2},$$

$$P(k \geq k_p) = \sum_{i=k_p}^n \binom{n}{i} \left(\frac{1}{2}\right)^{n-i} \left(\frac{1}{2}\right)^i = 1 - F(k_l) + P(k_p) = \frac{\alpha}{2}.$$

Hipotezę odrzucamy, gdy $k_+ < k_l$ lub $k_+ > k_p$.

Tabele, z których możemy odszukać k_l i k_p są następujące

k_l	$P(k \leq k_l) = F(k_l)$	k_p	$P(k \geq k_p) = 1 - F(k_p) + P(k_p)$
0	0,000977	0	1
1	0,010742	1	0,999023
2	0,054688	2	0,989258
3	0,171875	3	0,945313
4	0,376953	4	0,828125
5	0,623047	5	0,623047
6	0,828125	6	0,376953
7	0,945313	7	0,171875
8	0,989258	8	0,054688
9	0,999023	9	0,010742
10	1	10	0,000977

Zastosujemy test dwustronny na poziomie istotności $\alpha = 0,1$. Znajdujemy $k_l \approx 2$ i $k_p \approx 8$. Nasze $k_+ = 7$, czyli $k_l < k_+ < k_p$. A zatem na poziomie istotności 0,1 nie możemy odrzucić hipotezy mówiącej, że rozkłady obu zmiennych są takie same.