Lab 2

(1) Prawdopodobieństwo dziedziczenia pewnej cechy wsród potomków badanego gatunku drapieżnego kota wynosi 0,7. Jakie jest prawdopodobieństwo, że

(a) wśród pięciu potomków co najwyżej trzy koty mają badaną cechę?

(b) wśród sześciu potomków co najmniej cztery koty mają badaną cechę?

(c) wśród siedmiu potomków dokładnie pięć kotów ma badaną cechę?

(2) Prawdopodobieństwo, że produkt poddawany próbie wytrzymałościowej nie wytrzyma tej próby, wynosi $p = 0,02$. Wyznaczyć prawdopodobieństwo, że wśród losowo wybranych 20 takich produktów co najmniej 2 produkty nie wytrzymały próby.

(3) Zakładając, że wzrost w populacji studentów ma rozkład $N(176,\ 10)$ obliczyć prawdopodobieństwo tego, że wzrost przypadkowo napotkanego studenta

(a) jest mniejszy niż 186 cm

(b) jest mniejszy niż 166 cm

(c) jest większy niż 170 cm

(d) jest wiekszy niż 200 cm

(e) należy do przedziału $(168,\ 174)$

(4) Zakładając, że wyniki w teście egzaminacyjnym (o zakresie punktacji od 0 do 100) ze statystyki mają rozkład normalny $N(58, 10)$ obliczyć:

(a) jaki procent studentów uzyska wynik większy niż 75 punktów,

(b) jaki procent studentów uzyska wynik mniejszy niż 50 punktów,

(c) poniżej jakiego wyniki student będzie zaliczony do grupy 5% najsłabszych studentów,

(d) jaki wynik muszą osiągnąć studenci, aby mogli być zaliczeni do grupy 5% najlepszych studentów.

(5) Stwierdzono, że przeciętny czas pracy drukarek w pewnej firmie ma rozkład normalny z wartością oczekiwaną wynoszącą 3 lata i odchyleniem standardowym równym 5 miesięcy. Obliczyć prawdopodobieństwo, ze drukarka będzie pracować

(a) krócej niż 2 lata,

(b) dłużej niż 3 lata.

Wprowadzenie:

I.Rozkład dwumianowy (Bernoulliego) -> Analiza mocy tekstu

<= - normlanie | > - 1-p

1.Co najwyżej P(X<=x)

2.Co najmniej P(X>=x), np co najmniej 4 to więcej niż 3

3.Dokładnie P(X=x), np dokładnie 5 to P(X<=5)-P(X<=4)

II.Rozkład normalny -> Kalkulator prawdopodobieństwa

< - normlanie | > - 1-p

N(x,y) x - średnia, y - odchylenie standardowe

Zad. 1

(a) P(X<=3)=0,47178

(b) P(X>=4)=0,74431

(c) P(X=5)=0,31765

Zad. 2

P(X>=2)=0,05(98)

Zad. 3

(a) Z(176,10) = 186   p = **0,841345**

(b) Z(176,10) = 166   p = **0,158655**

(c) Z(176,10) = 170   p = **0,725747**

(d) Z(176,10) = 200   p = **0,008198**

(e) P(168<X<174) = P(X<174) - P(X<=168) =

= 0,42074-0,211855 = **0,208885**

* Z(176,10) = 174   p = 0,42074
* Z(176,10) = 168   p = 0,211855

Zad. 4

(a) Z(58,10) = 75   p = 0,044565 = **4,5%**

(b) Z(58,10) = 50   p = 0,211855 = **21,2%**

(c) Z(58,10) = **41,5515**   p = 0,05

(d) Z(58,10) = **74,4485**   p = 0,05

Zad. 5

Z(36,5) = 24   p = **0,008198**

Z(36,5) = 36   p = **0,5**

Lab 3/4

(1) Przeprowadzono pomiary dobowych dawek promieniowania nadfioletowego (w MED) w 32 miejscowościach południowej Polski. Obserwacje prowadzono w czerwcu. Uzyskano następujące średnie miesięczne wartości dobowych dawek w tych miejscowościach: 14, 15, 13, 12, 21, 15, 17, 20, 7, 12, 5, 9, 12, 13, 19, 4, 10, 11, 14, 9, 12, 11, 15, 14, 17, 13, 15, 16, 12, 14, 17, 14. Utworzyć szereg rozdzielczy przedziałowy i histogram, wyznaczyć podstawowe miary statystyczne oraz podać ich interpretację, sporządzić wykres ramka-wąsy.

(2) Czas mocowania detalu toczonego na obrabiarce ma rozkład normalny. Zmierzono czasy mocowania dla $n = 10$ wylosowanych niezależnie robotników i otrzymano następujące wyniki (w sekundach): 10, 20, 16, 20, 18, 30, 24, 20, 17, 25. Oszacować metodą przedziałową przy współczynniku ufności 0,95 średni czas potrzebny na zamocowanie tego detalu na obrabiarce.

(3) Pracochłonność 6 losowo wybranych detali (w minutach) kształtowała się następująco: $6,3$; $5,9$; $6,2$; $5,8$; $5,7$; $6,1$. Przyjmując współczynnik ufnosci 0,90 zbudować przedział ufności określający zróżnicowanie pracochłonności w całej populacji produkowanych detali. Zakładamy, że rozkład badanej cechy jest normalny.

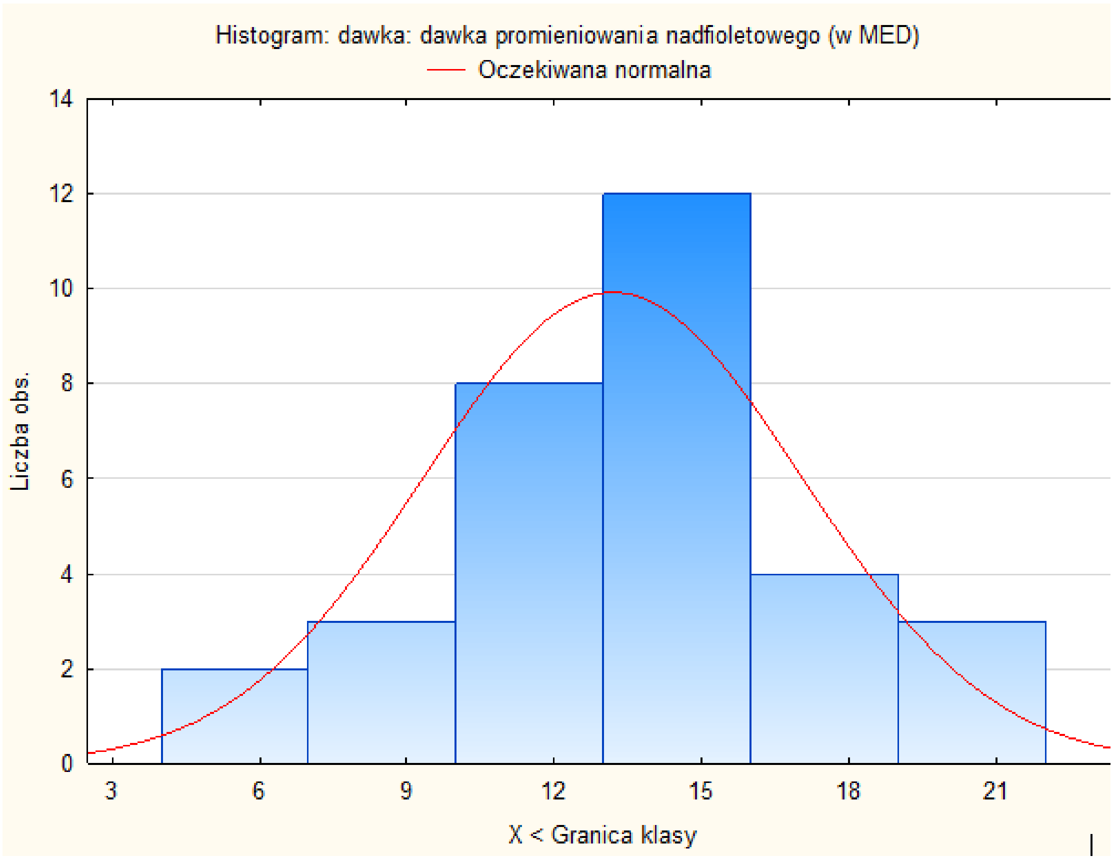
(4) W pewnym doświadczeniu fizycznym mierzy się czas występowania pewnego efektu świetlnego. Przeprowadzono $n = 1000$ niezależnych doświadczeń nad tym efektem i zbiór pogrupowanych wyników (w sekundach) jest następujący:

czas efektu	liczba doświadczeń
0,0-0,2	50
0,2-0,4	128
0,4-0,6	245
0,6-0,8	286
0,8-1,0	134
1,0-1,2	90
1,2-1,4	67

Przyjmując współczynnik ufności 0,95 oszacować metodą przedziałową średni czas trwania badanego efektu świetlnego.

Zad.1

	Tabela liczności: dawka: dawka promieniowania nadfioletowego (w MED) (zad1 (1))			
Od Do	Liczba	Skumulow. Liczba	Procent	Skumulow. Procent
4 <=x<7	2	2	6,25000	6,2500
7 <=x<10	3	5	9,37500	15,6250
10 <=x<13	8	13	25,00000	40,6250
13 <=x<16	12	25	37,50000	78,1250
16 <=x<19	4	29	12,50000	90,6250
19 <=x<22	3	32	9,37500	100,0000
Braki	0	32	0,00000	100,0000



Zmienna	Statystyki opisowe (zad1 (1))											
	Nważnych	Średnia	Mediana	Moda	Liczność Mody	Minimum	Maksimum	Dolny Kwartyl.	Górny Kwartyl.	Odch.std	Wsp.zmn.	Skośność
dawka	32	13,18750	13,50000	Wielokr.	5	4,000000	21,00000	11,50000	15,00000	3,855767	29,23804	-0,339415

Średnia dobowa dwaka promieniowania wynosi 13,2 MED.

Współczynnik zemienności 29,2% - odchylenie standardowe stanowi 29,2% wartości średniej.

Mediana 13,5: w co najmniej 50% miejscowości dawka promieniowa wynosi co najwyżej 13,5 MED (w co najmniej 50% miejscowości dawka promieniowania wynosi co najmniej 13,5 MED).

Kwartyl dolny 11,5: w co najmniej 25% miejscowości dawka promieniiowa wynosi co najwyżej 11,4 MED(i w co najmniej 75% mejscowości dawka promieniowania wynosi co najmniej 11,5 MED)

Kwartyl górny 15: w co najmniej 75% mejscowości dawka promieniowa wynosi co najwyżej 15 MED (i w co najmniej 25% miejscowości dawka promieniowania wynosi co najmniej 15 MED)

Współczynnik skośności -0.34
Jest asymetria lewostronna: w więcej niż 50% miejscowości dawka promieniownia przekracza średnią.

Miary statystyczne

Miary położenia - do określenia tej wartości zmiennej, wokół której skupiają się wszystkie pozostałe wartości

- *Wartość średnia* - średnia arytmetyczna
- *Mediana* - wartość środkowa w uporządkowanej niemalejąco próbie (co najmniej 50% jednostek ma wartosci cechy równe medianie lub niższe od niej, a co najmniej 50% - równe medianie lub wyższe od niej)
- *Moda (dominanta, wartość modalna)* - najcześciej powtarzająca się wartość
- *Kwartyl dolny (pierwszy)* - co najmniej 25% jednostek ma wartości cechy równe kwartylowi pierwszemu lub niższe od niego, a co najmniej 75% - równe mu lub wyższe.
- *Kwartyl górny (trzeci)* - co najmniej 75% jednostek ma wartosci cechy równe kwartylowi trzeciemu lub niższe od niego, a co najmniej 25% - równe mu lub wyższe.

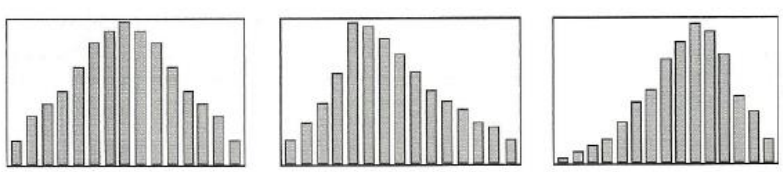
Miary rozproszenia (zmiennosci, zróżnicowania) - do badania stopnia zróżnicowania wartosci zmiennej

- *Rozstęp*
- *Rozstęp kwartylowy* - różnica między kwartylem trzecim i pierwszym (długość przedziału, w którym mieści się 50% środkowych obserwacji)
- *Wariancja*: $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$, $\hat{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
- *Odchylenie standardowe*: $s = \sqrt{s^2}$ - informuje o przeciętnym odchyleniu wartości cechy od średniej arytmetycznej
- *Współczynnik zmienności*: $V_s = \frac{s}{\bar{x}} \cdot 100\%$ - określa, jaki procent wartości średniej stanowi odchylenie standardowe, im większy współczynnik zmienności, tym zbiorowość jest mniej jednorodna.

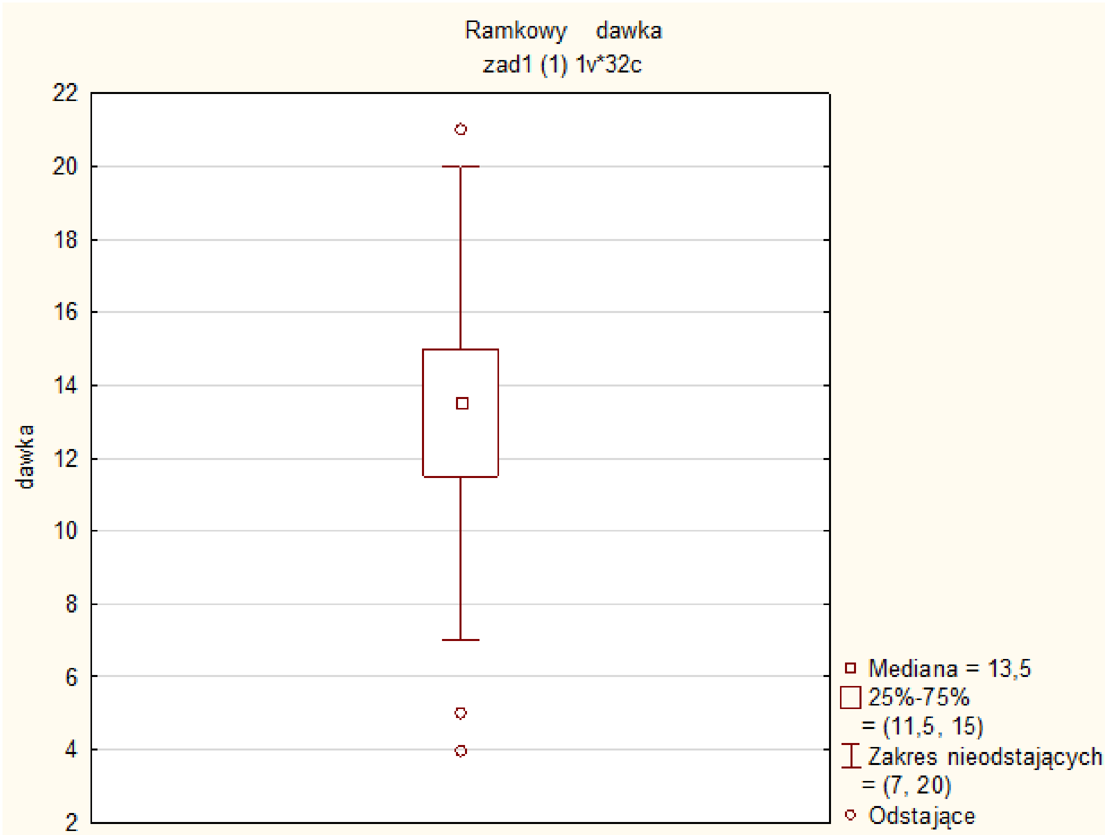
Miary asymetrii - do badania kierunku zróżnicowania wartości zmiennej

Współczynnik skośności (asymetrii) A_s

- $A_s = 0$ - rozkład symetryczny
- A_s wyraźnie różny od 0 - rozkład asymetryczny
  $A_s < 0$ asymetria lewostronna
  $A_s > 0$ asymetria prawostronna



wartość średnia = moda = mediana — rozkład symetryczny | moda < mediana < wartość średnia — asymetria prawostronna | wartość średnia < mediana < moda — asymetria lewostronna



Zad2

Zmienna	Statystyki opisowe (Arkusz1) Nważnych	Średnia	Ufność -95,000%	Ufność 95,000%	Minimum	Maksimum	Odch.std
Zmn1	10	20,00000	16,08183	23,91817	10,00000	30,00000	5,477226

P(16,08< m < 23,92)=0,95
Przedział(16,1;23,9) z prawdopodobieństwem równym 0,95 pokrywa nieznaną wartość średnią czasu mocowania detlau na obrabiarce.

Zad3

Zmienna	Statystyki opisowe (Arkusz1) Średnia	Wariancja	Odch.std	P. ufności odch. std. -90,000%	P. ufności odch. std. +90,000%
Zmn2	6,000000	0,056000	0,236643	0,159036	0,494409

Przedział (0,16;0,49) z prawdopodobieństwem rówym 0,90 pokrywa nieznaną wartość odchylenia standardowego pracochłonności detali.

Zad.4

czas efektu		środek przedziału x	liczba doświadczeń n	xi*ni	(xi-średnia)^2*ni
		xi	ni		
0	0,2	0,1	50	5	16,404992
0,2	0,4	0,3	128	38,4	17,78941952
0,4	0,6	0,5	245	122,5	7,3156608
0,6	0,8	0,7	286	200,2	0,21159424
0,8	1	0,9	134	120,6	6,91705856
1	1,2	1,1	90	99	16,4249856
1,2	1,4	1,3	67	87,1	26,35644928
alfa	0,05				
n	1000				
średnia	0,6728				
wariancja	0,09151167				
odch. Stand.	0,30250896				
u_alfa	1,95996398				
a.lewo	0,65405064				
b.prawo	0,69154936	Przedział(0,65;0,69) z prawdopodobieństwem równym 0,95			
		pokrywa nieznaną wartość srednią czasu występowania efektu świetlnego			

Przedziały ufności dla średniej

- cecha X ma w populacji generalnej rozkład $N(m, \sigma)$, m, σ - nieznane

$$P\left(\overline{x} - t_{\alpha,n-1} \cdot \frac{s}{\sqrt{n-1}} < m < \overline{x} + t_{\alpha,n-1} \cdot \frac{s}{\sqrt{n-1}}\right) = 1 - \alpha$$

$$P\left(\overline{x} - t_{\alpha,n-1} \cdot \frac{\hat{s}}{\sqrt{n}} < m < \overline{x} + t_{\alpha,n-1} \cdot \frac{\hat{s}}{\sqrt{n}}\right) = 1 - \alpha$$

$t_\alpha : \ P(|t| > t_{\alpha,n-1}) = \alpha,$

t ma rozkład t-Studenta o $n-1$ stopniach swobody $\quad (t_{\alpha,n-1} = \textbf{rozkł.t.odwr}(1 - \frac{\alpha}{2}; n-1))$

- cecha X ma w populacji generalnej rozkład dowolny, n - duże

$$P\left(\overline{x} - u_\alpha \cdot \frac{\hat{s}}{\sqrt{n}} < m < \overline{x} + u_\alpha \cdot \frac{\hat{s}}{\sqrt{n}}\right) = 1 - \alpha$$

$u_\alpha : \ P(|U| > u_\alpha) = \alpha, \quad U \sim N(0,1) \quad (u_\alpha = \textbf{rozkł.normalny.s.odwr}(1 - \frac{\alpha}{2}))$

Przedziały ufności dla wariancji i odchylenia standardowego

- cecha X ma w populacji generalnej rozkład $N(m, \sigma)$

$$P\left(\frac{ns^2}{\chi^2_{\frac{\alpha}{2},n-1}} < \sigma^2 < \frac{ns^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right) = 1 - \alpha$$

$P(\chi^2 < \chi^2_{1-\frac{\alpha}{2},n-1}) = P(\chi^2 > \chi^2_{\frac{\alpha}{2},n-1}) = \frac{\alpha}{2}$.

χ^2 ma rozklad chi-kwadrat o $n - 1$ stopniach swobody

($\chi^2_{\frac{\alpha}{2},n-1} =$ **rozkł.chi.odwr.ps**$(\frac{\alpha}{2}; n-1)$, $\chi^2_{1-\frac{\alpha}{2},n-1} =$ **rozkł.chi.odwr.ps**$(1-\frac{\alpha}{2}; n-1)$)

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (dla szeregu rozdzielczego $\bar{x} = \frac{1}{n}\sum \dot{x}_i \cdot n_i$)

$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ (dla szeregu rozdzielczego $s^2 = \frac{1}{n}\sum(\dot{x}_i - \bar{x})^2 \cdot n_i$)

$\hat{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ (dla szeregu rozdzielczego $\hat{s}^2 = \frac{1}{n-1}\sum(\dot{x}_i - \bar{x})^2 \cdot n_i$)

Lab 5

(1) Na plantacji agrestu pobrano w sposób losowy 26 owoców agrestu a następnie zważono je.Odnotowano nastepujace wyniki (w gramach):

$$3,5;\ \ 5,2;\ \ 4,8;\ \ 4,6;\ \ 4,4;\ \ 4,5;\ \ 6,2;\ \ 5,8;\ \ 4,2;\ \ 3,8;\ \ 6,0;\ \ 7,0,\ \ 4,2;$$

$$4,8;\ \ 5,3;\ \ 5,2;\ \ 6,8;\ \ 7,0;\ \ 7,2;\ \ 6,8;\ \ 4,5;\ \ 5,5;\ \ 4,2;\ \ 3,5;\ \ 5,5;\ \ 6,5.$$

Sprawdzić, czy na poziomie istotności $\alpha = 0,05$ można twierdzić, że waga owoców agrestu ma rozkład normalny.

(2) W pewnym biochemicznym doświadczeniu bada się czas życia pewnych żywych komórek w pewnym środowisku. Dokonano 8 pomiarów i otrzymano następujące czasy życia tych komórek w badanym środowisku (w godz):

$$4,7,\ \ 5,3,\ \ 4,0,\ \ 3,8,\ \ 6,2,\ \ 5,5,\ \ 4,5,\ \ 6,0.$$

Przyjmując poziom istotności $\alpha = 0,05$ zweryfikować hipotezę, że średni czas życia komórek w tym środowisku jest większy niż 4 godziny.

(3) Spośród dzieci uczęszczających do przedszkoli wylosowano dwie próby złożone z 10 chłopców oraz 8 dziewcząt. Następnie dokonano pomiaru czasu ich snu po obiedzie. Dla chłopców otrzymano następujące wyniki (w minutach):

$$65,\ \ 40,\ \ 60,\ \ 72,\ \ 38,\ \ 45,\ \ 55,\ \ 75,\ \ 80,\ \ 30,$$

analogicznie dla dziewcząt:

$$65,\ \ 70,\ \ 60,\ \ 40,\ \ 80,\ \ 85,\ \ 82,\ \ 78.$$

Na poziomie istotności $\alpha = 0,05$ zweryfikować hipotezę o jednakowym czasie snu w obu grupach.

(4) W jednym z pomieszczeń fabrycznych temperatura powinna utrzymywac się na poziomie 4^oC. W ciągu roku dokonano 60 pomiarów temperatury w tym pomieszczeniu, na podstawie których stwierdzono, że średnia temperatura wynosiła $3,74^oC$, a odchylenie standardowe $s = 1,4^oC$. Czy można uznać, że średnia temperatura w tym pomieszczeniu spełnia normę? Przyjąć poziom istotności $\alpha = 0,05$. Zakładamy, że badana zmienna ma rozkład normalny.

(5) W wylosowanej próbie 100 pracownic dużego zakładu przemysłowego, średni czas przebywania ich na zwolnieniach lekarskich w ciągu roku wyniósł $\bar{x} = 38$ dni, a odchylenie standardowe $s = 16$ dni. Czy można na tej podstawie twierdzić, że średni roczny czas zwolnień lekarskich dla pracownic tego zakładu jest dłuższy niż 31 dni? Przyjąć poziom istotności $\alpha = 0,01$. Zakładamy, że badana zmienna ma rozkład normalny.

(6) Spośród studentów pewnego wydziału wylosowano niezależnie 10 studentów IV roku i otrzymano dla nich następujące średnie oceny uzyskane w sesji egzaminacyjnej na I roku studiów (x_i) oraz na IV roku studiów (y_i):

x_i	3,5	4,0	3,8	4,6	3,9	3,0	3,5	3,9	4,5	4,1
y_i	4,2	3,9	3,8	4,5	4,2	3,4	3,8	3,9	4,6	4,0

- Zbadać, czy średnia ocen na IV roku zależy od średniej ocen na I roku, przeanalizować istotność tej zależności.
- Znaleźć oszacowanie liniowej funkcji regresji cechy Y względem cechy X.

!= p<alfa odrzucamy H0 na H1

p>alfa H0 zostaje (obustronne)

> lub < to samo tylko p/2(jednostronne)

Przy badaniu hipotezy:

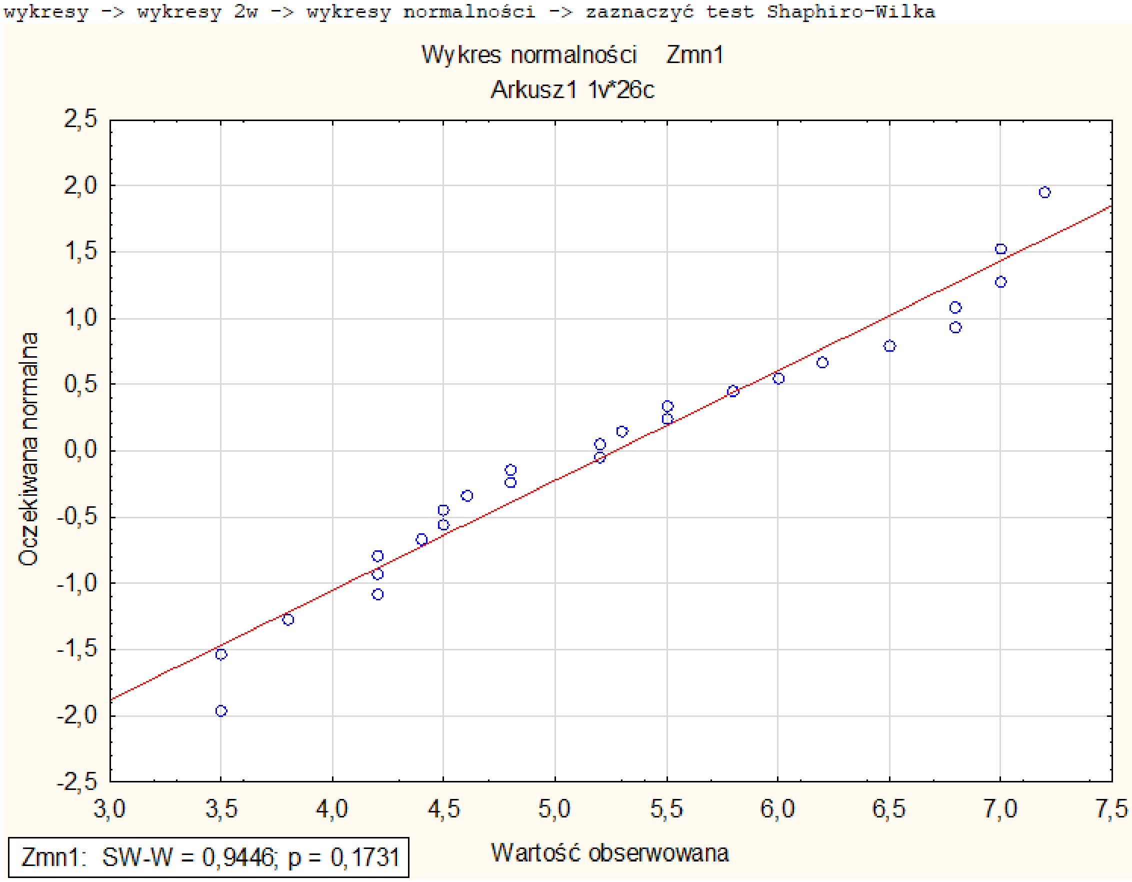
1. badamy rozkład normalny -> wykres normalności -> test wilka

2. próby niezależne to badamy wariancje -> statystyki podstawowe i tabele -> Test t dla (zależy co badamy) -> Testy jednorodności wariancji

3. badamy hipotezę p

Zad.1

wykresy -> wykresy 2w -> wykresy normalności -> zaznaczyć test Shaphiro-Wilka



Testy t-Studenta w programie Statistica

- test t dla pojedyńczej próby

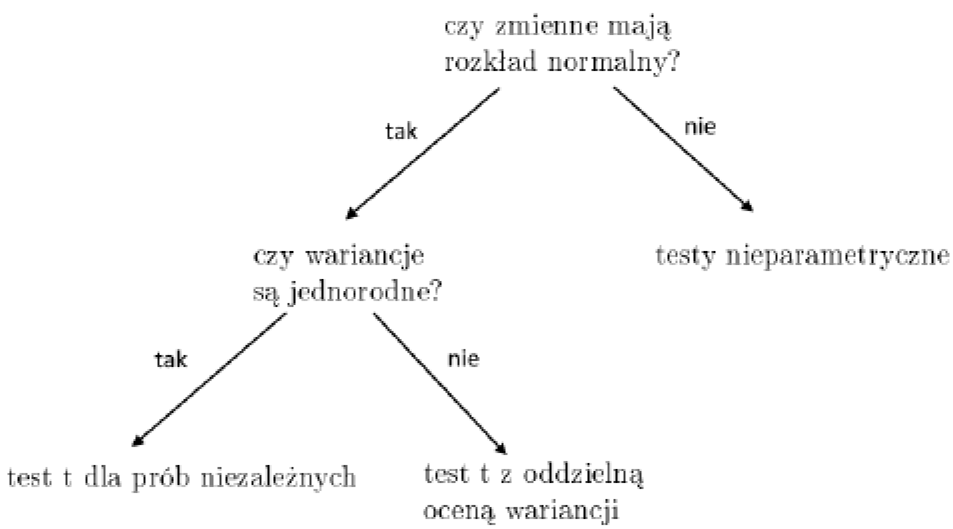
  $H_0: \ m = m_0$ (m_0 - wartość hipotetyczna)
  $H_1: \ m \neq m_0$ lub $m < m_0$ lub $m > m_0$

  założenie: normalność rozkładu badanej zmiennej

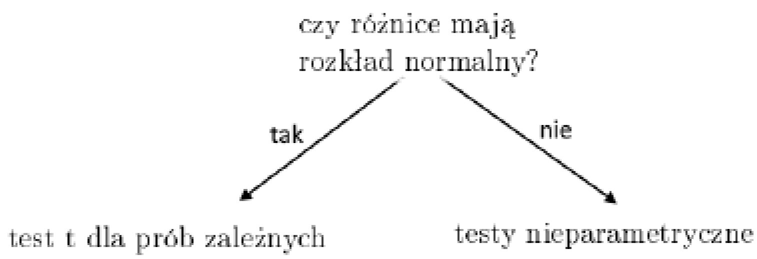
- test t dla dwóch prób niezależnych

  $H_0: \ m_1 = m_2$
  $H_1: \ m_1 \neq m_2$ lub $m_1 < m_2$ lub $m_1 > m_2$

  założenia:
  - normalność rozkładu
  - jednorodność wariancji



- test t dla dwóch prób zależnych
  $H_0: \ m_1 = m_2$
  $H_1: \ m_1 \neq m_2$ lub $m_1 < m_2$ lub $m_1 > m_2$
  założenie: normalność rozkładu różnic



Zad.2

1)
H0 - średni czas życia komórek jest równy 4
H1 - średni czas życia komówek jest większy od 4

2)
robimy rozkład t-studenta bo takie sie robi dla hipotez opartych na średnich

3)

Żeby móc zrobić testy t-studenta trzeba sprawdzić czy badana zmienna ma rozkład normalny:

{
H0 - ma rozkład normalny
H1 - nie ma rozkładu normalnego
wykres robimy jak w poprzednim zadaniu i otrzymujemy:

p > alfa więc nie ma podstaw do odrzucenia H0(tej z rozkładem normalnym) - na poziomie istotniości 0.05 można przyjąć, że rozkład jest normalny
}

4)
Po sprawdzeniu że rozkład długości życia komórek jest normalny możemy zrobić testy t-studenta :

Statystyka -> statystyki podstawowe i tabele -> Test t dla pojedynczej próby -> w polu testuj średnie względem wpisać liczbę względem której robimy test (w tym zadaniu 4)

	Test średnich względem stałej wartości odniesienia (Arkusz3)							
Zmienna	Średnia	Odch.st.	Ważnych	Bł. std.	Odniesienie Stała	t	df	p
Zmn1	5,000000	0,891227	8	0,315096	4,000000	3,173632	7	0,015629

5)
Odrzucamy H0 na rzecz H1, bo p/2 < alfa
(Średni czas życia komórek jest większy niż 4h)

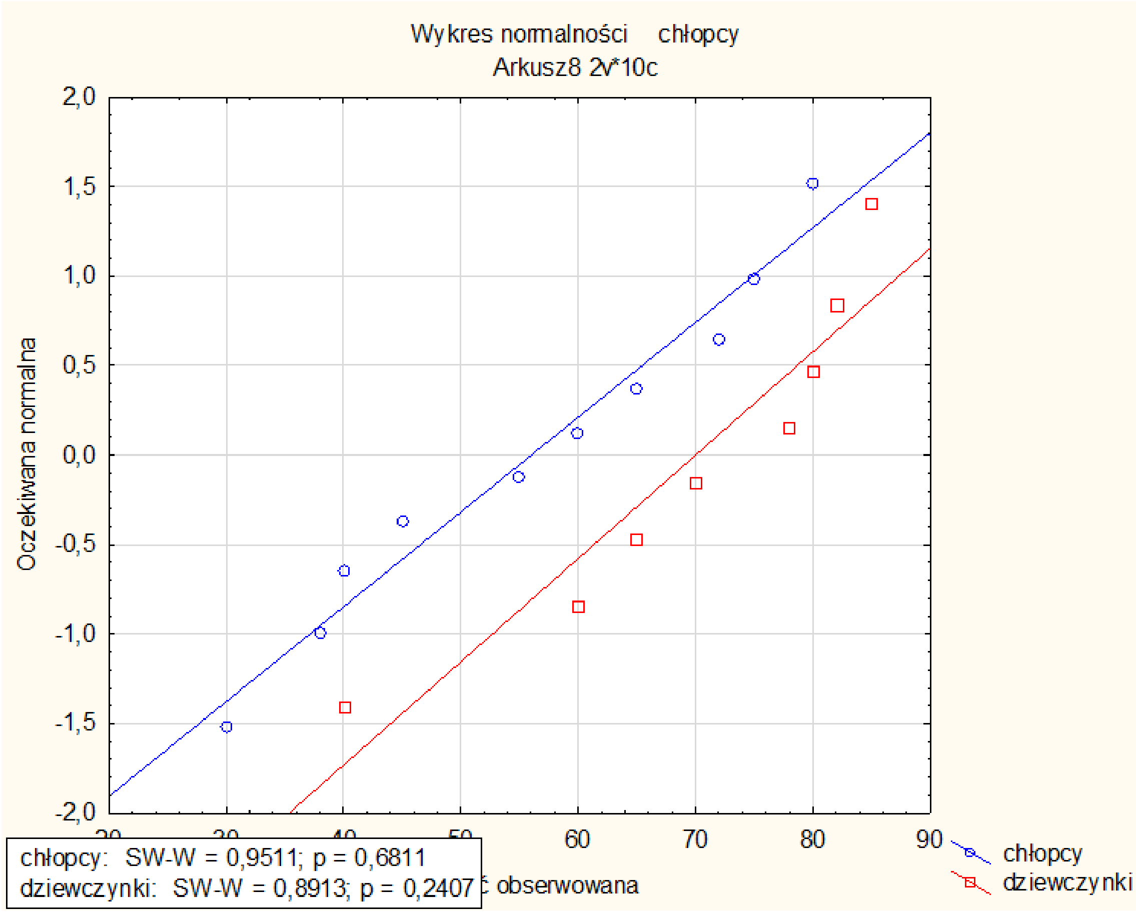
Zad.3

1)
H0 - czas snu w obu grupach jest taki sam (porównywalny) -> m1 = m2
H1 - czas snu w obu grupach jest różny -> m1!=m2

2) Sprawdzamy rozkłady normalne
H0c - ma rozkład normalny (chłopcy)

H1c - nie ma rozkładu normalnego (chłopcy)

H0d - ma rozkład normalny (dziewczynki)
H1d - nie ma rozkładu normalnego (dziewczynki)



pch > alfa -> nie ma podstaw do odrzucenia H0c
pdz > alfa -> nie ma podstaw do odrzucenia H0d

W obu przypadkach mamy rozkłady normalne.

3)
testy dla grup niezależnych
założenie: rozkłady zmiennych są normalne
równość wariancji
wtedy test t dla prób niezależnych
jeżeli wariancje są różne -> test t z uwzględnieniem wariancji

4) Sprawdzamy jednorodność wariancji
Statystyka -> statystyki podstawowe i tabele -> Test t dla prób niezależnych wzgl. zmiennych -> w karcie opcje zaznaczemy test levena i tego drugiego żeby dostać wariancje

	Testy dla prób niezależnych (Arkusz7) Uwaga: Zmienne traktowane są jako niezależne próby.						
Grupa 1 wz. Grupy 2	Średnia Grupa 1	Średnia Grupa 2	t	df	p	Nważnych Grupa 1	Nważnych Grupa 2
Chłopcy vs. Dupy	56,0000(	70,0000(	-1,81607	16	0,08814	10	8

iloraz F Wariancje	p Wariancje	Levene'a F(1,df)	df Levene'a	p Levene'a	Brn-Fors F(1,df)	df Brn-Fors	p Brn-Fors
1,331907	0,721336	0,619002	16	0,442917	0,546807	16	0,470336

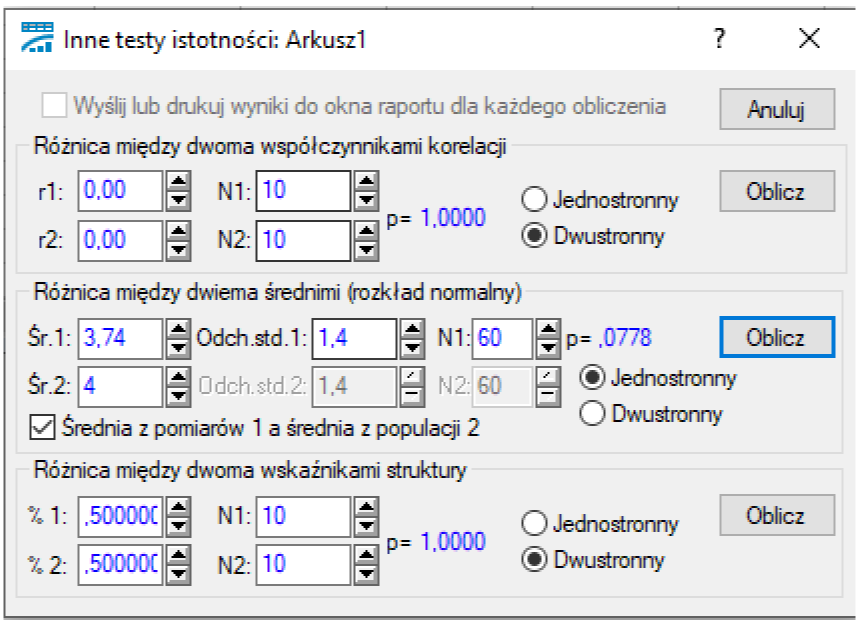
Dla testu Browna-Forsytha p=0.47
dla testu Levene'a p=0.44
dla testu F p=0.72;
we wszystkich testach p>alfa zatem nie ma podstaw do odrzucenia hipotezy o równości wariancji

5) Teraz test t-studenta odczytujemy z tabelki powyżej
tam p = 0.088144 > 0.05 zatem nie ma podstaw do odrzucenia hipotezy H0 o jednakowym czasie snu w obu grupach.

gdybyśmy wzieli H1 jako m1<m2 wychodzi nam p/2 < 0.05 w związku z czym odrzucamy H0

jeżeli byłoby czas snu i płeć wtedy sprawdzamy normalność:
wykresy -> wykresy skategoryzowane -> wykresy normalności i tam zaznaczamy test shearera wilka

Zad 4



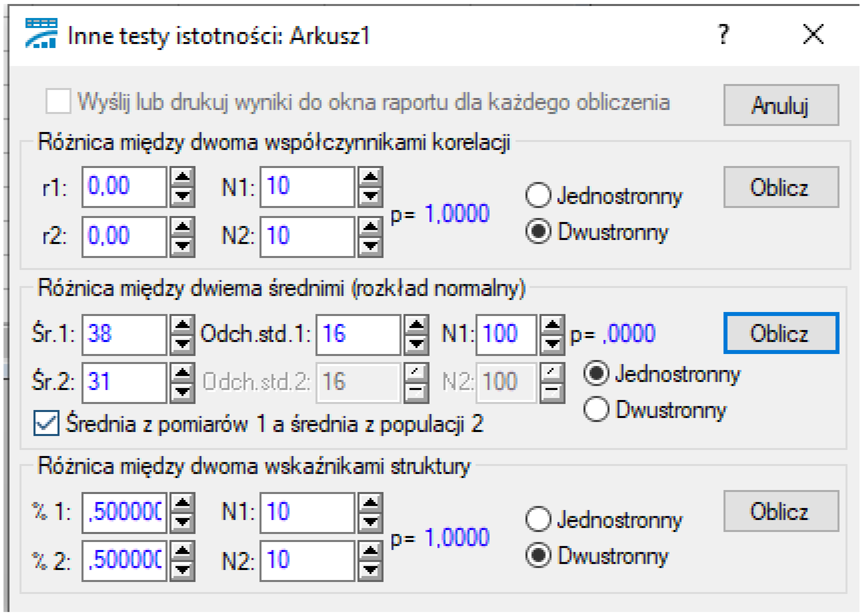
Ho m=4 (Średnia temp)

H1 m!=4

p= 0,1556 > a

Odrzucamy H0 na rzecz H1.

Zad 5



h0 m = 31

h1 m > 31

p/2 = 0,000 < a

Nie ma podstaw do odrzucenia H0.

Lab 6

Wprowadzenie:

Statystyki podstawowe i tabele -> macierze korelacji -> dwie listy zmiennych -> więcej -> 2W rozrzutu ...wracamy -> wyswietl dokładną tabelę -> podsumowanie

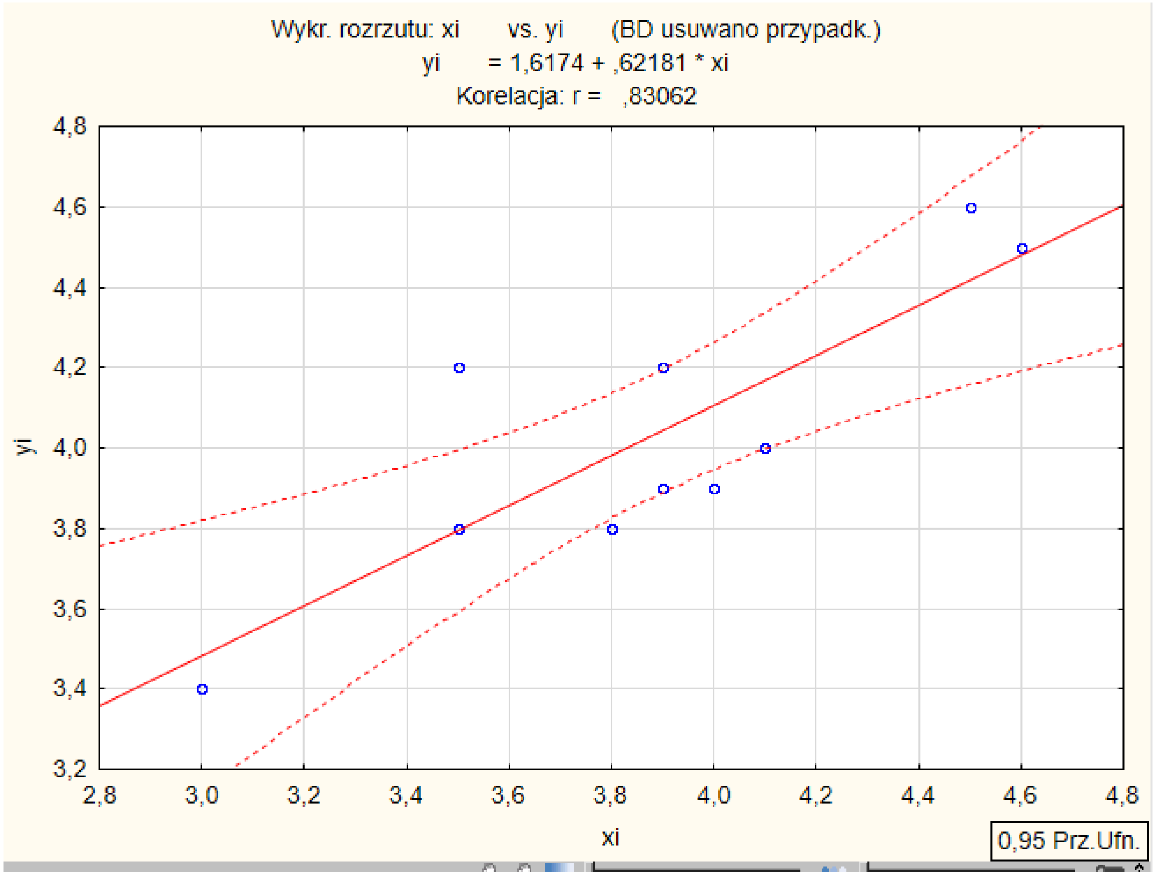
Współczynnik korelacji ***r(X,Y)*** wynosi ... co świadczy, o (silnej zależności, jak jest bliżej 1, słabej jak jest bliżej 0)

Współczynnik determinacji ***r2*** wynosi ... , co oznacza, że ... jest wyjaśniona w ..% przez ...

H0: wsp korelacji = 0

H1: wsp korelacji != 0

Zad 6



Korelacje (Arkusz1)
Oznaczone wsp. korelacji są istotne z p < ,05000
(Braki danych usuwano przypadkami)

Zmn. X & Zmn. Y	Średnia	Odch.st.	r(X,Y)	r2	t	p	Ważnych	Stała zal: Y	Nachyle zal: Y	Stała zal: X	Nachyle zal: X
xi	3,880000	0,475628									
yi	4,030000	0,356059	0,830619	0,689928	4,219063	0,002920	10	1,617387	0,621807	-0,591499	1,109553

Współczynnik korelacji ***r(X,Y)*** wynosi 0,83 co świadczy, o silnej zależności średniej ocen na IV roku od sredniej ocen na I roku.

Współczynnik determinacji ***r2*** wynosi 0,69 , co oznacza, że średnia ocen na IV roku jest wyjaśniona w 69% przez średnią ocen na I roku.

H0: wsp korelacji = 0

H1: wsp korelacji != 0

p = 0,029 < a= 0,05 zatem współczynnik korelacji jest istotnie rozny od zera.

- Znaleźć oszacowanie liniowej funkcji regresji cechy Y względem cechy X.

Aby znaleźć oszacowanie liniowej funkcji regresji musimy obliczyć błąd standardowy estymacji:

Statystyka -> regresja wieloraka (Ważna jest kolejność zmiennych, najpierw podejmy zmienna zależna, czyli IV rok i dalej I rok, klikamy ok i podsumowanie)

N=10	Podsumowanie regresji zmiennej zależnej: yi (Arkusz1) R= ,83061926 R^2= ,68992835 Popraw. R2= ,65116940 F(1,8)=17,800 p<,00292 Błąd std. estymacji: ,21029					
	b*	Bł. std. z b*	b	Bł. std. z b	t(8)	p
W. wolny			1,617387	0,575690	2,809475	0,022862
xi	0,830619	0,196873	0,621807	0,147380	4,219063	0,002920

Zależność średniej ocen na IV roku od średniej na I roku można zapisać wzorem:

y = 0,62x + 1,62 +/- 0,21 (+/- to bład estymacji)

Kolos_1

(1) Wadliwość procesu produkcyjnego wynosi 10%. Obliczyć prawdopodobieństwo, że na dziesięć wylosowanych produktów będzie mniej niż 8 dobrych.

(2) Procentowa zawartość tłuszczu w mleku 50 krów wynosi:
3.35, 4.16, 3.24, 4.23, 3.42, 3.73, 3.56, 3.98, 3.70, 4.47, 3.94, 3.92, 3.62, 3.53, 3.93, 4.16, 3.22, 4.10, 3.72, 4.26, 3.92, 3.66, 3.78, 3.96, 3.81, 4.28, 3.50, 3.39, 3.83, 4.27, 4.26, 3.71, 3.93, 4.27, 4.06, 3.78, 3.96, 3.89, 3.93, 4.06, 3.99, 3.77, 4.22, 3.78, 3.66, 3.41, 3.53, 3.54, 4.08, 3.44

(a) Dla powyższych danych skonstruować szereg rozdzielczy (7 przedziałów, zaczynając o 3.2) oraz histogram.

(b) Na poziomie istotności 0,01 zweryfikować hipotezę, że średnia zawartość tłuszczu w mle jest większa, niż 3, 7%.

(c) Wyznaczyć współczynnik zmienności i podać jego interpretację.

(d) Wyznaczyć kwartyl dolny i podać jego interpretację.

3) Dzienny przebieg taksówek pewnej korporacji charakteryzował szereg rozdzielczy:

przebieg w km	0-40	40-80	80-120	120-160	160-200
liczba taksówek	26	82	65	28	8

Obliczyć średnią arytmetyczną oraz odchylenie standardowe.

Kolos_2

(1) Wśród ziaren pszenicy znajduje się $0,6\%$ ziaren chwastów. Jakie jest prawdopodobieństwo, że wśród 50 losowo wybranych ziaren znajduje się co najmniej 45 ziaren pszenicy?

(2) Na podstawie danych pogrupowanych dotyczących wielkości gospodarstw rolnych w pewnej wsi, oszacować przedziałowo średnią wielkość gospodarstwa. Przyjąć poziom ufności 0,90.

wielkość gospodarstw w ha	(0,2>	(2,4>	(4,6>	(6,8>	(8,10>	(10,12>	(12,14>	(14,16>
liczba gospodarstw (n_i)	5	5	6	29	23	11	9	4

(3) W dwóch niezależnych próbach osób , które poddały się dwóm różnym kuracjom odchudzającym, otrzymano (X - ubytek wagi w kg):

$$n_1 = 16, \quad \bar{x}_1 = 5,4, \quad s_1 = 2,4,$$
$$n_2 = 11, \quad \bar{x}_2 = 3,4, \quad s_2 = 2,8.$$

Czy są podstawy by twiedzić na poziomie istotności 0,05, że średni ubytek wagi przy stosowaniu pierwszej kuracji jest wyższy niż w przypadku drugiej kuracji? Zakładamy, że rozkład ubytków wagi w każdym przypadku jest normalny.