

dr Robert Dargiewicz
r.dargiewicz@amisns.edu.pl

<https://as.kiero.net>

Parametryczne i nieparametryczne
statystyki opisowe. Profilaktyka określania
i rozwiązywania problemów badawczych.

Przegląd narzędzi statystycznych umożliwiających
poprawne wykonanie analizy i prezentacji wyników
w formie tabelarycznej oraz graficznej.



TAJNY JĘZYK NAUKOWCÓW

Instrukcja dla czytelników

Na prośbę naszych czytelników, narzekających na trudności w zrozumieniu języka, jakim pisane są artykuły naukowe, rozszyfrowujemy najczęściej używane wyrażenia.

1. *Od dawna wiadomo, że...* – nie chciało mi się szukać pracy źródłowej.

2. *... można dostrzec wyraźną tendencję...* – trudno wyciągnąć sensowne wnioski.

3. *... o wielkim znaczeniu teoretycznym i praktycznym* – mnie się to wydaje dość ciekawe.

4. *Znalezienie ostatecznej odpowiedzi na te pytania nie było na razie możliwe* – eksperyment się nie udał, ale i tak spróbuję opublikować wyniki.

5. *Trzy spośród zestawów danych wybrano do szczegółowej analizy* – pozostałych 20 zestawów nie nadawało się do niczego.

6. *Prezentowane są typowe wyniki* – są to najlepsze wyniki.

7. *Oczywiście* – nie sprawdzałem tego, ale...

8. *Najdokładniejsze wyniki otrzymał mgr Kikut* – mgr Kikut to mój asystent.

9. *Uznaje się, że...* – uważam, że...

10. *Powszechnie uznaje się, że...* – oprócz mnie jeszcze kilku facetów tak uważa.

11. *Jest jasne, że wiele pracy należy włożyć, zanim pełna odpowiedź stanie się możliwa* – nic z tego nie rozumiem.

12. *... dane są poprawne z dokładnością do rzędu wielkości...* – dane są bardzo niedokładne.

13. *Oczekuje się, że niniejszy artykuł pobudzi zainteresowanie tą dziedziną* – ten artykuł jest marny, ale inne w omawianej dziedzinie są równie słabe.

14. *Wybrano stop bizmutu z ołowiem, ponieważ właśnie dla tego materiału oczekiwany efekt powinien zaistnieć najwyraźniej* – innego stopu nie mieliśmy.

15. *... został przypadkowo lekko uszkodzony w czasie badań* – spadło nam na podłogę.

16. *... obchodziliśmy się z nadzwyczajną ostrożnością* – nie spadło nam na podłogę.

17. *Urządzenie automatyczne* – ma wyłącznik.

18. *Urządzenie półportatywne* – wyposażone w korbkę.

19. *Portatywne* – wyposażone w dwie korbki.

20. *Typowe wyniki zostały przedstawione na stronie...* – przedstawiono najlepsze wyniki.

21. *Staranna analiza posiadanych wyników...* – Trzy strony notatek zamazały się, gdy potrąciłem szklankę z piwem.

22. *... bardzo ważne pole odkrywczych badań* – bezużyteczny i głupi temat zasugerowany przez zwierzchników.

23. *Składałem podziękowania mgr. Łabądkowi za pomoc w pracy doświadczalnej, a dr. Pikucie za cenne dyskusje* – Łabądek odwalił czarną robotę, a Pikuta wyjaśnił mi, co oznaczają wyniki.

24. *Można by podać w wątpliwość, czy...* – tę wątpliwość sam wymyśliłem, bo mam na to świetną odpowiedź.

Od dawna wiadomo, że...

- znaczy tyle że nie chciało mi się szukać pracy źródłowej ale każdy poważny naukowiec dokładnie wie że tak właśnie jest – lub być powinno

...można dostrzec wyraźną tendencję...

- równoważne z określeniem że trudno wyciągnąć jakieś sensowne wnioski

Znalezienie ostatecznej odpowiedzi na te pytania nie było na razie możliwe

- eksperyment się nie udał ale i tak spróbujemy opublikować wyniki

Przedstawione są typowe wyniki

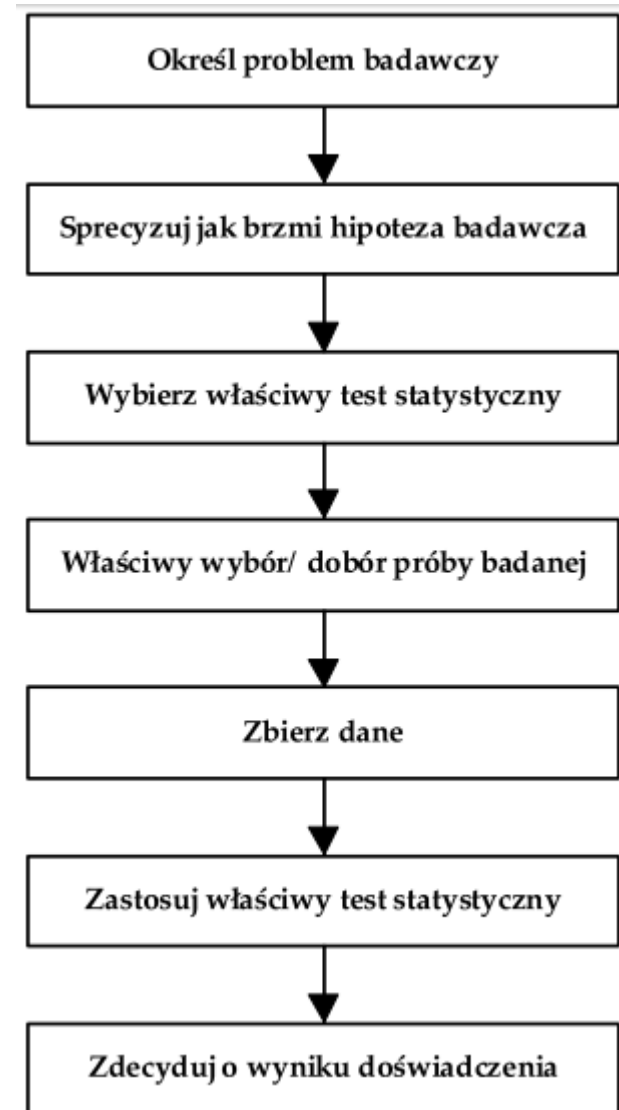
- przedstawiono najlepsze wyniki pasujące do danej tezy.

Powszechnie uznaje się, że...

- oprócz mnie jeszcze kilku facetów tak myśli...

Schemat postępowania w planowaniu doświadczenia

Przystępując do wykonania eksperymentu naukowego w myśl sprawdzenia pomysłu, koncepcji czy hipotezy musimy wiedzieć, jak brzmi ta hipoteza i czego mamy bronić lub co obalać.



Jak brzmi hipoteza?

Jak brzmi pytanie badawcze?

2-Poziom na notę do 4.0 - zaprezentować odpowiedź na dowolne zagadnienie:

- D. Czy badane kobiety różnią się zmierzonymi wynikami od badanych mężczyzn?
- E. Czy badani mężczyźni różnią się zmierzonymi wynikami od badanych kobiet?
- F. Czy występują różnice w wynikach badanych ze względu na płeć?
- G. Czy występują różnice w wynikach badanych ze względu na BMI>25?

3-Poziom na notę do 4,5 - zweryfikować dowolne zagadnienie:

- H. Czy uległy zmianie wyniki badanych kobiet po podaniu leku X?
- I. Czy uległy zmianie wyniki badanych mężczyzn po podaniu leku X?
- J. Czy uległy zmianie wyniki badanych po podaniu leku X?

4-Poziom na notę do 5.0 - zweryfikować dowolną hipotezę:

- X. Nadwaga badanych (BMI>25) zmniejsza skuteczność badanego leku X na obniżenie ciśnienia.
- Y. Płeć badanych nie wpływa na skuteczność badanego leku X na obniżenie ciśnienia.
- Z. Wiek badanych nie wpływa na skuteczność badanego leku X na obniżenie ciśnienia.

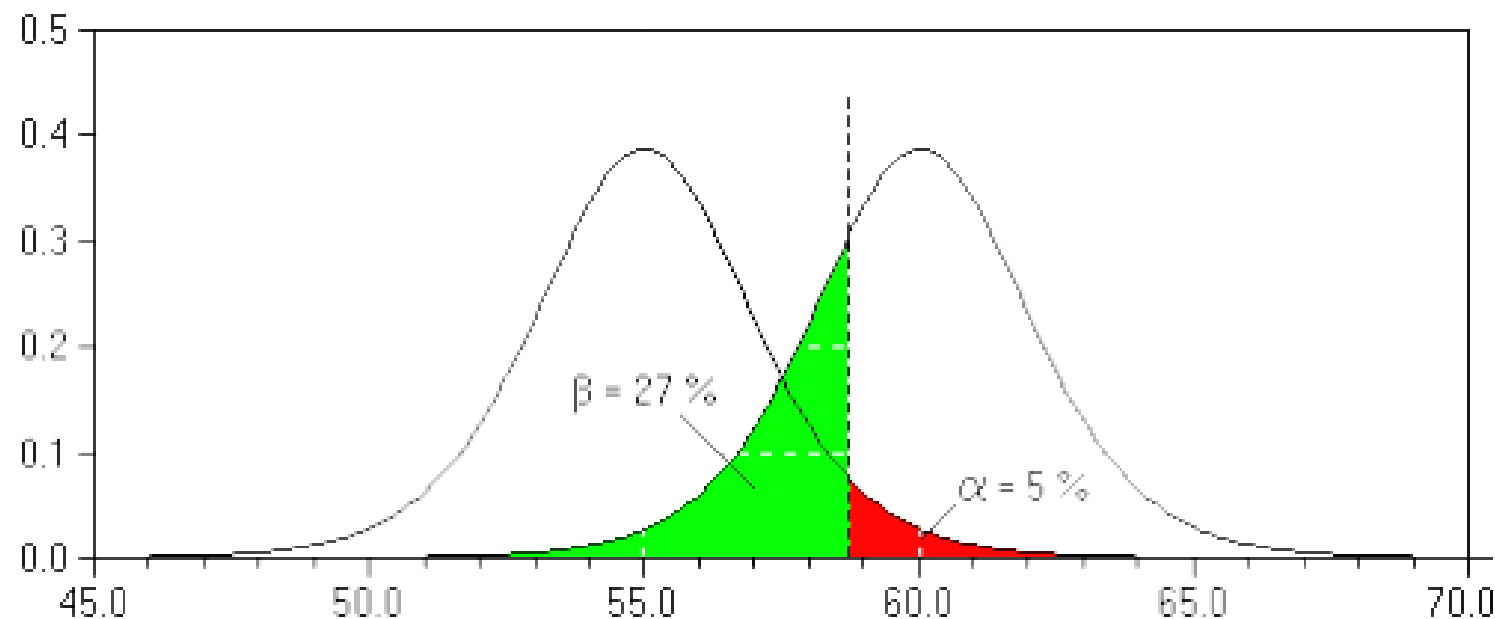
BŁĘDY PIERWSZEGO I DRUGIEGO RODZAJU

wynik testu	Rzeczywistość	
	H_0	H_1
H_0	ok	β
H_1	α	ok

błąd pierwszego rodzaju α polega na tym, że wynik testu wykazuje istnienie istotnej statystycznie różnicy, podczas gdy w rzeczywistości ta różnica nie jest istotna. Prawdopodobieństwo błędu I rodzaju nazywamy poziomem istotności.

błąd drugiego rodzaju β polega na stwierdzeniu na podstawie przeprowadzonego testu statystycznego, że średnie nie różnią się istotnie statystycznie, podczas gdy w rzeczywistości taka różnica istnieje. Określa on stopień akceptowanego przez eksperymentatora ryzyka zagubienia rzeczywiście istniejącej różnicy między średnimi. Na pojawienie się błędu drugiego rodzaju ma istotny wpływ zbyt mała liczebność próby oraz duże rozproszenie pomiarów w próbie.

BŁĘDY PIERWSZEGO I DRUGIEGO RODZAJU



Wartości α i β są ze sobą powiązane. Zmniejszenie prawdopodobieństwa α powoduje wzrost prawdopodobieństwa β .

DOBÓR TESTU STATYSTYCZNEGO

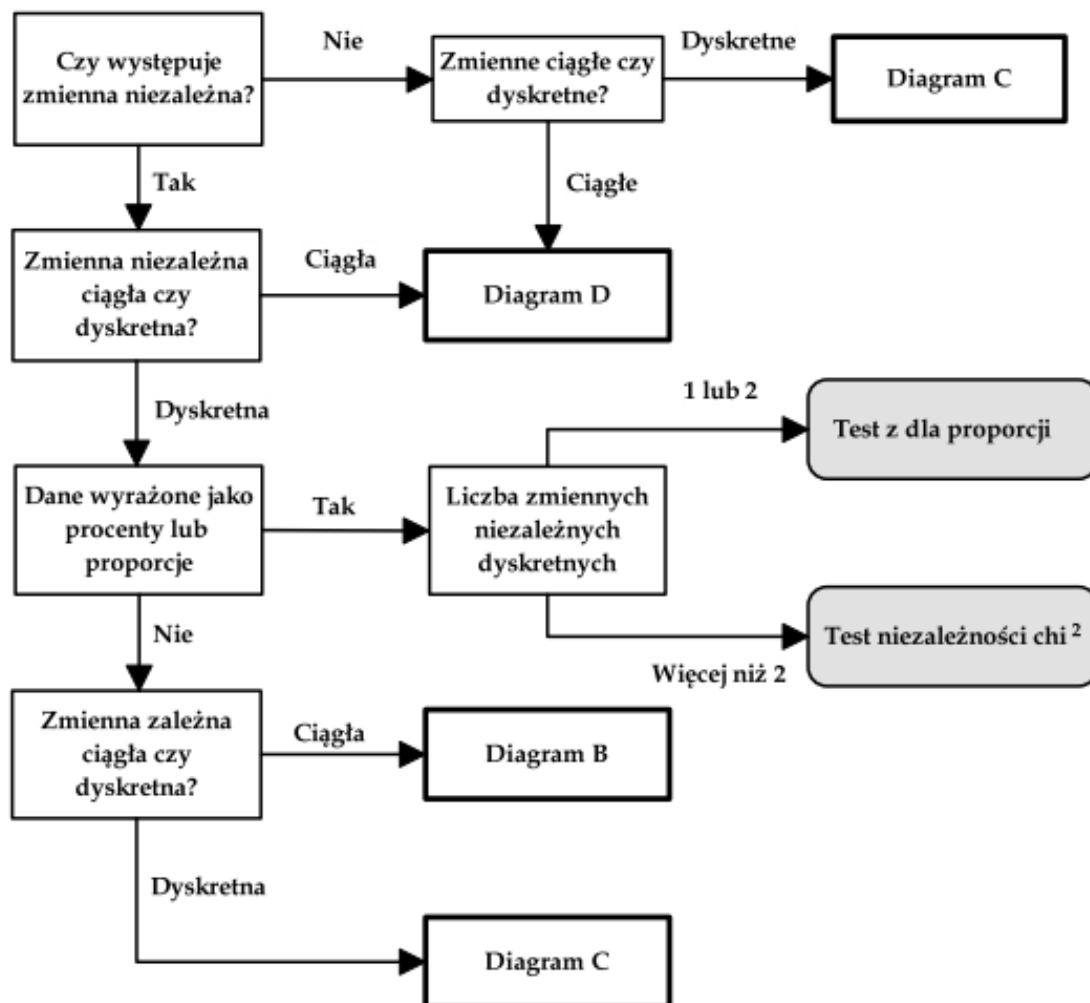
- czy mamy do czynienia z problemem badania istotności różnic czy też problemem badania zależności?
- w jakiej skali pomiarowej zostały wykonane pomiary (interwałowej, porządkowej, nominalnej) ?
- jeżeli dane wyrażone są w skali interwałowej to czy pochodzą z populacji o rozkładzie normalnym?
- ile grup porównujemy jednocześnie (dwie, więcej niż dwie) ?
- jaki model badania został wykorzystany przy zbieraniu danych (zmienne niepowiązane czy powiązane)?

LICZEBNOŚĆ BADANEJ PRÓBY

- PRÓBA O ZBYT WIELKIEJ LICZEBNOŚCI –
KONIECZNOŚĆ ZAANGAŻOWANIA
DODATKOWYCH OSOBNIKÓW, PODWYŻSZENIE
KOSZTÓW I CZASU TRWANIA BADANIA
- PRÓBA O ZBYT MAŁEJ LICZEBNOŚCI -
UNIEMOŻLIWIA WYKRYCIE WAŻNYCH Z
PUNKTU WIDZENIA KLINICZNEGO ZJAWISK –
ZWIĘKSZONE RYZYKO UZYSKIWANIA WYNIKÓW
FAŁSZYWIE NEGATYWNYCH (

Właściwy test statystyczny

Diagram A



Cezary Watała

*Zakład Zaburzeń Krzepnięcia Krwi
Uniwersytetu Medycznego w Łodzi*

Biostatystyka

– wykorzystanie metod statystycznych
w pracy badawczej
w naukach biomedycznych

Copyright © 2012 by α-medica press
All rights reserved

Wszystkie prawa zastrzeżone
ISBN 978-83-7522-072-8

Wydanie II

strona 32 do 35

Właściwy test statystyczny

Diagram B

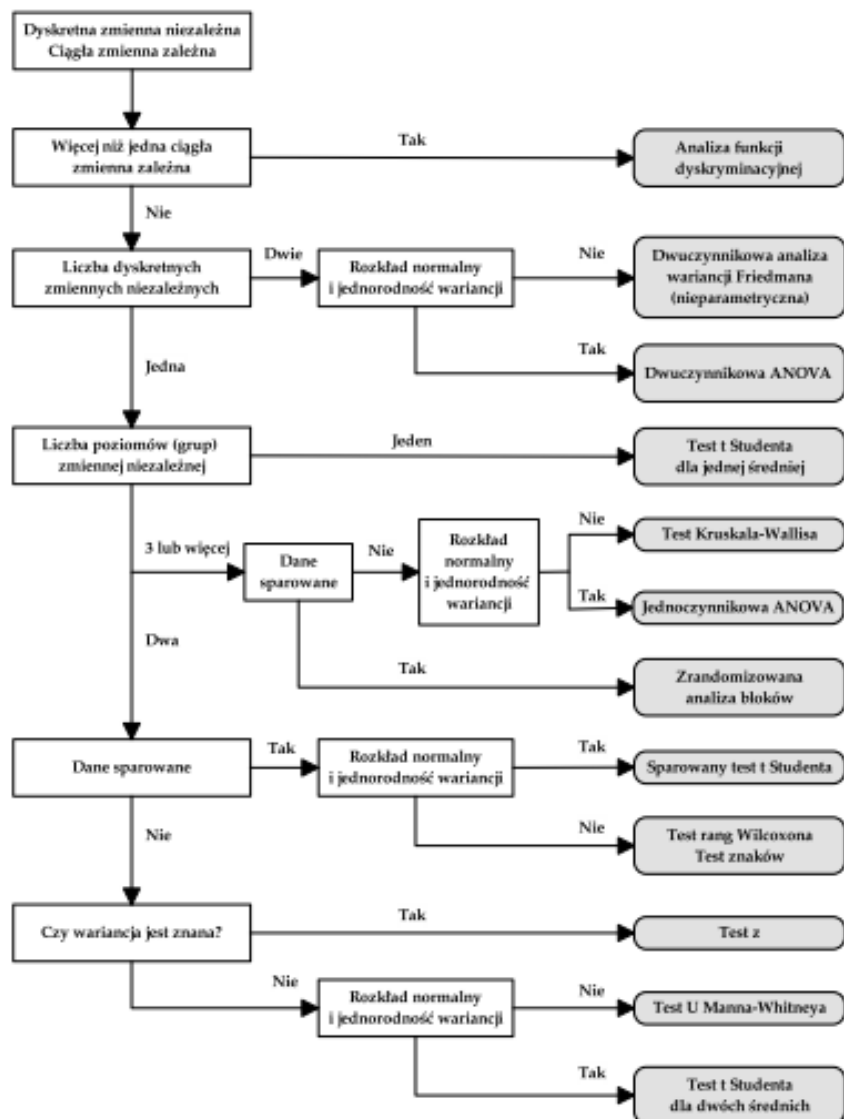
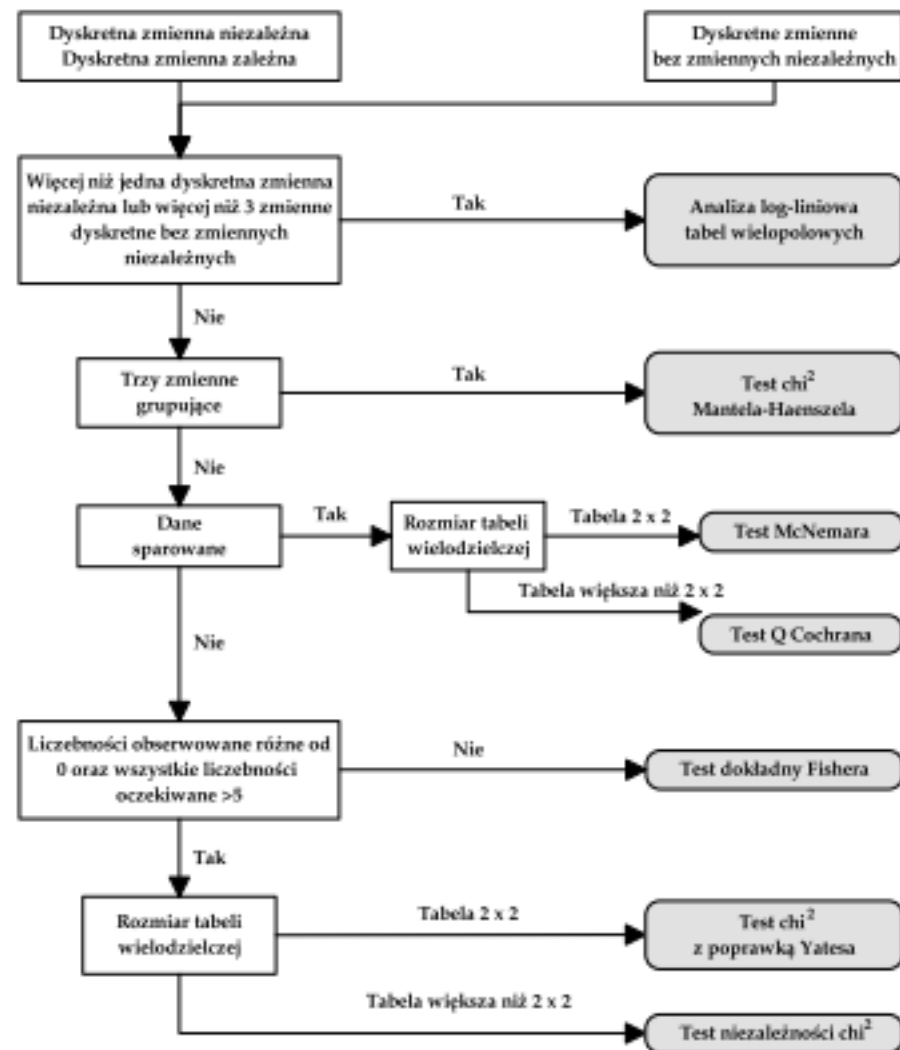
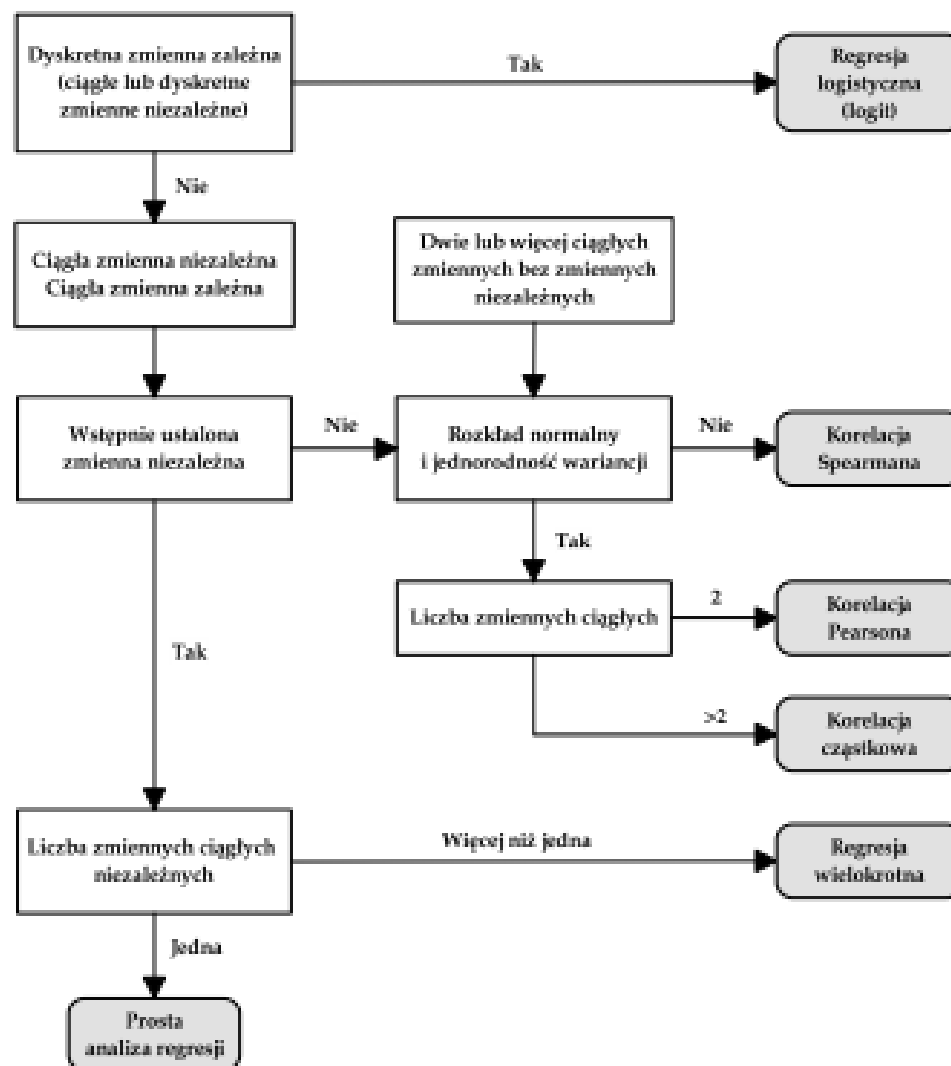


Diagram C



Właściwy test statystyczny

Diagram D



WYBÓR NAJCZĘŚCIEJ STOSOWANYCH TESTÓW W BADANIACH MEDYCZNYCH I WARUNKI ICH STOSOWANIA.

skala / liczba grup	warunki dodatkowe	dwie grupy	dwie grupy	więcej niż dwie grupy	więcej niż dwie grupy
		Zmienne niepowiązane	zmienne powiązane	zmienne niepowiązane	zmienne powiązane
interwałowa	normalność rozkładu	test t-Studenta (niepowiąz.)	test t-Studenta (powiązane)	analiza wariancji	analiza wariancji
interwałowa	brak normalności rozkładu	test Manna-Whitneya	test Wilcoxona	test Kruskala-Wallisa	test Friedmana
porządkowa	-	test Manna-Whitneya	test Wilcoxona	test Kruskala-Wallisa	test Friedmana
nominalna	-	test chi-kwadrat lub dokładny test Fishera	test znaków lub test McNemara	test chi-kwadrat lub dokładny test Fishera	test Q-Cochrana

MODEL BADANIA

- Model zmiennych niepowiązanych (niezależnych)
 - dwie lub więcej prób pobranych z dwu lub więcej różnych populacji.
- Model zmiennych powiązanych (sparowanych, zależnych)
 - ten sam pacjent (zjawisko itp.) badane są dwukrotnie lub większą ilość razy.

SPRAWDZANIE ZAŁOŻEŃ

- **NORMALNOŚĆ**
- **HOMOGENICZNOŚĆ**
(homoscedastyczność)
- **ZWIĄZEK LINIOWY**

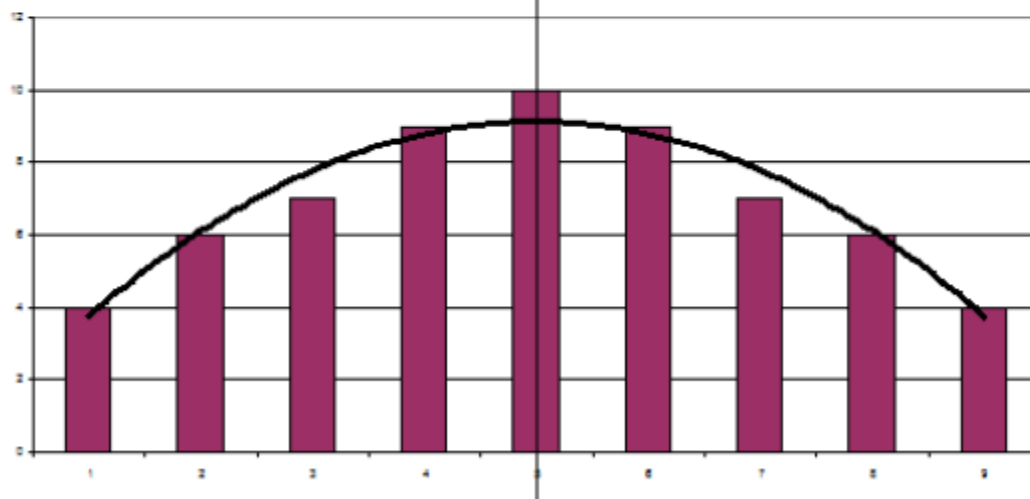
Kształt rozkładu, miary opisowe i normalność rozkładu

ANALIZA ROZKŁADÓW - ROZKŁAD NORMALNY

OBLICZAMY:
Xsr, D, Me, Q1, Q3, SD, rozpiętość wyników

1. Rozkład symetryczny

$X = do$



cechy
rozkładu

$do = x$

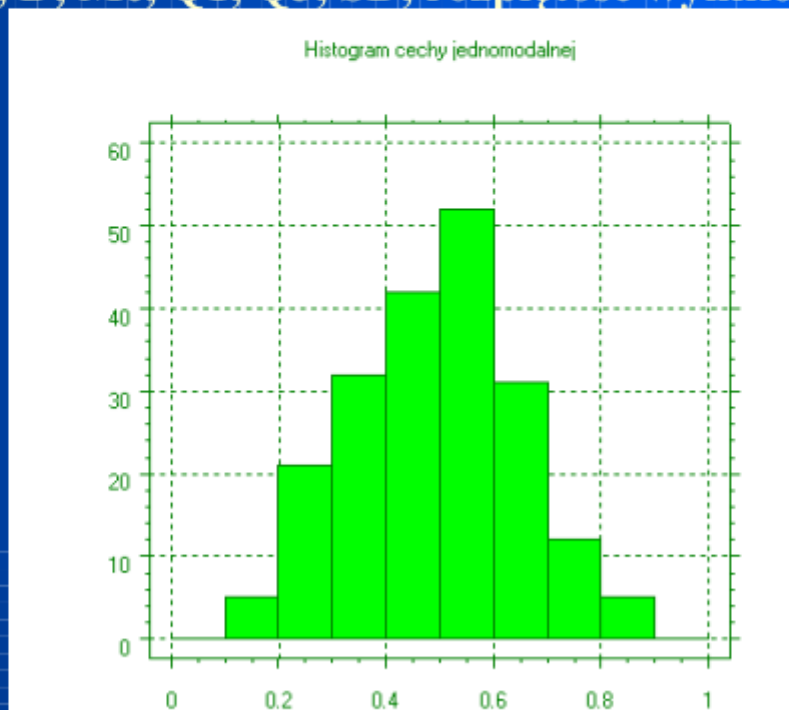
$Q_3 - me =$
 $= me - Q_1$

Kształt rozkładu, miary opisowe i normalność rozkładu

ANALIZA ROZKŁADÓW - ROZKŁAD NORMALNY

OBLICZAMY:

$\bar{X}$, D, Me, Q1, Q3, SD, rozpiętość wyników

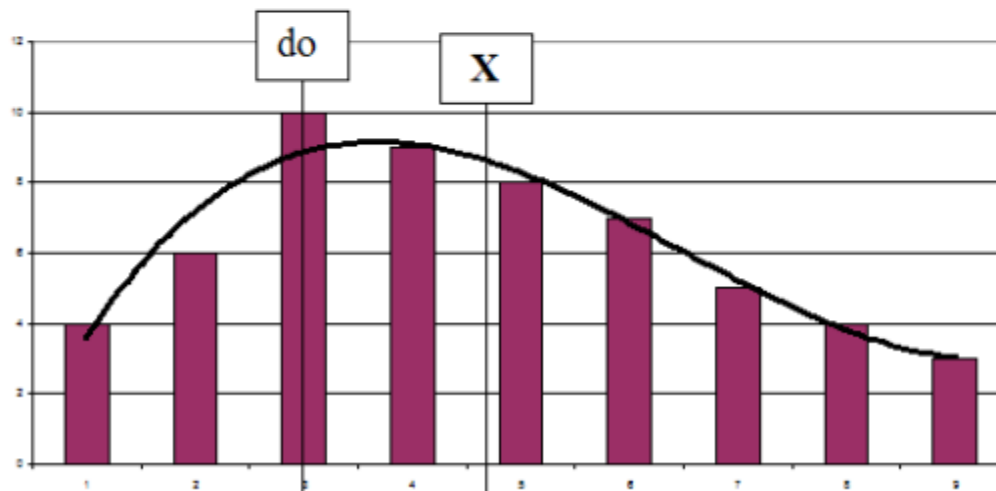


Kształt rozkładu, miary opisowe i normalność rozkładu

ANALIZA ROZKŁADÓW - UMIARKOWANIE ASYMETRYCZNE

OBLICZAMY:
 $\bar{X}$, D, Me, Q1, Q3, SD, rozpiętość wyników

2. Asymetria prawostronna



cechy
rozkładu

$do < \bar{x}$

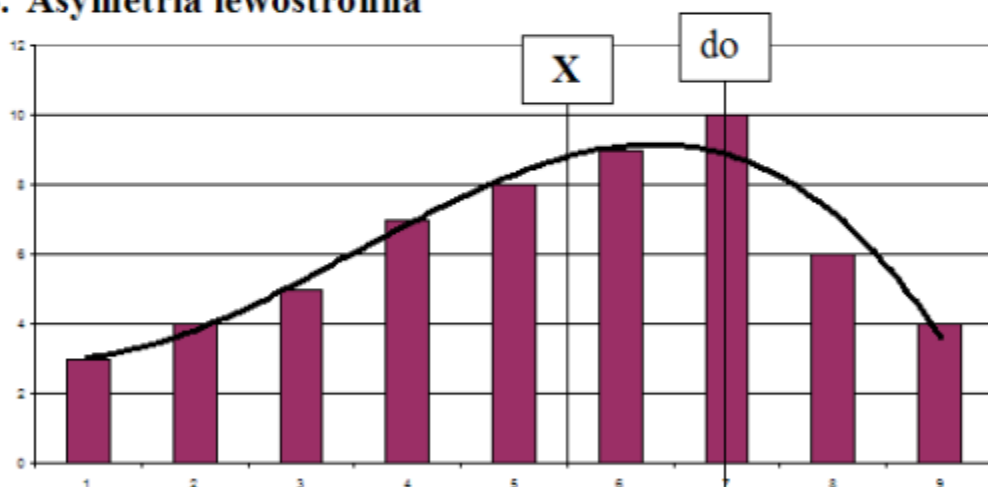
$Q_3 - me >$
 $me - Q_1$

Kształt rozkładu, miary opisowe i normalność rozkładu

ANALIZA ROZKŁADÓW - UMIARKOWANIE ASYMETRYCZNE

OBLICZAMY:
 $\bar{X}$, D, Me, Q1, Q3, SD, rozpiętość wyników

3. Asymetria lewostronna



cechy
rozkładu

$do > \bar{x}$

$Q_3 - me <$
 $me - Q_1$

Silę asymetrii wyznacza współczynnik $A = (\bar{x} - do)/S$

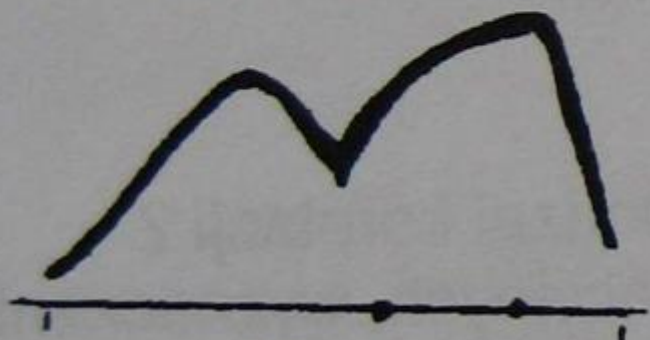
Kształt rozkładu, miary opisowe i normalność rozkładu

ANALIZA ROZKŁADÓW - ROZKŁADY WIELOMODALNE

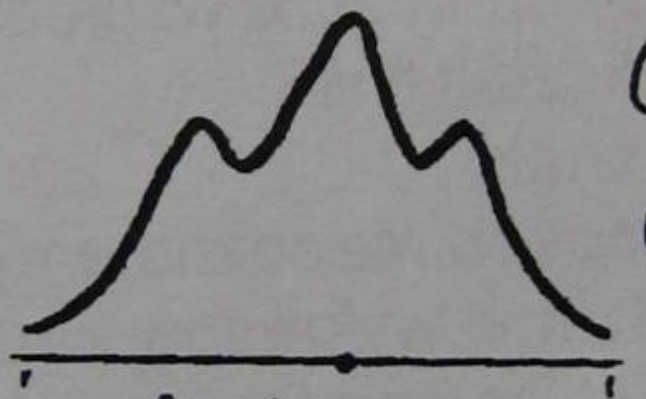
OBLICZAMY:

a) D, Me, Q1, Q3, b) $\bar{X}$, D, Me, Q1, Q3, wsp. zmienności

3. Wielomodalne



a) dwumodalny



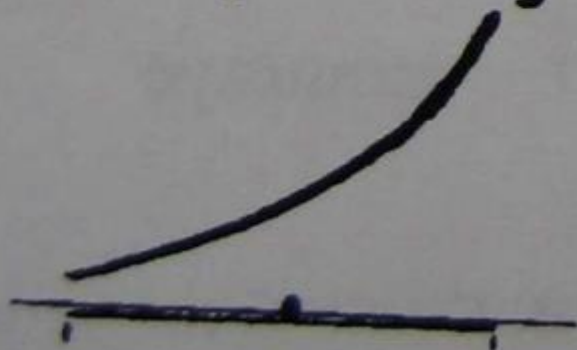
b) wielomodalny

Kształt rozkładu, miary opisowe i normalność rozkładu

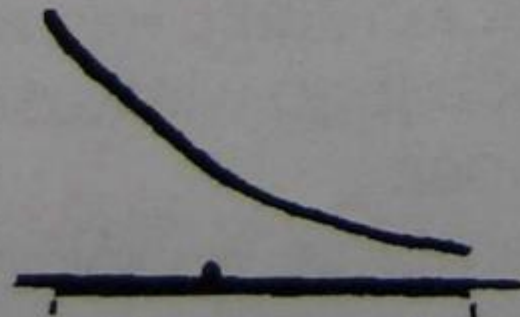
ANALIZA ROZKŁADÓW - SKRAJNIE ASYMETRYCZNE

OBLICZAMY:
D, Me, Q1, Q3,

4. Skrajnie asymetryczne



a) f. rosnąca

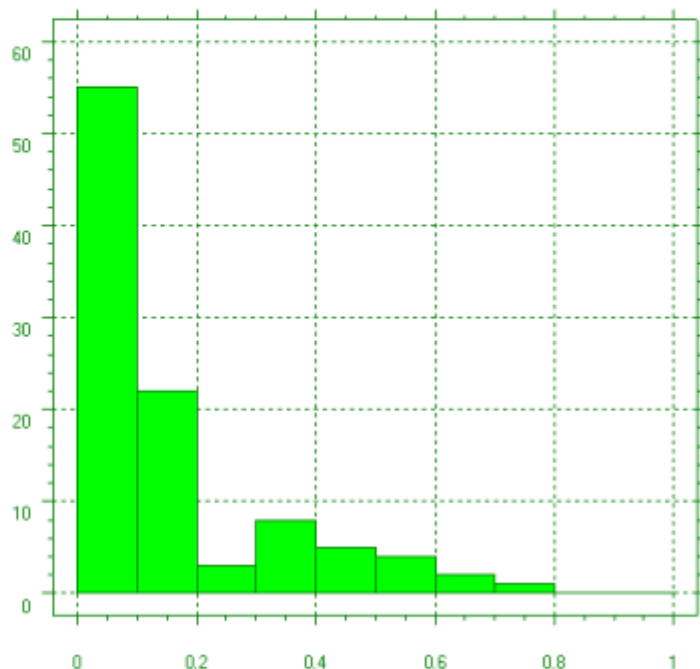


b) f. malejąca

Kształt rozkładu, miary opisowe i normalność rozkładu

ANALIZA ROZKŁADÓW - SKRAJNIE ASYMETRYCZNE

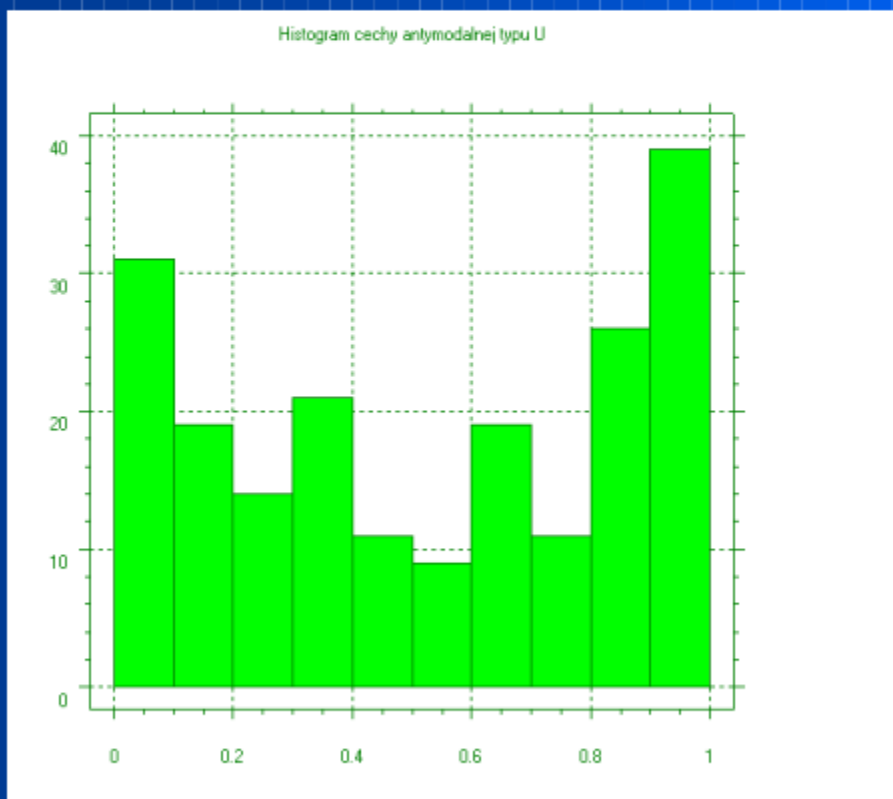
Histogram cechy antymodalnej typu J



OBLICZAMY:
D, Me, Q1, Q3,

Kształt rozkładu, miary opisowe i normalność rozkładu

ANALIZA ROZKŁADÓW - ROZKŁAD "U"



OBLICZAMY:
D, Me, Q1, Q3,

Kształt rozkładu, miary opisowe i normalność rozkładu

O przystawalności rozkładu danych doświadczalnych do rozkładu normalnego wnioskujemy najczęściej przeprowadzając jeden z testów normalności np.:

- test zgodności χ^2 ,
- test Kolmogorowa-Smirnowa,
- **test W Shapiro-Wilka**

(interpretacja polega na weryfikacji prawdopodobieństwa wystąpienia wartości krytycznej i popełnienia błędu np. przyjmujemy, że dla $p < 0.05$ – przyjmujemy brak zgodności, czyli odrzucamy Hipotezę 0)

Testy pozwalają obliczyć prawdopodobieństwo „p” tego, że próba pochodzi z populacji o rozkładzie normalnym.

Cezary Watała

*Zakład Zaburzeń Krzepnięcia Krwi
Uniwersytetu Medycznego w Łodzi*

Biostatystyka

– wykorzystanie metod statystycznych
w pracy badawczej
w naukach biomedycznych

Copyright © 2012 by α -medica press
All rights reserved

Wszystkie prawa zastrzeżone
ISBN 978-83-7522-072-8

Wydanie II

strona 183 do 186

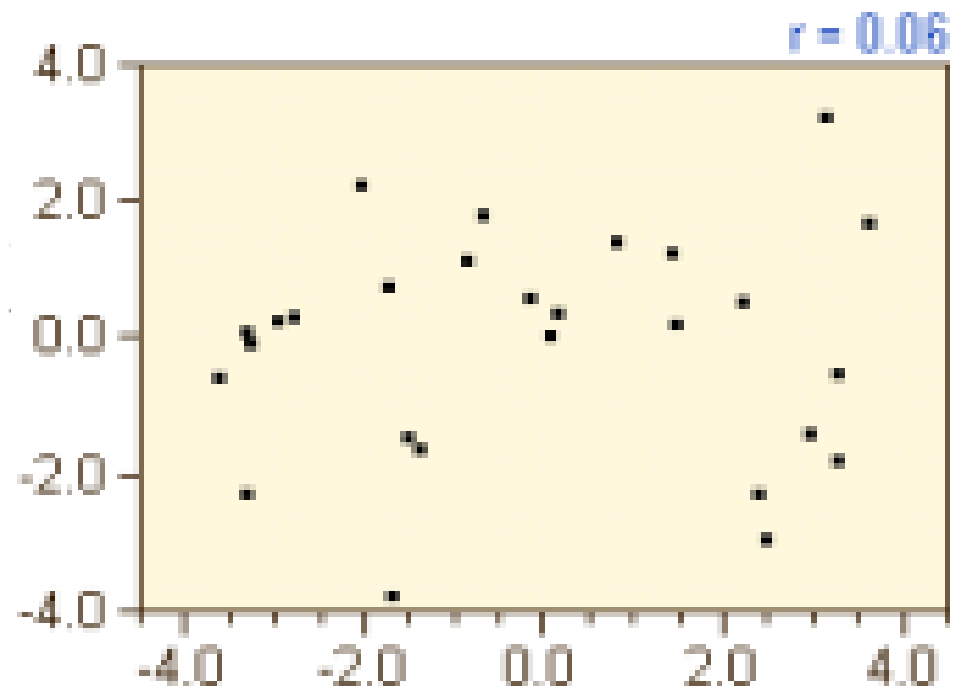
ZAGADNIENIE KORELACJI I REGRESJI

metody korelacyjne badają istnienie ewentualnego związku między dwoma lub większą liczbą zmiennych oraz szacują siłę i istotność statystyczną tego związku.

Aby zbadać czy istnieje zależność między dwoma parametrami na skali interwałowej należy stworzyć wykres przedstawiający rozrzut par punktów doświadczalnych (tzw. scattergram), pozwala on w pierwszym przybliżeniu określić kształt badanego związku.

Na osi OX - parametr, który jest określany jako zmienna niezależna.

Na osi OY - parametr, który jest określany jako zmienna zależna



ANALIZA WSPÓŁZALEŻNOŚCI CECH

CEL:

zbadanie charakteru zależności oraz ustalenie siły związku

ZALEŻNOŚCI POMIĘDZY:

A. Dwoma cechami mierzalnymi
(wzrost i długość skoku z miejsca)

B. Cechą mierzalną i cechą niemierzalną
(wykształcenie i obwód w bicepsie)

C. Dwoma cechami niemierzalnymi
(pochodzenie społeczne i przekonania polityczne)

ANALIZA WSPÓŁZALEŻNOŚCI CECH

ANALIZA WSPÓŁZALEŻNOŚCI CECH MIERZALNYCH

ZALEŻNOŚCI POMIĘDZY:

A. Dwoma cechami mierzalnymi

Tabela 22

Lp.	Wysokość ciała w cm x_1	Wyniki w skoku wzwyż w cm y_1
1	167	126
2	163	123
3	161	122
4	162	122
5	159	118
6	160	119
7	161	121
8	167	125
9	169	128
10	171	126

ANALIZA WSPÓŁZALEŻNOŚCI CECH

ANALIZA WSPÓŁZALEŻNOŚCI CECH MIERZALNYCH - CECHY MIERZALNE

Każdej obserwacji odpowiadają dwie wartości (każdej studentce odpowiada „pewna” wysokość ciała i uzyskany przez nią wynik w skoku wzwyż).

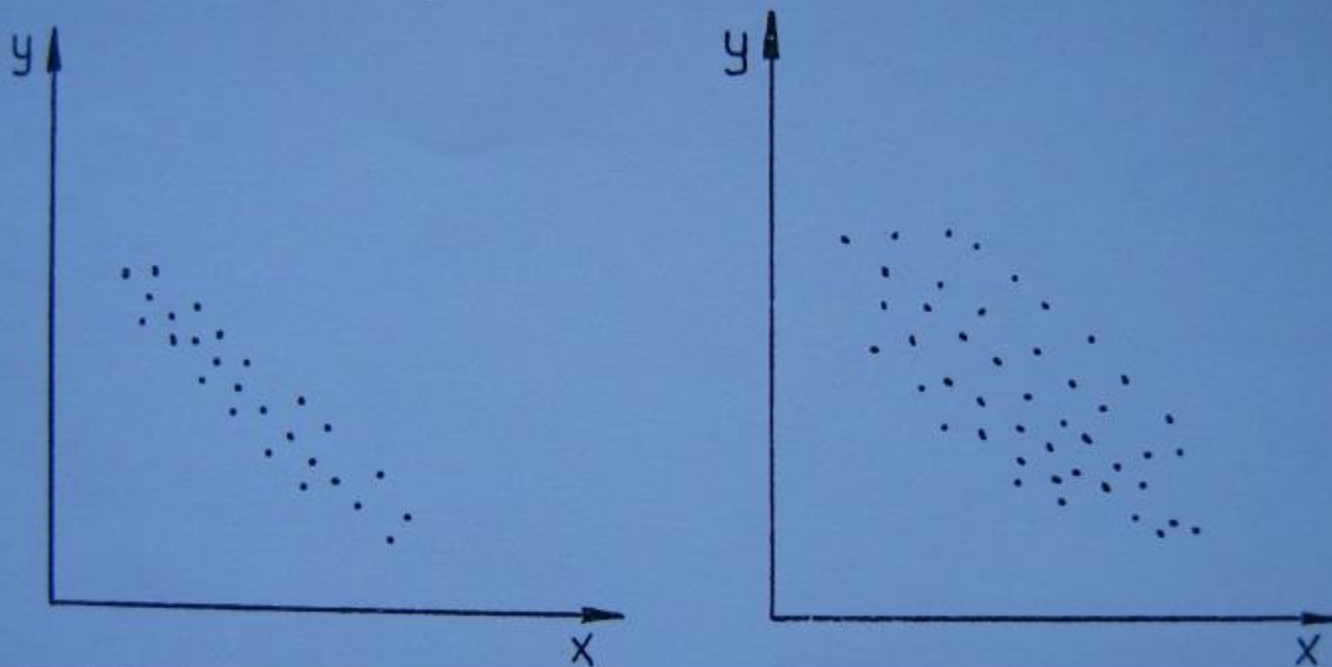
Szereg taki przyjęto nazywać dwucechowym szeregiem statystycznym. Chcąc przedstawić graficznie dwucechowy szereg postępujemy



Zależność prostoliniowa dodatnia ($p = 0,5$) (wraz ze wzrostem wartości jednej cechy rośnie wartość drugiej cechy)

ANALIZA WSPÓŁZALEŻNOŚCI CECH

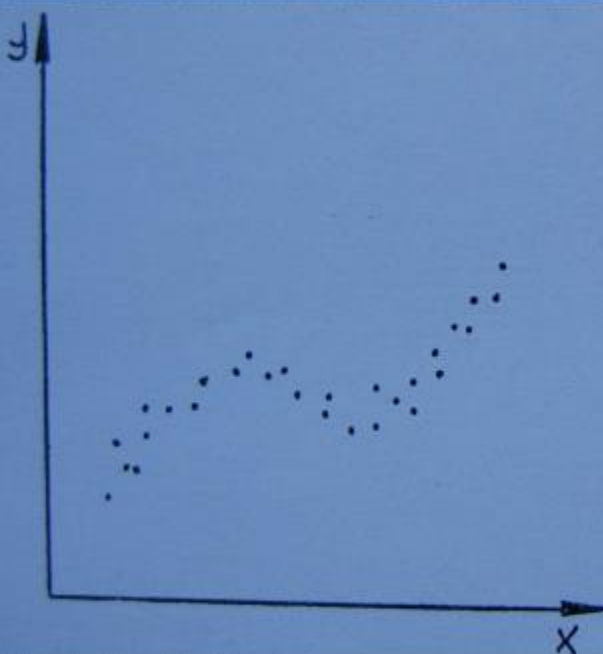
ANALIZA WSPÓŁZALEŻNOŚCI CECH MIERZALNYCH - CECHY MIERZALNE



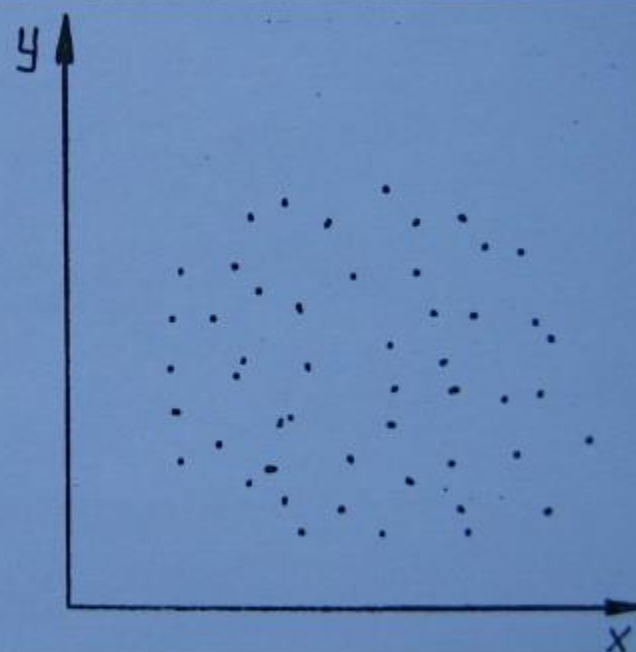
Zależność prostoliniowa ujemna ($p = -0,3$) (wraz ze wzrostem wartości jednej cechy maleje wartość drugiej cechy)

ANALIZA WSPÓŁZALEŻNOŚCI CECH

ANALIZA WSPÓŁZALEŻNOŚCI CECH MIERZALNYCH - CECHY MIERZALNE



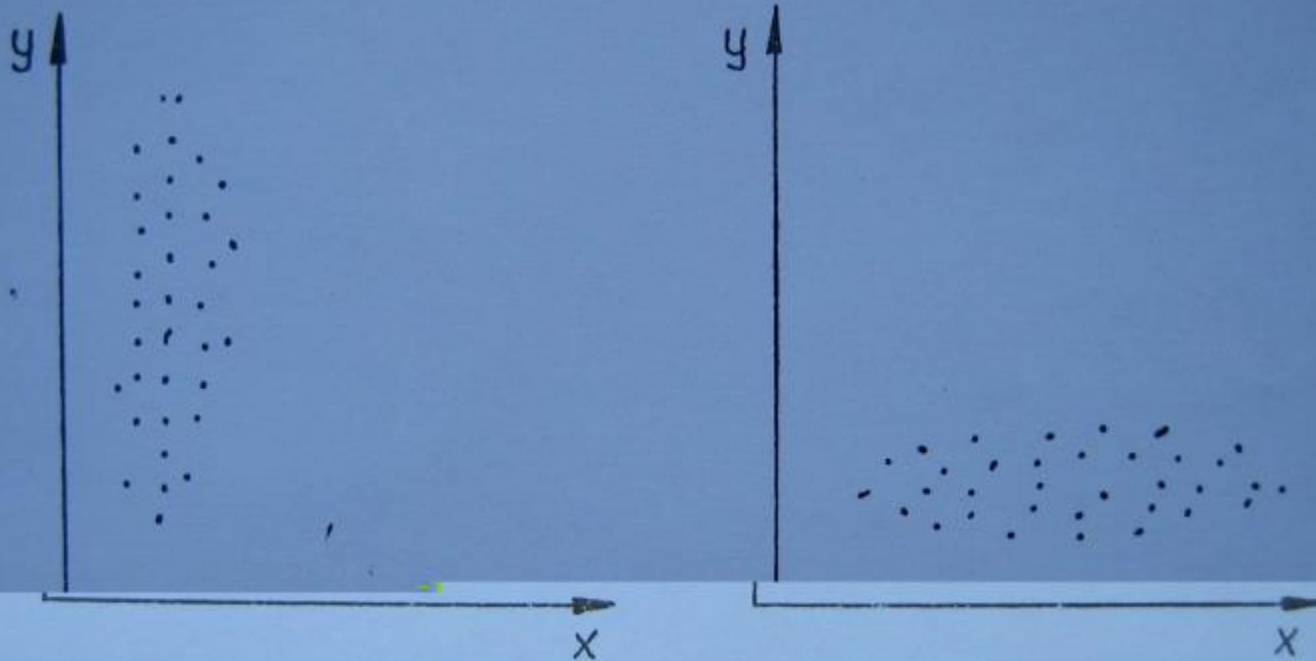
Zależność krzywoliniowa



Brak zależności lub słaba zależność

ANALIZA WSPÓŁZALEŻNOŚCI CECH

ANALIZA WSPÓŁZALEŻNOŚCI CECH MIERZALNYCH - CECHY MIERZALNE



Brak zależności lub słaba zależność

ANALIZA WSPÓŁZALEŻNOŚCI CECH

ANALIZA WSPÓŁZALEŻNOŚCI CECH MIERZALNYCH - **CECHY MIERZALNE**

Korelacja - nazwa pochodzi z łacińskiego „correlatio” - wzajemny stosunek

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n \cdot S_x S_y}$$

gdzie:

x_i — wartości cechy X

$\bar{x}$ — średnia arytmetyczna tej cechy

S_x — odchylenie standardowe cechy X

$y_i, \bar{y}, S_y$ — jak wyżej dla cechy Y

Współczynnik korelacji liniowej Pearsona

ANALIZA WSPÓŁZALEŻNOŚCI CECH

ANALIZA WSPÓŁZALEŻNOŚCI CECH MIERZALNYCH - CECHY MIERZALNE

Współczynnik ten przyjmuje wartości od -1 do $+1$, czyli $-1 \leq r \leq 1$
Jeżeli

$r_{xy} > 0$ — cechy są skorelowane dodatnio, tym silniej, im r_{xy} jest bliższy 1

$r_{xy} < 0$ — cechy są skorelowane ujemnie, tym silniej, im r_{xy} jest bliższy -1

A zatem znak współczynnika informuje o kierunku zależności, a jego wartość bezwzględna o sile tego związku.

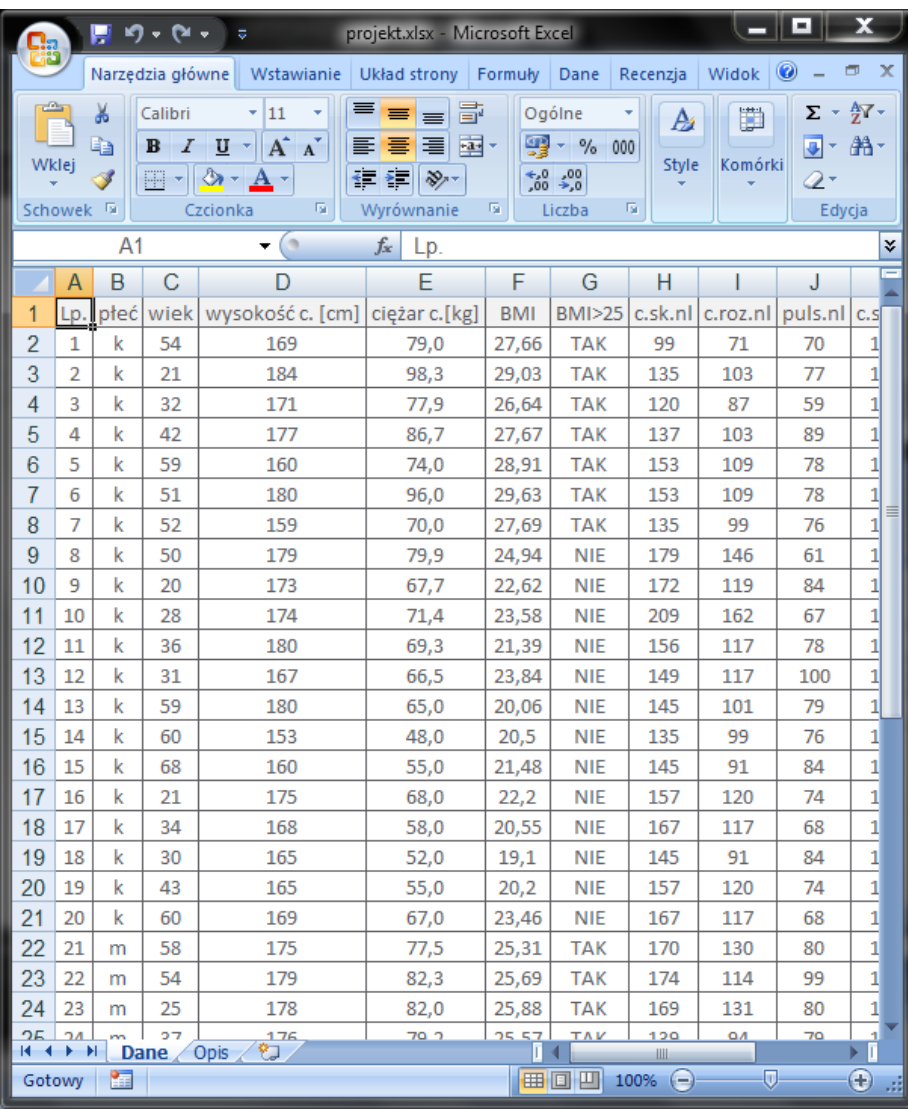
Współczynnik korelacji przybiera wartość -1 lub $+1$ tylko wtedy, gdy zależność między cechami X i Y wyraża się funkcją liniową.

W obu przypadkach każdej wartości x odpowiada tylko jedna wartość y .
Guilford*) proponuje przyjąć następującą interpretację współczynnika korelacji liniowej:

	$r_{xy} < 0,20$	— zależność prawie nie znacząca
$0,20 \leq r_{xy} < 0,40$	$r_{xy} < 0,40$	— zależność wyraźna lecz mała (korelacja niska)
$0,40 \leq r_{xy} < 0,70$	$r_{xy} < 0,70$	— zależność istotna (korelacja umiarkowana)
$0,70 \leq r_{xy} < 0,90$	$r_{xy} < 0,90$	— zależność znaczna (korelacja wysoka)
$0,90 \leq r_{xy} < 1,00$	$r_{xy} < 1,00$	— zależność bardzo pewna (korelacja bardzo wysoka)

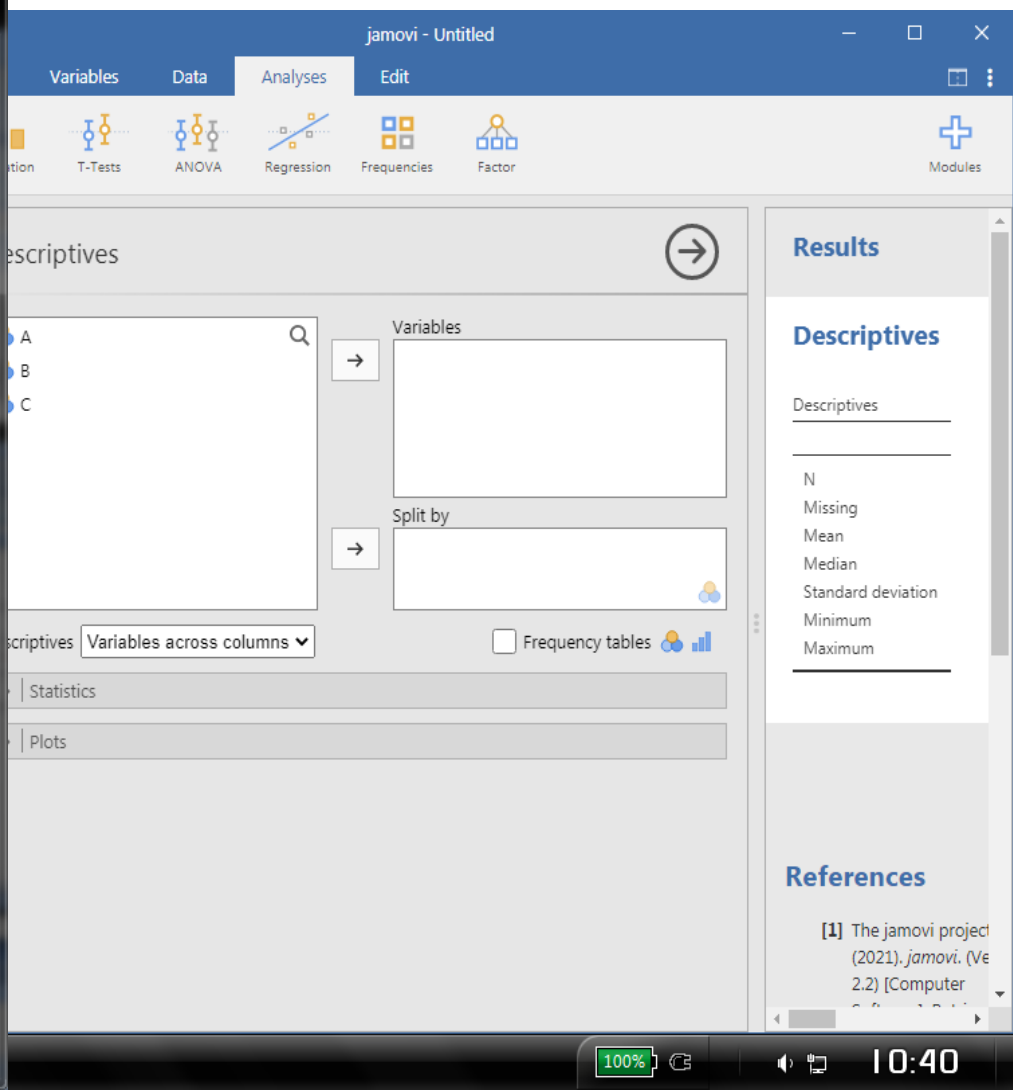
Przykłady

Excel



	A	B	C	D	E	F	G	H	I	J	K
	Lp.	płeć	wiek	wysokość c. [cm]	ciężar c. [kg]	BMI	BMI>25	c.sk.nl	c.roz.nl	puls.nl	c.s.
2	1	k	54	169	79,0	27,66	TAK	99	71	70	1
3	2	k	21	184	98,3	29,03	TAK	135	103	77	1
4	3	k	32	171	77,9	26,64	TAK	120	87	59	1
5	4	k	42	177	86,7	27,67	TAK	137	103	89	1
6	5	k	59	160	74,0	28,91	TAK	153	109	78	1
7	6	k	51	180	96,0	29,63	TAK	153	109	78	1
8	7	k	52	159	70,0	27,69	TAK	135	99	76	1
9	8	k	50	179	79,9	24,94	NIE	179	146	61	1
10	9	k	20	173	67,7	22,62	NIE	172	119	84	1
11	10	k	28	174	71,4	23,58	NIE	209	162	67	1
12	11	k	36	180	69,3	21,39	NIE	156	117	78	1
13	12	k	31	167	66,5	23,84	NIE	149	117	100	1
14	13	k	59	180	65,0	20,06	NIE	145	101	79	1
15	14	k	60	153	48,0	20,5	NIE	135	99	76	1
16	15	k	68	160	55,0	21,48	NIE	145	91	84	1
17	16	k	21	175	68,0	22,2	NIE	157	120	74	1
18	17	k	34	168	58,0	20,55	NIE	167	117	68	1
19	18	k	30	165	52,0	19,1	NIE	145	91	84	1
20	19	k	43	165	55,0	20,2	NIE	157	120	74	1
21	20	k	60	169	67,0	23,46	NIE	167	117	68	1
22	21	m	58	175	77,5	25,31	TAK	170	130	80	1
23	22	m	54	179	82,3	25,69	TAK	174	114	99	1
24	23	m	25	178	82,0	25,88	TAK	169	131	80	1
25	24	m	27	176	78,2	25,57	TAK	129	84	79	1

Jamovi



jamovi - Untitled

Variables Data Analyses Edit

Descriptives T-Tests ANOVA Regression Frequencies Factor Modules

Descriptives

Variables

Split by

Frequency tables

Results

Descriptives

Descriptives

N

Missing

Mean

Median

Standard deviation

Minimum

Maximum

References

[1] The jamovi project (2021). jamovi. (Version 2.2) [Computer software].

https://as.kiero.net

[Przejdź do treści](#)

 [ZALOGUJ](#)

Kiero.net | *lokujemy wyłącznie na Twoją obsługę 4 pliki cookie* | *stosujemy dyrektywy RODO*

[START](#) [CZAT](#)

AMiSNS 2022/2023		Tematów	Postów	Ostatni post
	Informatyka i Biostatystyka 2-semestr Lekarski Moderowane przez: Administrator	3	3	01-03-2023 11:30 przez kiero
	Statystyka medyczna Pielęgniarstwo	4	4	02-03-2023 20:58 przez kiero
	Informatyka i Biostatystyka Ratownictwo	3	3	02-03-2023 20:58 przez kiero



- Nowe treści od Twojej ostatniej wizyty.



- Brak nowych treści od Twojej ostatniej wizyty.

[Szukaj](#)

Dziękuję za uwagę