

PORADNIK STATYSTYCZNY

Uwagi wstępne:

Należy zwrócić uwagę na dokładność podawanych wartości %%

- do 1 miejsca po przecinku, tylko wyjątkowo do 2 (przy bardzo dużej liczebności materiału).
- Średnie i odchylenia standardowe – max do 2, ale w obu przypadkach taka sama dokładność.
- Wartości funkcji testowych – do 2 miejsc.
- Poziomy istotności – podaje się dokładny poziom do 3 miejsc po przecinku ale:
 - jeśli $p > 0,05$ można nie podawać wartości p i tylko napisać **ns** (*non significant*)
 - jeśli cyfra znacząca p jest dopiero na 4. lub dalszym miejscu po przecinku należy napisać $p < 0,001$; zapis $p = 0,000$ jest błędny, bo błąd nigdy nie jest zerowy.
- Istotności można też oznaczyć w postaci poziomów, zapisanych gwiazdkami :
 - * $p < 0,05$ ($1\% < \alpha < 5\%$)
 - ** $p < 0,01$ ($0,1\% < \alpha < 1\%$)
 - *** $p < 0,001$ ($\alpha < 0,1\%$)

UWAGA: Sposób wykonania wszystkich przedstawionych poniżej testów statystycznych został przedstawiony na zajęciach z „podstaw statystyki”, a ponadto szczegółowe instrukcje, jak uzyskać wszelkie zaprezentowane w tym poradniku wyniki znajdują się w skrypcie do ćwiczeń autorstwa profesora K. Boryśławskiego oraz w konspektach do ćwiczeń przesłanych Państwu przez prowadzącego zajęcia.

Jak przeprowadzić poprawną analizę statystyczną danych do pracy dyplomowej?

1. Czynności wstępne

Po zakończeniu zbierania materiału i utworzeniu bazy danych (najlepiej w arkuszu Excel) należy przejść do analizy statystycznej. Pierwszą czynnością, jaką należy wykonać jest przejrzanie bazy danych pod kątem poprawności rekordów (sprawdzenie, czy jest w niej wszystko, co będzie potrzebne do dalszych analiz oraz czy baza nie zawiera błędów lub braków danych, tam gdzie powinny się one znaleźć). Następnie należy zaimportować bazę danych do wybranego programu statystycznego, gdzie będą wykonywane analizy (np. Statistica, PAST, JASP itp.). W przypadku arkusza kalkulacyjnego Statistica, po imporcie danych, pierwszą czynnością jest oczywiście uporządkowanie typów zmiennych, dzieląc je na liczbowe oraz

tekstowe (w tym przypadku należy ustawić opcję „podw. precyzj.” i sprawdzić wartości tzw. etykiet tekstowych). W przypadku zmiennych nominalowych tj. płeć, nie ma większego znaczenia, która wartość tekstowa zostanie skojarzona, z którą etykietą – poziomy tego typu zmiennej są równoważne. Uwagę należy natomiast zwrócić na zmienne tekstowe o charakterze porządkowym (np. kategoria wysokości ciała [niski-średni-wysoki], stopień otyłości [norma-nadwaga-otyłość], ocena z przedmiotu [ndst-dst-db-bdb] itd.). Wówczas wartość etykiet tekstowych ma kolosalne znaczenie!

Przykład:

Etykieta tekstowa	Wartość etykiety tekstowej	
	ŹLE	DOBRCZE
Niski	101	101
Wysoki	102	103
Średni	103	102

Będzie to miało szczególne znaczenia w dalszych etapach analiz w trakcie generowania korelacji/regresji oraz wykresów – program nie rozumie, że wysoki jest większy od średniego, tylko, że 103 jest większe od 102 itd.

Po sprawdzeniu ułożenia etykiet tekstowych dla zmiennych tekstowych porządkowych można zacząć przyglądać się danym.

Dla wszystkich zmiennych ilościowych (ciągłych [interwałowych/ilorazowych]) należy utworzyć histogramy oraz obliczyć wskaźniki asymetrii i kurtozy, aby zobaczyć jak kształtują się rozkłady dla poszczególnych cech. Przy okazji należy zobaczyć czy nie pojawiają się jakieś obserwacje skrajne (ang. *outliers*), które znacząco odbiegają od zakresu zmienności danej cechy. Jeżeli takowe zostaną wykryte należy sprawdzić z czego wynikają – mogą to być błędy w zapisie wyników (np.: dla jednego z uczestników zapisano wartość wysokości ciała 1781 cm, zamiast 178,1 cm co oznaczałoby, że w próbie badawczej mamy olbrzyma 😊), albo rzeczywiste obserwacje odstające (np.: wartość BMI równa 40 – w grupie znalazła się osoba chorobliwie otyła). Takie przypadki należy usunąć z bazy danych, gdyż będą przekłamywać wszystkie dalsze wyniki analiz.

Dla zmiennych jakościowych (porządkowych i nominalowych) należy utworzyć tabele licznosci, aby sprawdzić czy kategorie cech są w miarę równoliczne (będzie to ważne później, podczas wykonywania testów statystycznych, gdzie zmienne te będą pełnić rolę zmiennych niezależnych – szczególnie przy testach parametrycznych, które wymagają równoliczności

porównywanych grup). Można to (mniej więcej) osiągnąć łącząc dwie (lub więcej) podobnych do siebie, sąsiednich) grup w jedną bardziej liczną lub poprzez eliminację grup o znikomej liczności.

Przykład:

Kategoria BMI	Liczność (<i>n</i>)	
	ŹLE	DOBRCZE
Norma	45	45
Nadwaga	32	32
Otyłość I st.	24	24 + 6 = 30
Otyłość II st.	6	

2. Charakterystyka materiału badawczego – licznosci i podstawowe statystyki opisowe

Wycieszczone i uporządkowane dane stanowią podstawę analiz. Najpierw zawsze należy zaprezentować **podstawowe statystyki opisowe**, które zazwyczaj pokazuje się w pierwszym podrozdziale wyników („charakterystyka materiału”). Dla zmiennych jakościowych (tekstowych) należy pokazać odpowiednie tabele licznosci prezentujące liczebność (*n*) oraz odsetek (%), jaki stanowią poszczególne podgrupy względem całego zabranego materiału.

Przykład:

Tabela 1. Tabela licznosci kategorii wykształcenia badanych.

Wykształcenie badanych	<i>n</i>	%
Podstawowe	10	12,5
Zawodowe	15	18,7
Średnie	30	37,5
Wyższe	25	31,3
SUMA	80	100,0

Można przedstawić tabele licznosci również z uwzględnieniem drugiej zmiennej jakościowej, na przykład płci. Jest to bardziej poprawne, ponieważ płeć jest bardzo ważną, specyficzną cechą.

Przykład:

Tabela 2. Tabela liczości kategorii wykształcenia badanych z podziałem na płeć.

Wykształcenie badanych	Mężczyźni		Kobiety		RAZEM
	<i>n</i>	%	<i>n</i>	%	
Podstawowe	6	14,0	4	10,8	10
Zawodowe	10	23,3	5	13,5	15
Średnie	15	34,8	15	40,6	30
Wyższe	12	27,9	13	35,1	25
SUMA	43	100,0	37	100,0	80

Wartości odsetkowe (%) przedstawione w powyższym przykładzie oznaczają odsetki w obrębie płci. Można oczywiście również przedstawić odsetki z całości materiału lecz wtedy porównanie wartości odsetkowych (np. przy wykorzystaniu wsk. struktury) między płciami nie będzie możliwe.

Dla zmiennych ilościowych (ciągłych) należy zaprezentować statystyki opisowe przedstawiające miary przeciętności (tendencji centralnej/położenia), zmienności (rozproszenia/dyspersji), a także miary kształtu rozkładu, tj. średnią, odchylenie standardowe, min-max asymetrię, kurtozę itp. Tabele takie, jak Tabela 3 i Tabela 4 mogą znaleźć się również w rozdziale dotyczącym analizy wyników pracy.

Przykład:

Tabela 3. Statystyki opisowe cech antropometrycznych mężczyzn.

Cecha	<i>n</i>	\bar{X}	<i>s</i>	min	max	<i>As</i>	<i>K</i>
Obwód ramienia lewego [cm]	60	27,3	7,2	22,1	35,2	-0,9	0,6
Obwód ramienia prawego [cm]	58	31,1	9,3	25,1	40,3	0,4	0,6
Obwód uda lewego [cm]	50	50,1	9,3	45,4	62,9	0,4	-0,5
Obwód uda prawego [cm]	45	56,2	11,8	49,2	67,2	0,2	-0,1

Dla powyższej tabeli dane przedstawić można z podziałem na płeć, jeżeli badaniu poddani są zarówno mężczyźni, jak i kobiety.

Przykład:

Tabela 4. Statystyki opisowe cech antropometrycznych z uwzględnieniem płci badanych.

Cecha	Mężczyźni						Kobiety					
	<i>n</i>	\bar{X}	<i>s</i>	min-max	<i>As</i>	<i>K</i>	<i>n</i>	\bar{X}	<i>s</i>	min-max	<i>As</i>	<i>K</i>
Obwód ramienia lewego [cm]	60	27,3	7,2	22,1-35,2	-0,9	0,6	70	22,6	6,3	18,5-26,4	0,6	-0,8
Obwód ramienia prawego [cm]	58	31,1	9,3	25,1-40,3	0,4	0,6	69	25,4	7,7	20,6-29,5	-0,5	0,4
Obwód uda lewego [cm]	50	50,1	9,3	45,4-62,9	0,4	-0,5	60	43,7	8,4	38,9-50,4	0,6	-1,3
Obwód uda prawego [cm]	45	56,2	11,8	49,2-67,2	0,2	-0,1	55	48,7	10,9	40,5-54,0	0,3	-0,6

Jeżeli dla powyższego przykładu wszystkie pomiary zostały wykonane dla wszystkich kobiet i wszystkich mężczyzn wówczas nie trzeba uwzględniać kolumny z liczebnością (*n*) – wystarczy napisać w nagłówkach kolumn z podpisami płci ile było kobiet, a ilu mężczyzn.

3. Analiza materiału - przedstawianie głównych wyników pracy dyplomowej

Analiza danych ma na celu **zweryfikowanie hipotez badawczych** pracy dyplomowej. Najważniejszym zadaniem przed wykonaniem analiz jest wybranie właściwych testów statystycznych pozwalających odpowiedzieć na pytania badawcze. Przypominam, że testy dzielą się w uproszczeniu na dwie grupy:

- parametryczne (za ich pomocą łatwiej jest odrzucić hipotezę zerową, czyli wykryć istotne zjawisko, jeżeli takie w populacji ogólnej występuje),
 - nieparametryczne (trudniej jest za ich pomocą odrzucić hipotezę zerową, tzn. można nie wykryć zjawiska występującego rzeczywiście w populacji, nawet jeżeli w istocie ono istnieje).
- Zawsze w pierwszej kolejności należy rozważyć możliwość zastosowania testów parametrycznych. Trzeba jednakże pamiętać, że **testy parametryczne** wymagają spełnienia kilku warunków. Są to zazwyczaj:

- a. ilościowy charakter zmiennej zależnej

b. duża liczność każdej z porównywanych podgrup (w zależności od testu jest to zazwyczaj ~25-30 przypadków/grupę).

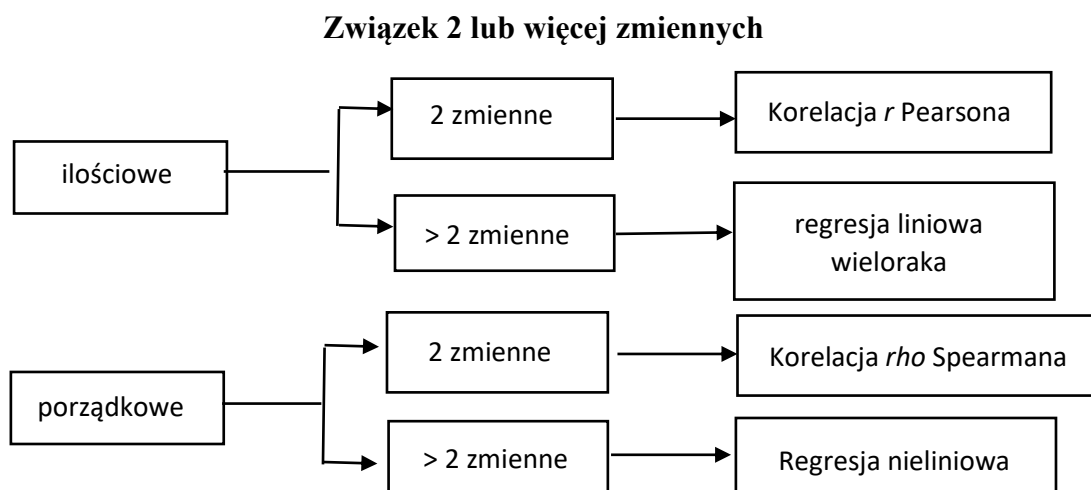
c. jednorodne (homogeniczne) wariancje w porównywanych grupach (czyli podobna zmienność dla porównywanych podgrup)

d. rozkład zmiennej zależnej zbliżony do rozkładu normalnego (Gaussa) w każdej z porównywanych podgrup

W związku z tym zawsze należy sprawdzić czy spełnione są te warunki. Warunki a i b są łatwe do sprawdzenia nawet dla laika. Problem pojawia się natomiast przy sprawdzaniu warunków c i d. Aby zweryfikować homogeniczność wariancji należy użyć testu **Levene’a** lub **Browna-Forsytha**. Jeżeli chodzi zaś o sprawdzenie normalności rozkładu to z pomocą przychodzi np. test **Shapiro-Wilka**. Przypominam, że w przypadku testów badających jednorodność wariancji i normalność rozkładu jest zupełnie na odwrót niż dla wszystkich innych testów statystycznych – cieszyć powinny wyniki nieistotne statystycznie ($p > 0.05$). Oznaczają one bowiem odpowiednio, jednorodne wariancje oraz rozkład normalny.

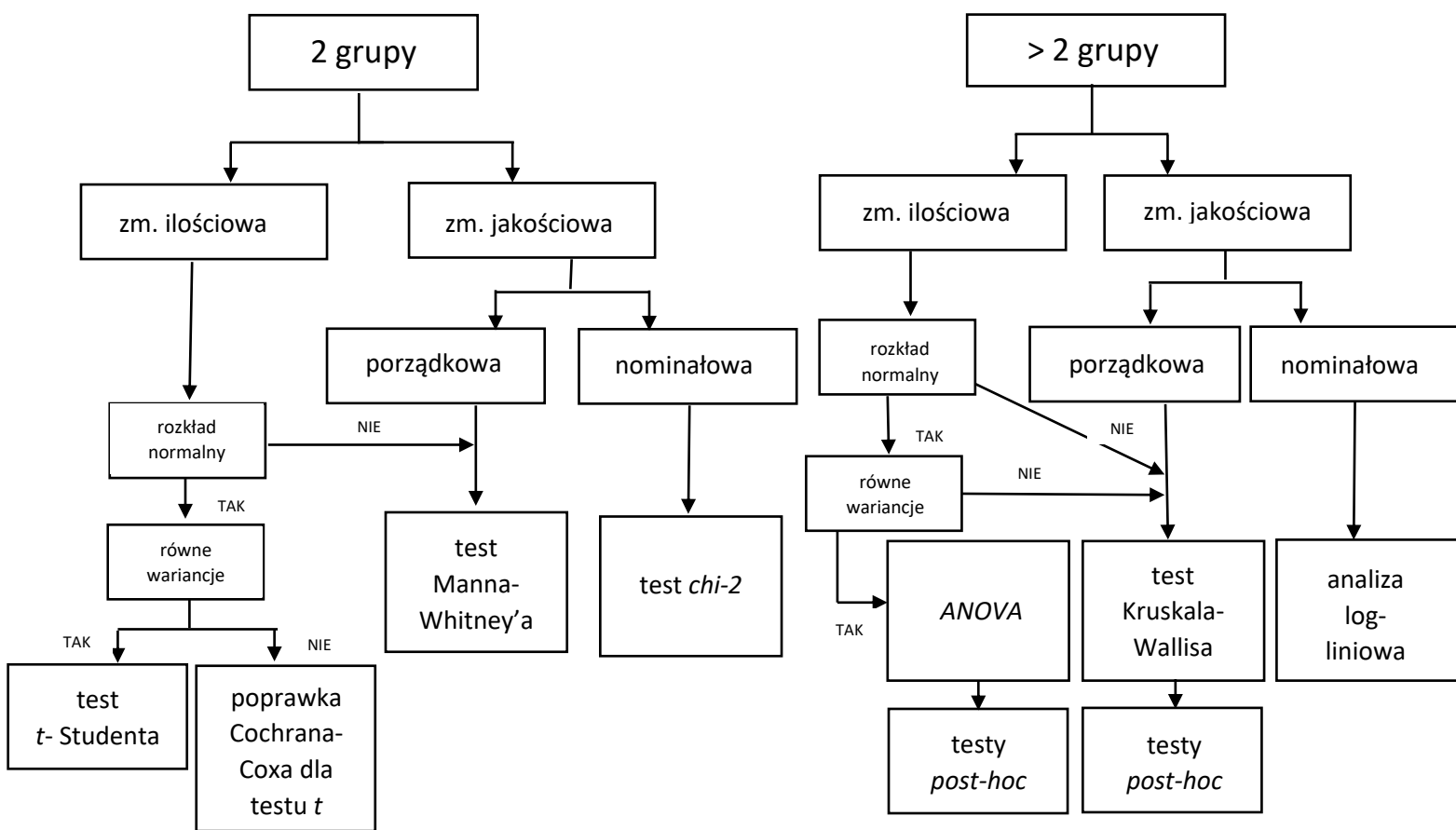
Jeżeli wszystkie warunki (a-d) są spełnione można stosować testy parametryczne. Jeżeli natomiast dane nie spełniają któregokolwiek z warunków trzeba zastosować testy nieparametryczne.

Zasadnicze pytanie – **który test kiedy zastosować** zależy od celu pracy. Poniżej kilka schematów ułatwiających wybór testu.



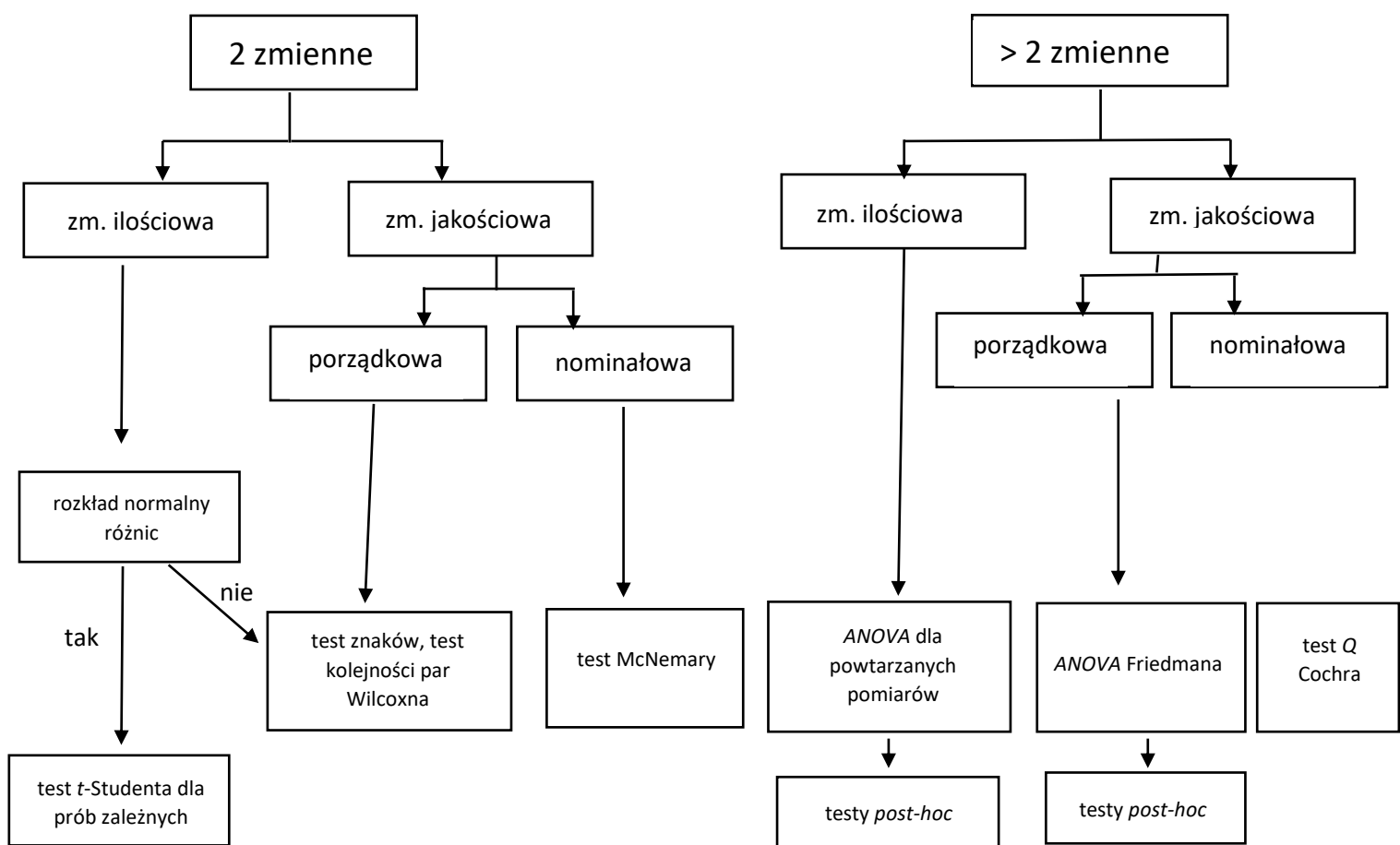
Ryc. 1. Wzajemne związki (korelacje) cech ilościowych i porządkowych (Źródło: Opracowanie własne).

Porównanie 2 lub więcej grup – próby niezależne



Ryc. 2. Schemat wyboru testów dla prób niezależnych (Źródło: Opracowanie własne).

Porównanie 2 lub więcej grup (zmiennych) – próby zależne



Ryc. 3. Schemat wyboru testów statystycznych dla prób zależnych (Źródło: Opracowanie własne).

I. Analiza korelacji i korelacji cząstkowych

SPOSÓB 1

Tabela 5. Korelacje r Pearsona siły dynamometrycznej dominującej kończyny górnej z cechami antropometrycznymi badanych.

Cecha	r	p
Wysokość ciała	0,15	ns
Masa ciała	0,82	0,012
Obwód ramienia	0,91	< 0,001

ns – wartość nieistotna statystycznie

Powyższą tabelę powinno się jednak pokazać z uwzględnieniem podziału na płeć. Wartości wsp. korelacji mogą być zupełnie inne u obu płci.

Tabela 5'. Korelacje r Pearsona siły dynamometrycznej dominującej kończyny górnej z cechami antropometrycznymi badanych mężczyzn i kobiet.

Cecha	Mężczyźni ($n=90$)		Kobiety ($n=60$)	
	r	p	r	p
Wysokość ciała	0,15	ns	0,05	ns
Masa ciała	0,86	0,011	0,74	0,024
Obwód ramienia	0,96	< 0,001	0,78	0,017

ns – wartość nieistotna statystycznie

SPOSÓB 2

Tabela 6. Korelacje r Pearsona dla parametrów fizycznych mówców i parametrów akustycznych ich głosu.

	F0	Głośność	Df
Wiek	-0,42***	-0,25*	-0,09
Wysokość ciała	-0,21*	0,18	-0,18
Masa ciała	-0,33**	0,40***	-0,26*

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Powyższą tabelę należy jednak pokazać z podziałem na płeć, gdyż głos jest cechą o silnym dymorfizmie płciowym i uśrednianie parametrów głosu dla kobiet i mężczyzn nie ma biologicznego sensu.

Tabela 6'. Korelacje r Pearsona dla parametrów fizycznych kobiet i mężczyzn i parametrów akustycznych ich głosu.

	F0	Głośność	Df
Mężczyźni (n=30)			
Wiek	-0,42***	-0,25*	-0,09
Wysokość ciała	-0,21*	0,18	-0,18
Masa ciała	-0,33**	0,40***	-0,26*
Kobiety (n=45)			
Wiek	-0,38***	-0,21*	-0,01
Wysokość ciała	-0,18	0,1	-0,11
Masa ciała	-0,3**	0,35**	-0,24*

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

SPOSÓB 3

Tabela 7. Macierz kwadratowa korelacji r Pearsona zdolności motorycznych badanych.

	Rzut piłką w przód	Skok w dal	Bieg na 100 m	Bieg na 400 m
Rzut piłką w przód	-	0,47**	0,24*	0,18
Skok w dal	0,47**	-	0,58***	0,51**
Bieg na 100 m	0,24*	0,58***	-	0,88***
Bieg na 400 m	0,18	0,51**	0,88***	-

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

lub można prościej

Tabela 7'. Macierz kwadratowa korelacji r Pearsona zdolności motorycznych badanych.

	Rzut piłką w przód	Skok w dal	Bieg na 100 m
Skok w dal	0,47**	-	-
Bieg na 100 m	0,24*	0,58**	-
Bieg na 400 m	0,18	0,51**	0,88***

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Te korelacje, jak i wszystkie inne dane powinno się jednak (o czym wspomniano już wcześniej) prezentować z podziałem na płeć

Tabela 7''. Macierz kwadratowa korelacji r Pearsona zdolności motorycznych mężczyzn i kobiet.

	Kobiety	Rzut piłką w przód	Skok w dal	Bieg na 100 m	Bieg na 400 m
Mężczyźni					
Rzut piłką w przód		-	0,33*	0,15	0,12
Skok w dal		0,47**	-	0,48**	0,43**
Bieg na 100 m		0,24*	0,58***	-	0,75***
Bieg na 400 m		0,18	0,51**	0,88***	-

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Dla korelacji ρ Spearmana oraz τ Kendalla można sporządzić analogiczne tabele.

W przypadku korelacji cząstkowych dane należy prezentować tak samo, jak w przypadku korelacji 0-rzędu, opisując jedynie w tytule tabeli, którego rzędu jest to korelacja cząstkowa i jaką zmienną lub jakie zmienne uwzględnia.

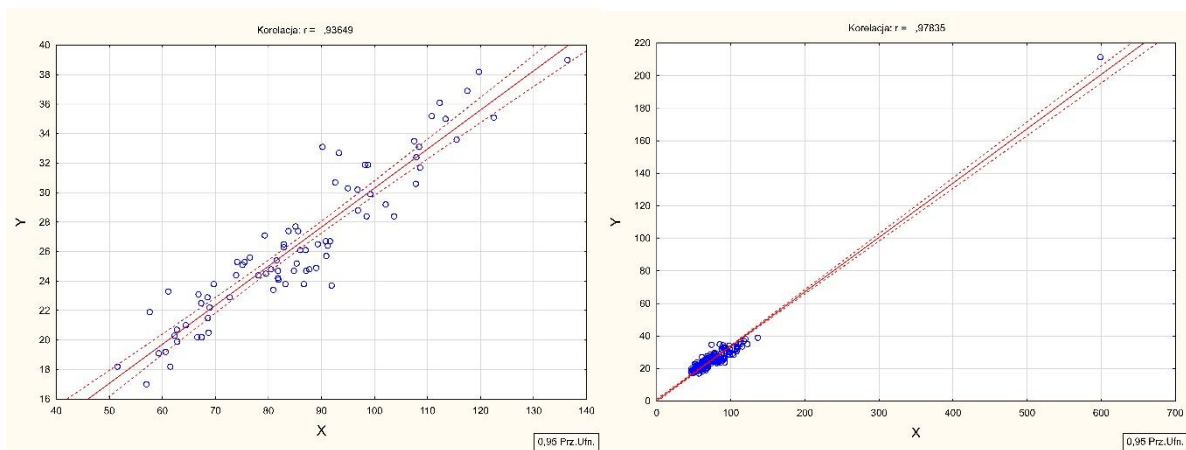
Tabela 8. Korelacja cząstkowa III-rzędu wysokości ciała, masy ciała, wieku i BMI z częstotliwością podstawową (F0).

	[5] F0
[1] Wysokość ciała	$r_{15,234} = -0,08$
[2]Wiek	$r_{25,134} = -0,23^*$
[3] Masa ciała	$r_{35,124} = -0,16$
[4] BMI	$r_{45,123} = -0,11$

* $p < 0,05$

np. $r_{15,234}$ oznacza korelację cząstkową III-rzędu wysokości ciała [1] z częstotliwością podstawową głosu (F0, [5]) z wyłączeniem wieku [2], masy ciała [3] i BMI [4]. Przy tego typu tabelach z korelacjami cząstkowymi ZAWSZE należy wyjaśnić, która cecha ma jaki numer.

WAŻNE: Dla istotnych statystycznie korelacji zawsze należy sprawdzić **wykres rozrzutu** (diagram korelacyjny), aby upewnić się, że nie są to korelacje fałszywe wynikające z występowania w próbie wartości odstających lub innego niż liniowy układu „chmury” punktów.



DOBRE

ŹLE

Ryc. 4. Wykres rozrzutu cechy Y względem cechy X.

II. Analiza związków cech nominalowych – test chi-kwadrat

Tabela 9. Samoocena atrakcyjności fizycznej badanego a ocena atrakcyjności fizycznej jego/jej partnera(-ki) [*Lo*– liczności obserwowane, *Lo-Le*– liczności resztowe].

Samoocena atrakcyjności		Ocena atrakcyjności partnera(-ki)			Σ wiersza
		niska	średnia	wysoka	
niska	% z wiersza	37,3	39,2	23,5	395
	<i>Lo</i>	147	155	93	
	<i>Lo-Le</i> *	14,67	31,29	-45,96	
średnia	% z wiersza	34,2	29,5	36,3	397
	<i>Lo</i>	136	117	144	
	<i>Lo-Le</i>	3,00	-7,33	4,33	
wysoka	% z wiersza	29,1	25,3	45,6	399
	<i>Lo</i>	116	101	182	
	<i>Lo-Le</i>	-17,67	-23,96	41,63	
Σ kolumny		399	373	419	1191

$$\chi^2=23,34, df=4, p=0,006$$

* Jeżeli wartość testu *chi-kwadrat* jest nieistotna ($p > 0,05$), wówczas nie trzeba przedstawiać w tabeli licznosci resztowych (*Lo-Le*). Ostateczna decyzja należy jednakże do opiekuna pracy!

III. Porównanie wartości odsetkowych – wskaźniki struktury

Tabela 10. Porównanie częstości występowania wybranego koloru oczu u mężczyzn i kobiet.

Kolor oczu	Mężczyźni		Kobiety		Poziom istotności różnicy częstości
	<i>n</i>	% z kolumny	<i>n</i>	% z kolumny	
brązowe	35	57,3	32	46,4	ns
niebieskie	17	27,9	26	37,7	0,024
zielone	9	14,8	11	15,9	ns
SUMA	61	100,0	69	100,0	

ns – wartość nieistotna statystycznie

IV. Porównanie średnich arytmetycznych dwóch prób niezależnych (2 grup)

IV.1. Test *t*-Studenta dla zmiennej zależnej ciągłej, gdy rozkład w obu porównywanych grupach jest normalny, a wariancje homogeniczne.

Tabela 11. Statystyki opisowe badanych cech w zależności od płci.

Cecha	Płeć				Istotność różnicy	
	Mężczyźni (<i>n</i> =300)		Kobiety (<i>n</i> =450)			
	\bar{X}	<i>s</i>	\bar{X}	<i>s</i>	<i>t</i>	<i>p</i>
wysokość ciała [cm]	182,3	15,6	165,1	10,6	2,89	0,022
obwód szyi [cm]	33,5	5,8	31,2	4,6	1,24	ns
obwód talii [cm]	100,5	25,7	85,7	17,8	3,05	0,013
grubość fałdu podłopatkowego [mm]	12,4	4,9	18,7	6,2	3,46	< 0,001

ns – różnica nieistotna statystycznie

IV.2. Test *t*-Studenta z poprawką na niezależną estymację wariancji (**poprawka Cochran-Coxa**) dla zmiennej zależnej ciągłej, gdy rozkład w obu porównywanych grupach jest normalny, ale wariancje nie są jednorodne.

Należy przyjąć ten sam wzór jak w tabeli 11.

IV.3. Test *U* Manna-Whitney'a. Porównanie dwóch median dla zmiennej zależnej porządkowej.

Tabela 12. Przeciętna ocena atrakcyjności głosu badanego przez słuchaczy płci przeciwnej w zależności od płci mówcy.

Cecha	Płeć						Istotność różnicy	
	Mężczyźni (<i>n</i> =300)			Kobiety (<i>n</i> =450)				
	<i>R</i> *	<i>Me</i>	<i>Q1-Q3</i>	<i>R</i>	<i>Me</i>	<i>Q1-Q3</i>	<i>U/Z</i> **	<i>p</i>
Ocena atrakcyjności głosu mówcy (1-5)	1125,6	3	2-4	1113,5	4	3-5	3,4	0,015

R – średnia rang

UWAGI

* Statistica podaje wartość sumy rang. Aby wyliczyć średnią rang (*R*) należy podzielić podaną przez program wartość przez licznosc odpowiedniej grupy.

** W zależności od liczności porównywanych grup podajemy wartość U (gdy w każdej z grup $n < 20$) i odpowiedni poziom p (Statistica nazywa to „2*1 str. dokł p”) lub Z , gdy licznosci są duże ($n > 20$) i właściwą wartość p (zawsze podajemy to co Statistica nazywa $Z_{\text{popr.}}$ oraz p stojący zaraz za tą wartością).

IV. 4. Test U Manna-Whitney’a. Porównanie dwóch średnich dla zmiennej zależnej ciągłej, gdy rozkład w co najmniej jednej z porównywanych grup nie jest rozkładem normalnym.

Tabela 13. Wartości cech antropometrycznych w zależności od płci badanego.

Cecha	Płeć						Istotność różnicy	
	Mężczyźni ($n=300$)			Kobiety ($n=450$)				
	R	\bar{X} *	s	R	\bar{X}	s	U/Z	p
wysokość ciała [cm]	112,5	182,3	15,6	98,9	165,1	10,6	2,3	0,032
obwód szyi [cm]	146,7	33,5	5,8	142,4	31,2	4,6	0,8	ns

R – średnia rang, ns – wynik nieistotny statystycznie

UWAGI

*dla zmiennych ciągłych, które nie spełniają warunku normalności rozkładu w obu porównywanych grupach można zaprezentować średnią arytmetyczną oraz odchylenie standardowe, zamiast mediany i rozstępu kwartylowego (choć podanie tych wartości w tej sytuacji jest rozwiązaniem jak najbardziej prawidłowym), ale konieczne jest sprawdzenie najpierw z czego wynika brak normalności rozkładu z grupie (grupach). Jeżeli jest to wynikiem skośności rozkładu, to należy zobaczyć czy kierunek skośności rozkładów w porównywanych grupach jest taki sam (np. rozkład wysokości ciała u kobiet i mężczyzn jest lewoskośny). Jeżeli tak jest, to można zaprezentować wartości średnich arytmetycznych, bo przekłamanie tej miary położenia powinno być +/- podobne dla obydwu grup. Inaczej sytuacja się ma, gdy jedna z grup ma rozkład lewo-, a druga prawoskośny lub kiedy jedna z grup cechuje się rozkładem skośnym, a druga normalnym; wówczas wartości średniej arytmetycznej nie będą rzetelnymi miarami tendencji centralnej – wtedy najlepiej będzie pokazać mediany oraz rozstępy kwartyłowe, aby uniknąć zniekształcenia różnic w grupach.

V. Porównanie średnich arytmetycznych dwóch prób zależnych (2 zmiennych dla tych samych przypadków).

V.1. Test *t*-Studenta - pomiary powtarzane dla zmiennej zależnej ciągłej, gdy rozkład różnic wartości pomiarów jest rozkładem normalnym.

Tabela 14. Wartości parametrów fizjologicznych badanych mężczyzn i kobiet przed i po treningu na cykloergometrze.

Parametr fizjologiczny	Przed wysiłkiem		Po wysiłku		Istotność różnicy		
	\bar{x}	<i>s</i>	\bar{x}	<i>s</i>	Bezwzględna różnica średnich	<i>t</i>	<i>p</i>
mężczyźni (<i>n</i> =300)							
tętno [ud./min]	70	22	95	10	25	2,18	0,032
ciśnienie skurczowe [mmHg]	120	35	155	20	35	2,34	0,028
ciśnienie rozkurczowe [mmHg]	80	30	100	19	20	2,56	0,019
kobiety (<i>n</i> =450)							
tętno [ud./min]	75	28	105	18	30	2,29	0,03
ciśnienie skurczowe [mmHg]	122	41	163	29	41	2,61	0,017
ciśnienie rozkurczowe [mmHg]	84	35	115	24	31	3,05	0,012

V.2. Test kolejności par Wilcozona - pomiary powtarzane dla zmiennej zależnej ciągłej, gdy rozkład różnic wartości pomiarów nie jest rozkładem normalnym.

Tabela 15. Wartości parametrów fizjologicznych badanych kobiet przed i po treningu na cykloergometrze.

Parametr fizjologiczny	Przed wysiłkiem		Po wysiłku		Bezwzględna różnica średnich	Istotność różnicy	
	\bar{x}	<i>s</i>	\bar{x}	<i>s</i>		<i>T/Z</i> *	<i>p</i>
tętno [ud./min]	73	24	98	15	25	3,2	< 0,001
ciśnienie skurczowe [mmHg]	121	39	160	23	39	2,5	0,019
ciśnienie rozkurczowe [mmHg]	82	34	110	26	28	2,9	0,011

*wartość T podajemy, gdy licznosc grupy jest mala ($n < 25$), natomiast wartosc Z w przypadku grup wiekszych ($n > 25$).

V.3. Test znaków - pomiary powtarzane dla zmiennej porządkowej.

Tabela 16. Ocena atrakcyjności zapachu osoby płci przeciwnej przed i po zaaplikowaniu właściwego feromonu w okolicy szyi dla kobiet i mężczyzn.

Ocena atrakcyjności zapachu (1-7)	Przed zaaplikowaniem		Po zaaplikowaniu		Różnica	Istotność różnicy	
	Me	Q1-Q3	Me	Q1-Q3		Z	p
Mężczyźni ($n=100$)	4	1-5	6	2-7	2	3,2	< 0,001
Kobiety ($n=189$)	5	2-7	7	4-7	2	3,6	< 0,001

VI. Porównanie więcej niż dwóch prób niezależnych

VI.1. Porównanie więcej niż dwóch średnich dla zmiennej zależnej ciągłej, gdy rozkład we wszystkich porównywanych grupach jest normalny a wariancje homogeniczne – analiza wariancji **ANOVA** (test F)

a) jednoczynnikowa analiza wariancji **ANOVA**

Tabela 17. Średni czas spędzany w pozycji siedzącej w ciągu doby, samopoczucie i wskaźnik aktywności fizycznej u osób o różnym chronotypie.

Cecha	Chronotyp						Istotność różnicy	
	Poranny ($n=18$)		Pośredni ($n=121$)		Wieczorny ($n=36$)			
	\bar{X}	s	\bar{X}	s	\bar{X}	s	F	p
Czas w pozycji siedzącej [min]	343,7	205,6	370,8	181,3	473,4	207,6	4,56	0,012
Samopoczucie (punkty)	13,5	4,3	10,7	4,5	8,3	4,9	9,7	< 0,001
Wsk. aktywności fizycznej (punkty)	4340,6	3275,9	2831,8	3253,9	3175,4	4497,7	1,44	ns

ns – wynik nieistotny statystycznie

Dla wyników istotnych statystycznie należy zaprezentować odpowiednie **testy post-hoc**.

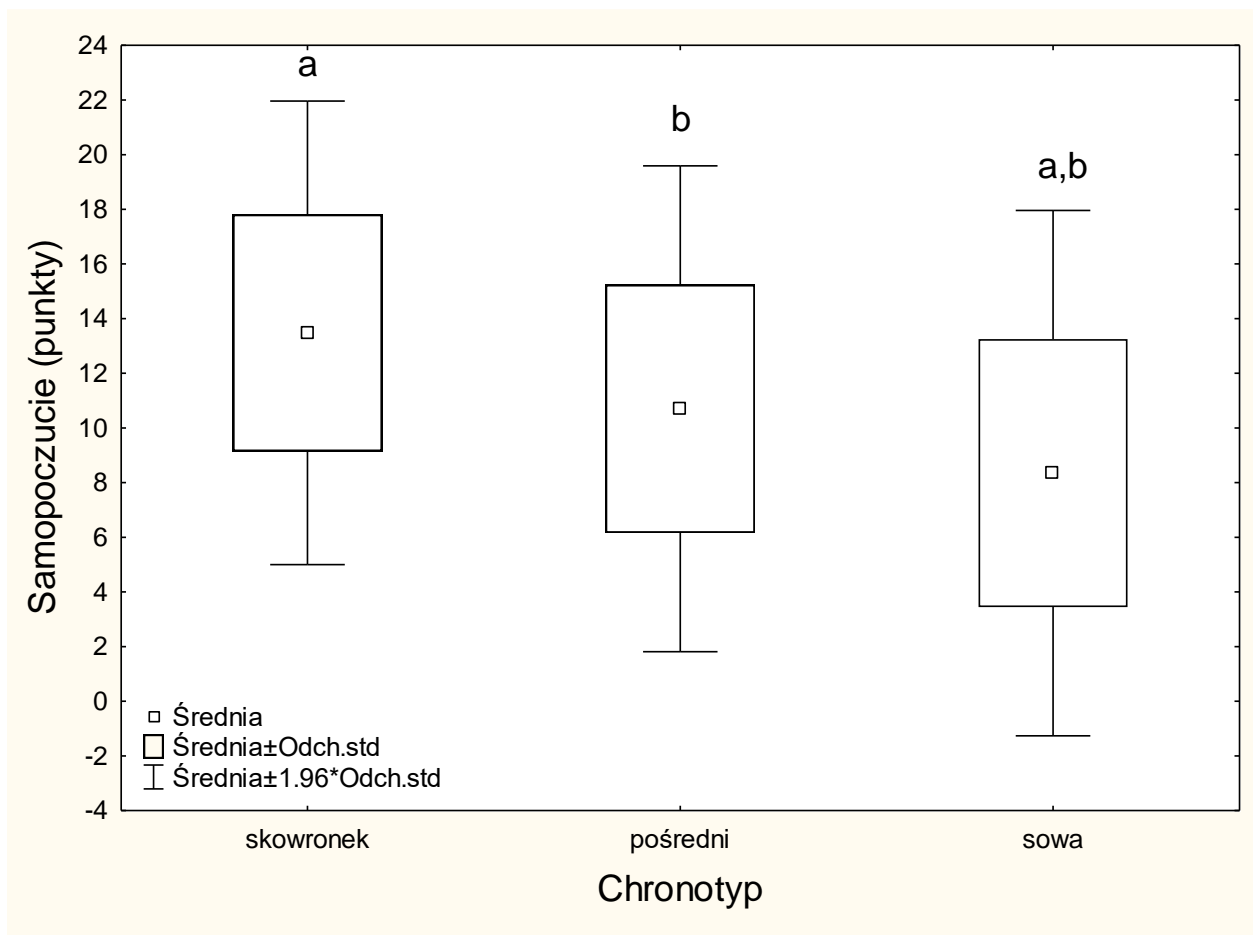
Można to zrobić w formie wykresu średnia-błądy lub w formie tabeli:

Tabela 18. Chronotyp a samopoczucie. Test *post-hoc* RIR Tukey'a dla nierównych liczości.

Chronotyp	{1} $\bar{X} = 13,5$	{2} $\bar{X} = 10,7$	{3} $\bar{X} = 8,4$
poranny{1}	-		
pośredni{2}	ns	-	
wieczorny {3}	< 0,001	0,046	-

ns – wynik nieistotny statystycznie

Wyniki testów *post-hoc* zamiast w postaci tabeli, jak pokazano to w Tabeli 19, przedstawić można za pomocą wykresów (np. ramka-wąsy).



Ryc. 5. Różnice samopoczucia pomiędzy osobami reprezentującymi odmienny chronotyp. Wyniki test *post-hoc* RIR Tukey'a: a: $p < 0,001$; b: $p = 0,046$.

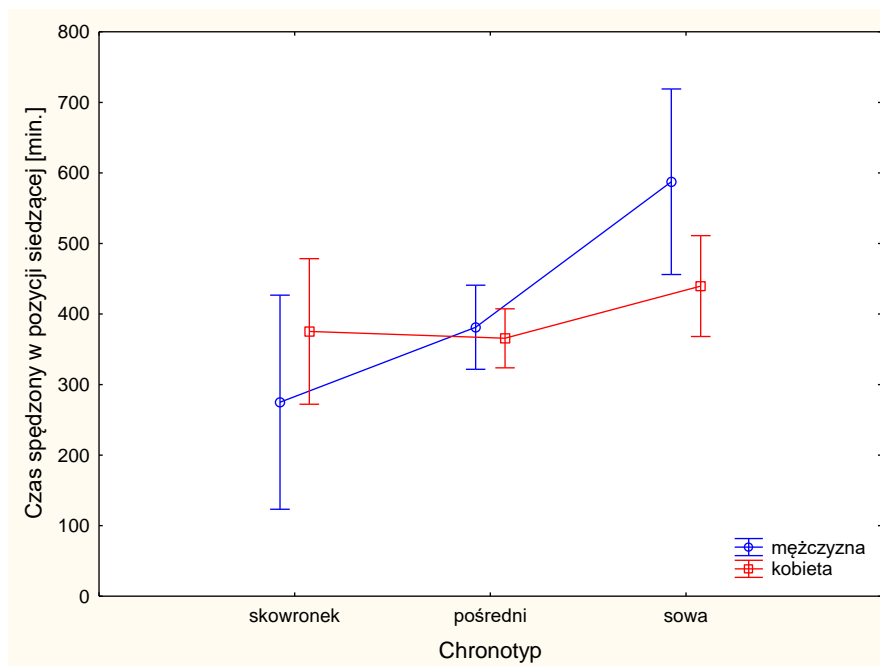
b) dwuczynnikowa analiza wariancji ANOVA

Tabela 19. Średni czas spędzany w pozycji siedzącej w ciągu doby w zależności od płci i chronotypu badanych.

	<i>F</i>	<i>df1, df2*</i>	<i>p</i>
Wyraz wolny	373,3	1,166	< 0,001
Płeć	0,3	1,166	ns
Chronotyp	6,7	2,166	0,002
Płeć*Chronotyp	2,3	2,166	ns

ns – wynik nieistotny statystycznie

**df1* i *df2* to liczba stopni swobody dla wariancji wewnątrz- i międzygrupowej. *df1* to jest to, co program Statistica podaje dla dwuczynnikowej ANOVA jako „stopnie swobody”, natomiast *df2* program podaje jako „błąd” (całkowita licznosc próby – liczba porównywanych kategorii). Istotne statystycznie czynniki z powyższego przykładu można następnie przedstawić pokazując średnią i odchylenie standardowe dla każdej z kategorii (grup), tak jak zaprezentowano to w przykładzie tabeli dla jednoczynnikowej analizy wariancji ANOVA (Tabela 17). Podobnie w przypadku testów *post-hoc* dla dwuczynnikowej ANOVA można wzorować się na przykładach podanych dla analizy wariancji jednoczynnikowej, dodatkowo prezentując efekt interakcji (płeć*chronotyp) za pomocą wykresu interakcji:



Ryc. 6. Czas spędzony w pozycji siedzącej w zależności od płci i chronotypu.

VI. 2. Porównanie więcej niż dwóch średnich dla zmiennej zależnej ciągłej, gdy rozkład jest normalny we wszystkich porównywanych grupach, ale wariancje nie są homogeniczne – test **Welcha**

Tabela 20. Średni czas spędzany w pozycji siedzącej w ciągu doby u osób o różnym chronotypie.

Cecha	Chronotyp						Istotność różnicy	
	Poranny (n=18)		Pośredni (n=121)		Wieczorny (n=36)		F	p
	\bar{X}	s	\bar{X}	s	\bar{X}	s		
Czas w pozycji siedzącej [min]	343,7	205,6	370,8	181,3	473,4	207,6	3,91	0,029

Testy *post-hoc* dla wyników istotnych w teście Welcha przedstawia się analogicznie, jak w przypadku jednoczynnikowej ANOVA

VI. 3. Porównanie więcej niż dwóch średnich dla zmiennej zależnej ciągłej, gdy rozkład nie jest normalny w co najmniej jednej z porównywanych grup – test **H Kruskala-Wallis**

Tabela 21. Średni czas spędzany w pozycji siedzącej w ciągu doby, samopoczucie i wskaźnik aktywności fizycznej u osób o różnym chronotypie.

Cecha	Chronotyp									Istotność różnicy	
	Poranny (n=18)			Pośredni (n=121)			Wieczorny (n=36)			H	p
	R	\bar{X} *	s	R	\bar{X}	s	R	\bar{X}	s		
Czas w pozycji siedzącej [min]	72,3	343,7	205,6	82,9	370,8	181,3	106,3	473,4	207,6	7,82	0,021
Samopoczucie (punkty)	143,2	13,5	4,3	109,2	10,7	4,5	80,0	8,3	4,9	16,46	<0,001
Wsk. aktywności fizycznej (punkty)	122,5	4340,6	3275,9	85,6	2831,8	3253,9	78,8	3175,4	4497,7	9,81	0,007

R – średnia rang

*w przypadku zmiennych ciągłych, dla których stosujemy test Kruskala-Wallisa, jako miary położenia i rozproszenia można przedstawić średnią arytmetyczną oraz odchylenie standardowe, pod warunkiem, że brak normalności rozkładu nie wynika z jego skośności lub gdy asymetria rozkładów jest taka sama dla każdej z porównywanych grup (np. we wszystkich kategoriach chronotypu rozkład zmiennej „samopoczucie” jest lewoskośny). W przeciwnym przypadku (gdy dla jednych grup rozkłady są lewo- a dla innych prawoskośne) należy zamiast średniej i odchylenia standardowego podawać wartość mediany i rozstępu kwartylowego ($Q1-Q3$).

VI. 4. Porównanie więcej niż dwóch grup dla zmiennej zależnej porządkowej – test *H* Kruskala-Wallisa

Tabela 22. Ocena atrakcyjności męskiego głosu i twarzy przez kobiety w różnych fazach cyklu menstruacyjnego.

Cecha	Faza cyklu menstruacyjnego									Istotność różnicy	
	przedowulacyjna (n=186)			owulacyjna (n=89)			poowulacyjna (n=163)				
	<i>R</i>	<i>Me</i>	<i>Q1-Q3</i>	<i>R</i>	<i>Me</i>	<i>Q1-Q3</i>	<i>R</i>	<i>Me</i>	<i>Q1-Q3</i>	<i>H</i>	<i>p</i>
Ocena atrakcyjności głosu (1-10)	122,5	7	6-7	178,9	7	6-8	155,6	4	4-5	15,65	<0,001
Ocena atrakcyjności twarzy (1-10)	112,5	6	3-8	156,8	9	5-9	146,7	7	5-8	20,11	<0,001

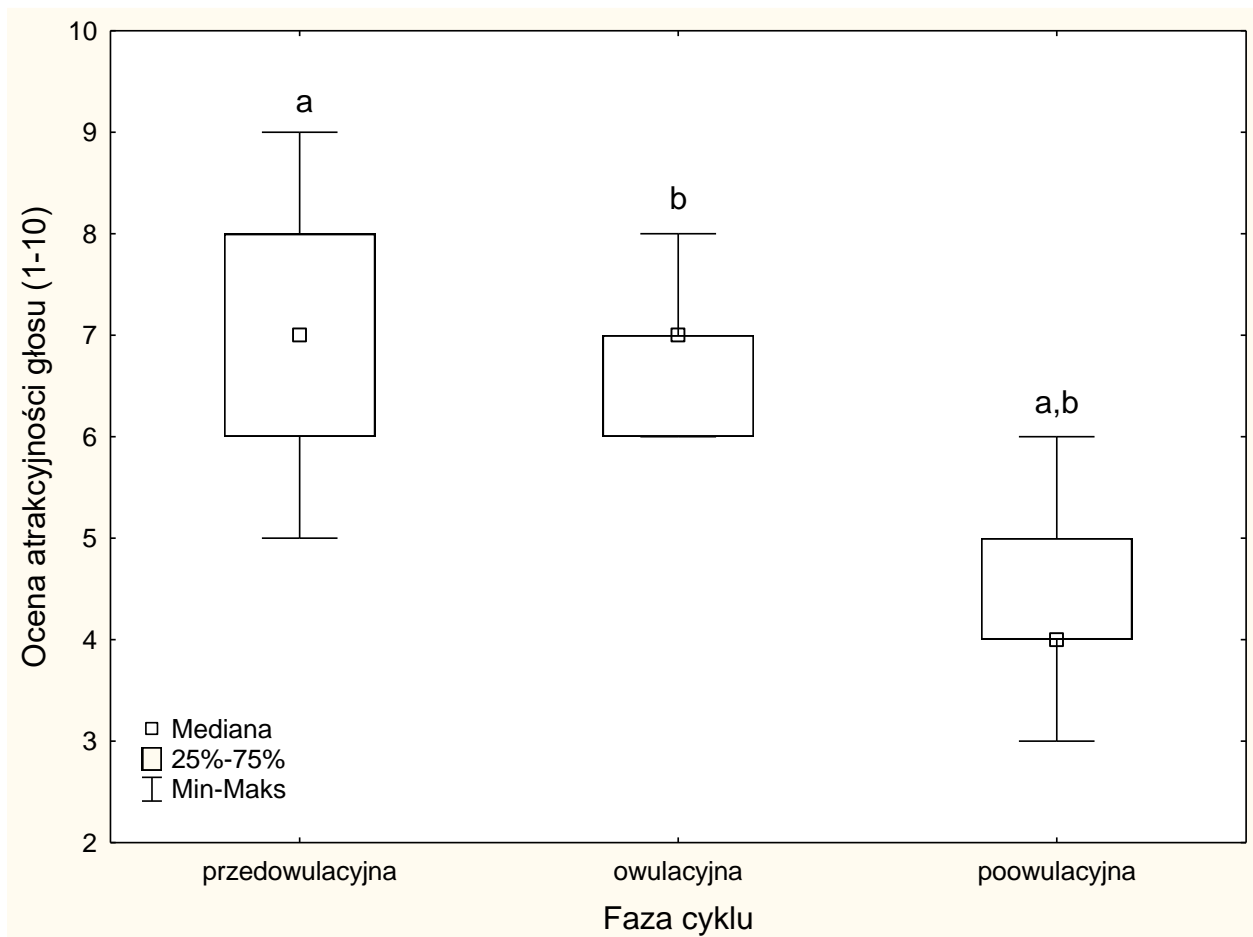
R – średnia rang

Dla testu Kruskala-Wallisa stosuje się test *post-hoc* Dunna. Można go pokazać w postaci wykresu ramka-wąsy lub tabeli:

Tabela 23. Ocena atrakcyjności męskiego głosu przez kobiety w różnych fazach cyklu miesięcznego. Test Dunna.

Faza cyklu menstruacyjnego	{1} <i>Me</i> = 7	{2} <i>Me</i> = 7	{3} <i>Me</i> = 4
przedowulacyjna {1}	-		
owulacyjna {2}	ns	-	
poowulacyjna {3}	0,037	0,013	-

ns – wynik nieistotny statystycznie



Ryc. 6. Różnice w ocenie atrakcyjności męskiego głosu przez kobiety w różnych fazach cyklu menstruacyjnego. Wyniki test *post-hoc* Dunna: a: $p=0,037$, b: $p=0,013$.

VII. Porównanie więcej niż dwóch prób zależnych

VII. 1. Porównanie więcej niż dwóch pomiarów powtarzanych dla zmiennej zależnej ciągłej – ANOVA dla powtarzanych pomiarów

Tabela 24. Procentowa zawartość tkanki tłuszczowej u 30 pacjentów poddanych dietoterapii w kolejnych etapach kuracji.

Cecha	Czas stosowania dietoterapii								Istotność różnicy	
	Na początku		Po 1 tygodniu		Po 2 tygodniach		Po miesiącu			
	\bar{X}	s	\bar{X}	s	\bar{X}	s	\bar{X}	s	F	p
Zawartość tkanki tłuszczowej [%]	22,8	7,0	22,7	7,0	22,1	6,9	20,6	6,8	696,99	< 0,001

Testy *post-hoc* dla istotnych wartości testu F dla pomiarów powtarzanych należy przedstawiać analogicznie, jak dla ANOVA dla prób niezależnych (w formie tabeli lub wykresu średnia-błędy). Jeżeli badania mają charakter badań laboratoryjnych z grupą kontrolną (tak jak w przykładzie powyżej, gdzie za grupę kontrolną można uznać osoby zmierzone pod kątem % zawartość tkanki tłuszczowej na początku dietoterapii) to dodatkowo można zastosować test *post-hoc* Dunnetta, który charakteryzuje się tym, że porównuje pomiary z grup badanych (w tym przypadku pomiary tkanki tłuszczowej po tygodniu, 2 tygodniach oraz miesiącu) tylko i wyłącznie do grupy kontrolnej, nie wykonując wzajemnych porównań pomiędzy nimi (czyli każda grupa z każdą, jak robi to na przykład test Tukey'a czy Scheffe'go). Test Dunnetta można przedstawić w formie tabeli lub wykresu średnia-błędy.

Tabela 25. Porównanie początkowej zawartości tkanki tłuszczowej (grupa kontrolna) z zawartością tłuszczu po 1, 2 i 4 tygodniach stosowania dietoterapii (grupy badane). Test Dunnetta.

Czas stosowania dietoterapii	Na początku $\bar{X} = 22,8$
Po 1 tygodniu $\bar{X} = 22,7$	ns
Po 2 tygodniach $\bar{X} = 22,1$	< 0,001
Po miesiącu $\bar{X} = 20,6$	< 0,001

ns – wynik nieistotny statystycznie

VII. 2. Porównanie więcej niż dwóch pomiarów powtarzanych dla zmiennej zależnej porządkowej – **ANOVA Friedmana**

Tabela 26. Ocena zadowolenia z życia pacjentów w kolejnych miesiącach uczęszczania na zajęcia w ramach psychoterapii.

Cecha	Czas stosowania psychoterapii												Istotność różnicy	
	Na początku			Po 1 miesiącu			Po 2 miesiącach			Po 3 miesiącach				
	<i>R</i>	<i>Me</i>	<i>Q1-Q3</i>	<i>R</i>	<i>Me</i>	<i>Q1-Q3</i>	<i>R</i>	<i>Me</i>	<i>Q1-Q3</i>	<i>R</i>	<i>Me</i>	<i>Q1-Q3</i>	χ^2	<i>p</i>
Ocena zadowolenia z życia (1-10)	2,5	3	1-4	3,8	5	3-6	4,3	6	3-8	5,1	8	4-8	48,71	< 0,001

R – średnia rang

Testy *post-hoc* dla ANOVA Friedmana należy przedstawiać analogicznie, jak testy dla istotnych wyników zwykłego testu ANOVA lub testu Kruskala-Wallisa.

VIII. Dodatkowe testy statystyczne – ocena zgodności „sędziów”

Jeżeli badanie dotyczy oceny pewnej cechy (np. atrakcyjności koloru tęczówki) dokonanej przez „sędziów” częstą praktyką jest sprawdzanie zgodności ich ocen. Służy ku temu współczynnik W Kendalla. Przyjmuje on wartości liczb rzeczywistych z zakresu $\langle 0;1 \rangle$, gdzie 0 oznacza kompletny brak zgodności „sędziowskiej” (oceny są przyznawane zupełnie losowo), a 1 całkowitą zgodność sędziowską (np. osobom o niebieskich tęczówkach przyznawane są przez „sędziów” najwyższe noty). W przypadku, gdy atrakcyjność tęczówek była oceniana w 4 grupach (1- kobiety oceniały oczy mężczyzn; 2- kobiety oceniały oczy kobiet; 3- mężczyźni oceniali oczy kobiet; 4- mężczyźni oceniali oczy mężczyzn) współczynniki zgodności sędziowskiej Kendalla przedstawić można w następujący sposób:

Tabela 27. Współczynniki zgodności sędziowskiej oceny atrakcyjności tęczówek męskich i żeńskich dokonanej przez „sędziów” obydwu płci.

Kolor oczu Grupa	tęczówki brązowe	tęczówki niebieskie	tęczówki zielone
Grupa 1			
W	0,35	0,56	0,23
χ^2	53,25	85,21	33,61
p	<0,001	<0,001	0,021
Grupa 2			
W	0,19	0,43	0,11
χ^2	57,15	129,68	33,51
p	<0,001	<0,001	0,020
Grupa 3			
W	0,26	0,32	0,06
χ^2	94,61	115,17	22,66
p	<0,001	<0,001	0,256
Grupa 4			
W	0,55	0,57	0,50
χ^2	72,67	86,64	75,77
p	<0,001	<0,001	<0,001

* wartości statystyki *chi-kwadrat* oraz poziomu p można znaleźć w teście ANOVA Friedmana.