

Raportowanie wyników testów statystycznych dwóch zmiennych

Spis treści

1	Struktura raportowania	2
1.1	Rozdział Metody	2
1.2	Rozdział Wyniki	3
1.3	Ogólne wskazania	4
1.4	Uwagi statystyczne	5
2	Testy istotności różnic dla danych niezależnych	6
2.1	Dwie grupy	6
2.1.1	Test Welcha (Studenta) dla dwóch prób	6
2.1.2	Jednostronny test Welcha (Studenta) dla dwóch prób	7
2.1.3	Test Manna–Whitneya	8
2.2	Więcej niż dwie grupy	9
2.2.1	Test jednoczynnikowa ANOVA	9
2.2.2	Test jednoczynnikowa ANOVA Welcha	11
2.2.3	Test Kruskala–Wallisa	13
3	Testy istotności różnic dla danych zależnych	15
3.1	Dwie grupy	15
3.1.1	Test Studenta dla dwóch prób zależnych	15
3.1.2	Test Wilcoxona	17
4	Testy siły związku	18
4.1	Test istotności współczynnika korelacji Pearsona	18
4.2	Test istotności współczynnika korelacji Spearmana	20
4.3	Test niezależności zmiennych χ^2	21

1. Struktura raportowania

1.1. Rozdział Metody

- **pytanie badawcze** - pytanie, na które spróbujemy znaleźć odpowiedź
 - Bezpośrednio związane z hipotezą badawczą, ale często poprzedzone dodatkowo słowem „czy”.
- **hipoteza badawcza** - zakładana odpowiedź na pytanie badawcze
 - Od hipotez testowych H_0 i H_1 różni się tym, że:
 - jej sformułowanie jest bardziej ogólne, często dotyczy zmiennych jeszcze sprzed operacjonalizacji, a nie konkretnych zależności w danych,
 - jest zgodna tylko z jedną z nich - częściej z H_1 niż z H_0 .
- **nazwy zastosowanych testów statystycznego** do zbadania problemu
 - Jeśli zmienna jest ilościowa, raportujemy o wykonaniu wybranych testów wstępnych, by sprawdzić założenia odpowiedniego testu parametrycznego, a następnie, na podstawie ich wyników, podajemy ostatecznie wybrany test właściwy. W przypadku pozostałych zmiennych, od razu podajemy nazwę testu właściwego.
 - Jeśli uzyskany wynik z testu dla więcej niż 2 prób jest istotny statystycznie, to uprzedzamy fakty i podajemy przy nim od razu nazwę testów post-hoc. Jeśli nie jest, pomijamy o nich informację i podajemy jedynie nazwę testu dla kilku prób.
- **opis wyników testów wstępnych**, jeśli były one wykonane
 - Dla testów wstępnych oraz testów post-hoc zwykle nie zamieszczamy w raporcie badanych hipotez.
 - Wyniki testów wstępnych można zamieścić w tym rozdziale, by w rozdziale „Wyniki” zamieścić już kluczowe i gotowe do interpretacji analizy.
 - Jeśli z powodu nienormalności rozkładu zmiennej wykonaliśmy skuteczną transformację danych, to należy napisać jakie to było przekształcenie i przedstawić wyniki testu wstępnego jedynie dla już przekształconych danych.
- **ewentualne dodatkowe informacje na temat zastosowania testów**, jeśli przyjęto jakieś założenia
 - Jeśli może być coś jeszcze nieoczywistego dla czytelnika, jak wykonano podane testy statystyczne, to należy o tym napisać w raporcie.

1.2. Rozdział Wyniki

- **hipotezy testowe** - zerowa i alternatywna -
obie hipotezy, które przeciwstawia sobie wybrany test statystyczny
- **wykres wspólnego rozkładu zmiennych**
biorących udział w teście statystycznym, wizualizujący badany problem
 - Pod wykresem powinien znaleźć się podpis rozpoczęty numerem wykresu. Jego postać w zależności od typu wykresu może być na przykład taka:
 - wykres bez zmiennej grupującej (wykres rozrzutu, mapa ciepła):
„Rozkład x i y względem siebie”
 - wykres ze zmienną grupującą:
 - „Rozkład x względem y ”
 - „Rozkład x w zależności od y ”
 - „Rozkład x dla/wśród/w poszczególnych y ”
 - „Rozkład x poszczególnych y ”
 - Jeśli na wykresie pojawiają się skróty lub coś innego wymaga wyjaśnienia, to należy to dodać do podpisu wykresu.
 - W szczególności, jeśli wykres słupkowy zawiera słupki błędów, to należy napisać, co one określają, np. „Słupki błędów: odchylenie standardowe”.
 - Wykres niebędący częścią wykresu panelowego nie powinien zawierać tytułu.
 - Wykres powinien zawierać podpisy osi i jeśli jest taka potrzeba, odpowiednie dla zmiennych jednostki umieszczone w kwadratowych nawiasach.
 - Wykres powinien być czytelny, w szczególności nie zawierać zbędnych dla jego zrozumienia elementów.
- **tabele z wynikami testów statystycznych** (właściwego i post-hoc)
 - Tabele raportujące wynik testu statystycznego powinny zawierać:
 - rozkład badanych zmiennych
Dla testów parametrycznych zwykle warto zawrzeć n , m , sd ,
a dla nieparametrycznych n , med , IQR .
 - wartości związane z testem:
stopnie swobody, statystyki testowe, wartości p
 - wielkość efektu (warto nawet jeśli wynik jest statystycznie nieistotny),
 - Nad wykresem powinien znaleźć się podpis rozpoczęty numerem tabeli.
 - Jeśli na wykresie pojawiają się skróty lub coś innego wymaga wyjaśnienia, to należy to zawrzeć w adnotacji pod tabelą.

- W szczególności, pod tabelą należy zawrzeć wyjaśnienia oznaczeń stopnia statystycznej istotności wyników - tylko występujących w danej tabeli lub wszystkich używanych w raporcie niezależnie od ich występowania w danej tabeli.

■ **opis wyników testów statystycznych** (właściwego i post-hoc)

- Wnioski z testu statystycznego powinny być zawarte w języku naturalnym, a odpowiednie statystyki i wartości związane z testem powinny być umieszczone w nawiasach.
- Jeśli wyniki przedstawione w zdaniu odnoszą się do zawartości tabeli (lub wykresu), to powinno się zamieścić w nim odpowiednie odwołanie. Jeśli zdania po nim następujące mają to samo odniesienie, nie trzeba już powtarzać tego odwołania.
- Jeśli wyniki testu istotności różnic są zamieszczone również w tabeli, to we wniosku przy wyniku nieistotnym statystycznie, można pominąć wielkość efektu, a umieścić go jedynie w tabeli.

1.3. Ogólne wskazania

- Wszystkie liczby w raporcie powinny mieć tę samą liczbę cyfr po przecinku, zazwyczaj wystarczające są dwie.
 - Wyjątek w raporcie mogą stanowić wartości p , które są na granicy któregoś ze stopni istotności (np. $< 0,05$ lub $< 0,001$) i dla których przy ustalonej liczbie cyfr może nie być jasne, w którym przedziale się one znajdują — wtedy można podać odpowiednio większą liczbę cyfr.
- Znaki dziesiętne w raporcie powinny być dopasowane do języka raportu - w języku polskim znakiem dziesiętnym jest przecinek, a różne liczby możemy w raporcie oddzielać średnikiem.
- Jeśli wartość p wynosi mniej niż 0,001, to we wniosku, opcjonalnie również w tabeli, zamiast dokładnej wartości podajemy „ $< 0,001$ ”.
- Dodanie do testów statystycznych interpretacji wielkości efektu jest wartościowe, ale nie konieczne.
- Skróty z pierwszym pojawieniem się w raporcie powinny zostać rozwinięte, np. 1. „analiza wariancji (ANOVA)”, 2. „ANOVA”.
- Dozwolone jest umieszczenie w raporcie testu Welcha pod nazwą testu Studenta.
- Wyniki testów statystycznych dla wielu prób zależnych raportuje się analogicznie do raportowania wyników testów dla wielu prób niezależnych.

1.4. Uwagi statystyczne

- Przy testach dla dwóch zmiennych obsługa braków danych nie jest konieczna, ponieważ funkcje automatycznie odrzucają wiersze, w których nie ma pary wartości, a nie wiąże się to z utratą żadnej innej pary wartości. Ma ona znaczenie dopiero, gdy zestawiamy ze sobą co najmniej trzy zmienne, np. w wieloczynnikowej analizie wariancji lub w modelu uczenia maszynowego. Niezależnie od liczby zmiennych wymagana jest jednak zamiana konkretnych wartości na braki danych, jeśli za takie można je uznać (np. „Nie wiem” \rightarrow np.NaN).
- Testy wstępne wymagane są jedynie dla zmiennych ilościowych.
- Testy post-hoc wymagane są jedynie, gdy wynik w teście dla wielu prób jest istotny statystycznie lub opcjonalnie, gdy jest tego bliski.
- Testy nieparametryczne, tak samo jak parametryczne, mają swoje założenia. Są one prostsze do spełnienia, ale również przed wykonaniem tych testów należy się upewnić, czy w danym przypadku tak jest, co zwykle można zrobić bez dodatkowych testów. Sprawdzanie tego nie jest jednak wymagane na projekcie.
- Zanim wykorzysta się test nieparametryczny dla zmiennej ilościowej, warto spróbować przekształcić dane do rozkładu normalnego za pomocą transformacji do tego służących (np. Boxa-Coxa) i ponownie wykonać test sprawdzający normalność rozkładu. Jeśli okaże się ona skuteczna, można wykonać test parametryczny na przekształconej zmiennej, jednak dla zrozumiałej interpretacji jego wyników statystyki umieszczone we wniosku i tabeli powinny odnosić się do oryginalnej zmiennej. Wizualizacja może być natomiast dla przekształconej zmiennej. Użycie transformacji w przypadku niespełnionych założeń nie jest jednak wymagane na projekcie.
- Wielkością efektu dla testu Manna-Whitneya i Wilcoxona może być CL (Common Language) lub r_{rb} - współczynnik korelacji rangowo-dwuseryjnej. Dla testów Studenta dla 2 prób może to być natomiast d Cohena lub r_{pb} - współczynnik korelacji punktowo-dwuseryjnej (przy czym ten drugi do uzyskania wymaga użycia osobnej funkcji).

2. Testy istotności różnic dla danych niezależnych

2.1. Dwie grupy

2.1.1. Test Welcha (Studenta) dla dwóch prób

W rozdziale Metody:

Pytanie badawcze: Czy kobiety i mężczyźni różnią się wskaźnikiem uwagi?

Hipoteza badawcza: Kobiety i mężczyźni różnią się wskaźnikiem uwagi.

Metoda sprawdzająca założenie testu Welcha/Studenta - normalność rozkładu:
test Shapiro-Wilka

Wynik testów wstępnych:

Na poziomie istotności 0,05 nie mamy podstaw do odrzucenia hipotezy zerowej ($W = 1,00$; $p = 0,89$), a więc przyjmujemy, że wskaźnik uwagi wśród grup poszczególnych płci wystarczająco spełnia założenie normalności w teście Studenta.

Wynik testów wstępnych, gdyby wynik był istotny statystycznie:

Wskaźnik uwagi wśród grup poszczególnych płci nie ma rozkładu normalnego ($W = 0,94$; $p < 0,001$).

Metoda: test Welcha/Studenta dla dwóch prób niezależnych

W rozdziale Wyniki:

[wykres rozkładu łącznego zmiennych]

Hipotezy testowe:

H_0 : Średni wskaźnik uwagi u kobiet jest równy średniemu wskaźnikowi uwagi u mężczyzn.

H_1 : Średni wskaźnik uwagi u kobiet jest różny od średniego wskaźnika uwagi u mężczyzn.

Tabela 1: Porównanie wskaźnika uwagi między płciami

płeć	wskaźnik uwagi			test Welcha/Studenta			
	n	m	sd	df	T	p	d
kobieta	116	0,59	0,14	239,17	-1,03	0,30	0,12
mężczyzna	184	0,60	0,14				

Wniosek:

Na poziomie istotności 0,05 nie możemy stwierdzić, czy wskaźnik uwagi u kobiet ($m = 0,59$; $sd = 0,14$) jest różny od wskaźnika uwagi u mężczyzn ($m = 0,60$; $sd = 0,14$) ($T(239,17) = -1,03$; $p = 0,30$) (Tab. 1).

Wniosek, gdyby wynik był istotny statystycznie:

Wskaźnik uwagi u kobiet ($m = \dots$; $sd = \dots$) jest różny od wskaźnika uwagi u mężczyzn ($m = \dots$; $sd = \dots$) ($T(\dots) = \dots$; $p = \dots$; $d = \dots$) (Tab. 1).

2.1.2. Jednostronny test Welcha (Studenta) dla dwóch prób

Pytanie badawcze: Czy kobiety i mężczyźni różnią się wskaźnikiem uwagi?

Hipoteza badawcza: Kobiety mają większy wskaźnik uwagi niż mężczyźni.

Metoda: jednostronny test Welcha/Studenta dla dwóch prób niezależnych

Hipotezy testowe:

H_0 : Średni wskaźnik uwagi u kobiet jest równy średniemu wskaźnikowi uwagi u mężczyzn.

H_1 : Średni wskaźnik uwagi u kobiet jest większy od średniego wskaźnika uwagi u mężczyzn.

Wniosek:

Na poziomie istotności 0,05 nie możemy stwierdzić, czy wskaźnik uwagi u kobiet ($m = 0,59$; $sd = 0,14$) jest większy od wskaźnika uwagi u mężczyzn ($m = 0,60$; $sd = 0,14$) ($T(239,17) = -1,03$; $p = 0,85$).

[pozostałe punkty analogicznie do dwustronnego testu Welcha]

2.1.3. Test Manna–Whitneya

W rozdziale Metody:

Pytanie badawcze: Czy kobiety i mężczyźni różnią się poziomem wykształcenia?

Hipoteza badawcza: Kobiety i mężczyźni różnią się poziomem wykształcenia.

[test wstępny, jeśli zmienna jest ilościowa]

Metoda: test Manna-Whitneya

W rozdziale Wyniki:

[wykres rozkładu łącznego zmiennych]

Hipotezy testowe:

H_0 : Mediana poziomu wykształcenia u kobiet jest równa medianie poziomu wykształcenia u mężczyzn.

H_1 : Mediana poziomu wykształcenia u kobiet jest różna od mediany poziomu wykształcenia u mężczyzn.

Tabela 2: Porównanie poziomu wykształcenia między płciami

płeć	poziom wykształcenia			test Manna–Whitneya		
	n	med	IQR	U	p	CL
kobieta	116	2	2	190	0,02	0,31
mężczyzna	184	3	2		*	

* – $p \leq 0,05$

Wniosek:

Kobiety różnią się poziomem wykształcenia ($med = 2$; $IQR = 2$) od mężczyzn ($med = 3$; $IQR = 2$) ($U = 190$; $p = 0,02$; $CL = 0,31$) (Tab. 2).

Wniosek, gdyby wynik nie był istotny statystycznie:

Na poziomie istotności 0,05 nie możemy stwierdzić, czy kobiety różnią się poziomem wykształcenia ($med = \dots$; $IQR = \dots$) od mężczyzn ($med = \dots$; $IQR = \dots$) ($U = \dots$; $p = \dots$) (Tab. 2).

2.2. Więcej niż dwie grupy

2.2.1. Test jednoczynnikowa ANOVA

W rozdziale Metody:

Pytanie badawcze:

Czy tematy treści różnią się zmiennością nastroju jaka następuje przy interakcji z nimi?

Hipoteza badawcza:

Tematy treści różnią się zmiennością nastroju jaka następuje przy interakcji z nimi.

Metoda sprawdzająca założenie testu ANOVA - normalność rozkładu:

test Shapiro-Wilka

Metoda sprawdzająca założenie testu ANOVA - homogeniczność wariancji:

test Levene'a

Wynik testów wstępnych:

Na poziomie istotności 0,05 nie mamy podstaw do odrzucenia hipotezy zerowej ($W = 1,00$; $p = 0,55$), a więc przyjmujemy, że wskaźnik zmienności nastroju w trakcie interakcji z treściami wśród poszczególnych ich tematów wystarczająco spełnia założenie normalności w teście ANOVA.

Na poziomie istotności 0,05 nie mamy podstaw do odrzucenia hipotezy zerowej ($W = 0,31$; $p = 0,82$), a więc przyjmujemy, że wskaźnik zmienności nastroju w trakcie interakcji z treściami wśród poszczególnych ich tematów wystarczająco spełnia założenie homogeniczności wariancji w teście ANOVA.

Metoda: jednoczynnikowa analiza wariancji (ANOVA) z testami post-hoc HSD Tukeya
[o testach post-hoc wspomnieć, o ile ANOVA jest istotna]

W rozdziale Wyniki:

[wykres rozkładu łącznego zmiennych]

Hipotezy testowe:

H_0 : Średni wskaźnik zmienności nastroju w trakcie interakcji z treściami jest taka sama dla wszystkich tematów tych treści.

H_1 : Średni wskaźnik zmienności nastroju w trakcie interakcji z treściami różni się w co najmniej jednej parze tematów tych treści.

Tabela 3: Jednoczynnikowa ANOVA dla wskaźnika zmienności nastroju w zależności od tematu treści

źródło zmienności	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η^2
między grupami	3	0,24	0,08	3,59	0,014	0,04
wewnątrz grup	296	6,73	0,02		*	

* – $p \leq 0,05$

Wniosek z testu ANOVA:

Tematy treści różnią się zmiennością nastroju jaka następuje przy interakcji z nimi ($F(3; 296) = 3,59$; $p = 0,014$; $\eta^2 = 0,04$) (Tab. 3).

Wniosek, gdyby wynik nie był istotny statystycznie:

Na poziomie istotności 0,05 nie można stwierdzić, czy tematy treści różnią się zmiennością nastroju jaka następuje przy interakcji z nimi ($F(\dots) = \dots$; $p = \dots$) (Tab. 3).

Tabela 4: Porównania parami dla wskaźnika zmienności nastroju w zależności od tematu treści

temat	zmienność nastroju			test HSD Tukeya			
	<i>n</i>	<i>m</i>	<i>sd</i>	<i>df</i>	<i>T</i>	<i>p</i>	<i>g</i>
radzenie sobie	84	0,54	0,16	296	2,70	0,04	0,41
uwaga	83	0,47	0,15			*	
radzenie sobie	84	0,54	0,16	296	-0,02	1,00	-0,00
budowa odporności	64	0,54	0,14				
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

* – $p \leq 0,05$; ** – $p \leq 0,01$; *** – $p \leq 0,001$

Wnioski z testów post-hoc:

Temat umiejętności radzenia sobie różni się zmiennością nastroju jaka następuje przy interakcji z nimi ($m = 0,54$; $sd = 0,16$) od tematu uwagi ($m = 0,47$; $sd = 0,15$) ($T(296) = 2,70$; $p = 0,04$; $g = 0,41$) (Tab. 4).

Temat uwagi różni się zmiennością nastroju jaka następuje przy interakcji z nimi od tematu zarządzania stresem ($m = 0,54$; $sd = 0,15$) ($T(296) = -2,63$; $p = 0,04$; $g = -0,43$).

Dla pozostałych par tematów, na poziomie istotności 0,05, nie można stwierdzić różnicy w zmienności nastroju jaka następuje przy interakcji z nimi.

2.2.2. Test jednoczynnikowa ANOVA Welcha

W rozdziale Metody:

Pytanie badawcze:

Czy tematy treści różnią się zmiennością nastroju jaka następuje przy interakcji z nimi?

Hipoteza badawcza:

Tematy treści różnią się zmiennością nastroju jaka następuje przy interakcji z nimi.

Metoda sprawdzająca założenie testu ANOVA Welcha - normalność rozkładu:

test Shapiro-Wilka

Wynik testów wstępnych:

Na poziomie istotności 0,05 nie mamy podstaw do odrzucenia hipotezy zerowej ($W = 1,00$; $p = 0,55$), a więc przyjmujemy, że wskaźnik zmienności nastroju w trakcie interakcji z treściami wśród poszczególnych ich tematów wystarczająco spełnia założenie normalności w teście ANOVA Welcha.

Metoda: jednoczynnikowa analiza wariancji (ANOVA) Welcha z testami post-hoc Gamesa-Howella

[o testach post-hoc wspomnieć, o ile ANOVA jest istotna]

W rozdziale Wyniki:

[wykres rozkładu łącznego zmiennych]

Hipotezy testowe:

H_0 : Średni wskaźnik zmienności nastroju w trakcie interakcji z treściami jest taka sama dla wszystkich tematów tych treści.

H_1 : Średni wskaźnik zmienności nastroju w trakcie interakcji z treściami różni się w co najmniej jednej parze tematów tych treści.

Tabela 5: Jednoczynnikowa ANOVA Welcha dla wskaźnika zmienności nastroju w zależności od tematu treści

df_1	df_2	F	p	η^2
3	161,72	3,73	0,012 *	0,04

* – $p \leq 0,05$

Wniosek z testu ANOVA:

Tematy treści różnią się zmiennością nastroju jaka następuje przy interakcji z nimi ($F(3; 161,72) = 3,73$; $p = 0,012$; $\eta^2 = 0,04$) (Tab. 5).

Wniosek, gdyby wynik nie był istotny statystycznie:

Na poziomie istotności 0,05 nie można stwierdzić, czy tematy treści różnią się zmiennością nastroju jaka następuje przy interakcji z nimi ($F(\dots) = \dots$; $p = \dots$) (Tab. 5).

Tabela 6: Porównania parami dla wskaźnika zmienności nastroju w zależności od tematu treści

temat	zmienność nastroju			test Gamesa-Howella			
	n	m	sd	df	T	p	g
radzenie sobie	81	0,54	0,16	159,51	2,66	0,04	0,41
uwagażność	83	0,47	0,15			*	
radzenie sobie	81	0,54	0,16	140,80	-0,02	1,00	-0,00
budowa odporności	83	0,54	0,14				
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

* – $p \leq 0,05$; ** – $p \leq 0,01$; *** – $p \leq 0,001$

Wnioski z testów post-hoc:

Temat umiejętności radzenia sobie różni się zmiennością nastroju jaka następuje przy interakcji z nimi ($m = 0,54$; $sd = 0,16$) od tematu uwagażności ($m = 0,47$; $sd = 0,15$) ($T(159,51) = 2,66$; $p = 0,04$; $g = 0,41$) (Tab. 4).

Temat uwagażności różni się zmiennością nastroju jaka następuje przy interakcji z nimi od tematu budowania odporności ($m = 0,54$; $sd = 0,15$) ($T(136,40) = -2,67$; $p = 0,04$; $g = -0,44$).

Temat uwagażności różni się zmiennością nastroju jaka następuje przy interakcji z nimi od tematu zarządzania stresem ($m = 0,54$; $sd = 0,15$) ($T(147,90) = -2,67$; $p = 0,04$; $g = -0,43$).

Dla pozostałych par tematów, na poziomie istotności 0,05, nie można stwierdzić różnicy w zmienności nastroju jaka następuje przy interakcji z nimi.

2.2.3. Test Kruskala–Wallisa

W rozdziale Metody:

Pytanie badawcze: Czy typy treści różnią się poziomem trudności?

Hipoteza badawcza: Typy treści różnią się poziomem trudności.

[test wstępny, jeśli zmienna jest ilościowa]

Metoda: test Kruskala-Wallisa z testami post-hoc Dunn

[o testach post-hoc wspomnieć, o ile test KW jest istotny]

W rozdziale Wyniki:

[wykres rozkładu łącznego zmiennych]

Hipotezy testowe:

H_0 : Mediana poziomu trudności jest taka sama dla wszystkich typów treści.

H_1 : Mediana poziomu trudności różni się w co najmniej jednej parze typów treści.

Tabela 7: Test Kruskala-Wallisa dla poziomu trudności w zależności od typu treści

df	H	p	η^2
3	25,71	0,00 ***	0,09

*** – $p \leq 0,001$

Wniosek z testu Kruskala-Wallisa:

Typy treści różnią się poziomem trudności ($H(3) = 25,71$; $p < 0,001$; $\eta^2 = 0,09$) (Tab. 7).

Wniosek, gdyby wynik nie był istotny statystycznie:

Na poziomie istotności 0,05 nie można stwierdzić, czy typy treści różnią się poziomem trudności ($H(\dots) = \dots$; $p = \dots$) (Tab. 7).

Tabela 8: Porównania parami dla poziomu trudności w zależności od typu treści

treść	poziom trudności			test Dunn
	n	med	IQR	p
artykuł	95	1	1	0,11
interaktywna	110	2	2	
artykuł	95	1	1	0,00 ***
quiz	12	3	0	
⋮	⋮	⋮	⋮	⋮

* – $p \leq 0,05$; ** – $p \leq 0,01$; *** – $p \leq 0,001$

Wnioski z testów post-hoc:

Treść typu quiz różni się poziomem trudności ($med = 3$; $IQR = 0$) od treści typu artykuł ($med = 1$; $IQR = 1$) ($p < 0,001$) (Tab. 8).

Treść typu quiz różni się poziomem trudności od treści typu interaktywnego ($med = 2$; $IQR = 2$) ($p < 0,001$).

Treść typu quiz różni się poziomem trudności od treści typu wideo ($med = 1$; $IQR = 2$) ($p < 0,001$).

Dla pozostałych par typów treści, na poziomie istotności 0,05, nie można stwierdzić różnicy w poziomie trudności.

3. Testy istotności różnic dla danych zależnych

3.1. Dwie grupy

3.1.1. Test Studenta dla dwóch prób zależnych

W rozdziale Metody:

Pytanie badawcze:

Czy pomiary stresu przed i po interakcji użytkownika z treścią różnią się od siebie?

Hipoteza badawcza:

Pomiary stresu przed i po interakcji użytkownika z treścią różnią się od siebie.

Metoda sprawdzająca założenie testu Studenta - normalność rozkładu:

test Shapiro-Wilka

Wynik testów wstępnych:

Na poziomie istotności 0,05 nie mamy podstaw do odrzucenia hipotezy zerowej ($W = 1,00$; $p = 0,89$), a więc przyjmujemy, że różnica między pomiarami stresu przed interakcją użytkownika z treścią i po tej interakcji wystarczająco spełnia założenie normalności w teście Studenta.

Wynik testów wstępnych, gdyby wynik był istotny statystycznie:

Różnica między pomiarami stresu przed interakcją użytkownika z treścią i po tej interakcji nie ma rozkładu normalnego ($W = 0,94$; $p < 0,001$).

Metoda: test Studenta dla dwóch prób zależnych

W rozdziale Wyniki:

[wykres rozkładu łącznego zmiennych]

Hipotezy testowe:

H_0 : Średnia różnica między pomiarami stresu przed i po interakcji użytkownika z treścią wynosi 0.

H_1 : Średnia różnica między pomiarami stresu przed i po interakcji użytkownika z treścią jest różna od 0.

Tabela 9: Porównanie pomiarów stresu przed i po interakcji

pomiar	stres			test Studenta			
	n	m	sd	df	T	p	d
przed interakcją	300	25,13	5,95	299	45,62	0,00	0,52
po interakcji	300	22,02	5,93			***	

*** – $p \leq 0,001$

Wniosek:

Pomiary stresu przed interakcją użytkownika z treścią ($m = 25,13$; $sd = 5,95$) i po tej interakcji ($m = 22,02$; $sd = 5,93$) różnią się od siebie ($T(299) = 45,62$; $p < 0,001$; $d = 0,52$) (Tab. 9).

Wniosek, gdyby wynik nie był istotny statystycznie:

Na poziomie istotności 0,05 nie możemy stwierdzić, czy pomiary stresu przed interakcją użytkownika z treścią ($m = \dots$; $sd = \dots$) i po tej interakcji ($m = \dots$; $sd = \dots$) różnią się od siebie ($T(\dots) = \dots$; $p = \dots$) (Tab. 9).

3.1.2. Test Wilcoxona

W rozdziale Metody:**Pytanie badawcze:**

Czy pomiary stresu przed i po interakcji użytkownika z treścią różnią się od siebie?

Hipoteza badawcza:

Pomiary stresu przed i po interakcji użytkownika z treścią różnią się od siebie.

[test wstępny, jeśli zmienna jest ilościowa]

Metoda: test Wilcoxona

W rozdziale Wyniki:

[wykres rozkładu łącznego zmiennych]

Hipotezy testowe:

H_0 : Mediana różnicy między pomiarami stresu przed i po interakcji użytkownika z treścią wynosi 0.

H_1 : Mediana różnicy między pomiarami stresu przed i po interakcji użytkownika z treścią jest różna od 0.

Tabela 10: Porównanie pomiarów stresu przed i po interakcji

pomiar	stres			test Wilcoxon		
	<i>n</i>	<i>med</i>	<i>IQR</i>	<i>W</i>	<i>p</i>	<i>CL</i>
przed interakcją	300	25,00	11,00	0,00	0,00	0,64
po interakcji	300	22,18	10,08		***	

*** – $p \leq 0,001$

Wniosek:

Pomiary stresu przed interakcją użytkownika z treścią ($med = 25,00$; $IQR = 11,00$) i po tej interakcji ($med = 22,18$; $IQR = 10,08$) różnią się od siebie ($W = 0,00$; $p < 0,001$; $CL = 0,64$) (Tab. 10).

Wniosek, gdyby wynik nie był istotny statystycznie:

Na poziomie istotności 0,05 nie możemy stwierdzić, czy pomiary stresu przed interakcją użytkownika z treścią ($med = \dots$; $IQR = \dots$) i po tej interakcji ($med = \dots$; $IQR = \dots$) różnią się od siebie ($W = \dots$; $p = \dots$) (Tab. 10).

4. Testy siły związku

4.1. Test istotności współczynnika korelacji Pearsona

W rozdziale Metody:

Pytanie badawcze:

Czy między wskaźnikami uwagi i zmienności nastroju istnieje związek?

Hipoteza badawcza:

Między wskaźnikami uwagi i zmienności nastroju istnieje związek.

Metoda sprawdzająca założenie współczynnika korelacji Pearsona

- normalność rozkładu: test Henze-Zirklera

Wynik testów wstępnych:

Na poziomie istotności 0,05 nie mamy podstaw do odrzucenia hipotezy zerowej ($HZ = 0,58$; $p = 0,63$), a więc przyjmujemy, że wskaźnik uwagi i wskaźnik zmienności nastroju w trakcie interakcji z treściami wystarczająco spełniają założenie normalności w teście istotności współczynnika korelacji Pearsona.

Metoda: współczynnik korelacji Pearsona

W rozdziale Wyniki:

[wykres rozkładu łącznego zmiennych]

Hipotezy testowe:

H_0 : Korelacja pomiędzy wskaźnikami uwagi i zmienności nastroju wynosi 0.

H_1 : Korelacja pomiędzy wskaźnikami uwagi i zmienności nastroju jest różna od 0.

Tabela 11: Statystyki opisowe i współczynnik korelacji dla wskaźników uwagi i zmienności nastroju

zmienna	rozkład zmiennej					test korelacji Pearsona		
	<i>n</i>	<i>m</i>	<i>sd</i>	<i>min</i>	<i>max</i>	<i>df</i>	<i>r</i>	<i>p</i>
wskaźnik uwagi	300	0,60	0,14	0,12	1,00	298	0,11	0,052
wskaźnik zmienności nastroju	300	0,52	0,15	0,08	0,92			

Wniosek:

Na poziomie istotności 0,05 nie można stwierdzić, czy między wskaźnikami uwagi i zmienności nastroju istnieje związek ($r(298) = 0,11$; $p = 0,052$) (Tab. 11).

Wniosek, gdyby wynik był istotny statystycznie:

Między wskaźnikami uwagi i zmienności nastroju istnieje związek o zgodnym/przeciwym kierunku ($r(\dots) = \dots$; $p = \dots$) - im wyższy wskaźnik uwagi, tym wyższy wskaźnik zmienności nastroju (Tab. 11).

4.2. Test istotności współczynnika korelacji Spearmana

W rozdziale Metody:

Pytanie badawcze:

Czy między trudnością treści i wskaźnikiem informacji zwrotnej istnieje związek?

Hipoteza badawcza:

Między trudnością treści i wskaźnikiem informacji zwrotnej istnieje związek.

[test wstępny, jeśli zmienna jest ilościowa]

Metoda: współczynnik korelacji Spearmana

W rozdziale Wyniki:

[wykres rozkładu łącznego zmiennych]

Hipotezy testowe:

H_0 : Korelacja pomiędzy trudnością treści i wskaźnikiem informacji zwrotnej wynosi 0.

H_1 : Korelacja pomiędzy trudnością treści i wskaźnikiem informacji zwrotnej jest różna od 0.

Tabela 12: Statystyki opisowe i współczynnik korelacji dla trudności treści i wskaźnika informacji zwrotnej

zmienna	rozkład zmiennej					test korelacji Spearmana	
	n	med	IQR	min	max	r_S	p
trudność treści	300	2	2	1	3	0,14	0,014
wskaźnik informacji zwrotnej	300	3	2	1	5		*

* – $p \leq 0,05$

Wniosek:

Między trudnością treści i wskaźnikiem informacji zwrotnej istnieje związek o zgodnym kierunku ($r_S = 0,14$; $p = 0,014$) - im większa trudność treści, tym wyższy wskaźnik informacji zwrotnej (Tab. 12).

Wniosek, gdyby wynik nie był istotny statystycznie:

Na poziomie istotności 0,05 nie można stwierdzić, czy między trudnością treści i wskaźnikiem informacji zwrotnej istnieje związek ($r_S(\dots) = \dots$; $p = \dots$) (Tab. 12).

4.3. Test niezależności zmiennych χ^2

W rozdziale Metody:

Pytanie badawcze: Czy między płcią i zawodem istnieje związek?

Hipoteza badawcza: Między płcią i zawodem istnieje związek.

Metoda: test niezależności χ^2 ze współczynnikiem korelacji V Craméra

W rozdziale Wyniki:

[wykres rozkładu łącznego zmiennych]

Hipotezy testowe:

H_0 : Płeć i zawód są od siebie niezależne.

H_1 : Płeć i zawód są od siebie zależne.

Tabela 13: Tabela krzyżowa i test niezależności χ^2 dla płci i zawodu

płeć	zawód (n)					test niezależności χ^2			
	inżynier	lekarz	uczeń	nauczyciel	bezrobotny	df	χ^2	p	V
kobieta	2	4	6	6	3	4	3,43	0,49	0,26
mężczyzna	8	3	5	8	5				

[jeśli tabela krzyżowa ma wiele pól, można przedstawić ją osobno]

Wniosek:

Na poziomie istotności 0,05 nie można stwierdzić, czy między płcią i zawodem, istnieje związek ($\chi^2(4) = 3,43$; $p = 0,49$; $V = 0,26$) (Tab. 13).

Wniosek, gdyby wynik był istotny statystycznie:

Między płcią i zawodem istnieje związek ($\chi^2(...) = \dots$; $p = \dots$; $V = \dots$) (Tab. 13). Największa zależność występuje w zawodzie inżyniera - mężczyźni częściej zostają inżynierami niż kobiety/pozostali przedstawiciele płci [gdyby było więcej poziomów].

Wniosek, gdyby wynik był istotny statystycznie i tabela krzyżowa była wymiaru 2×2 :

Między płcią i byciem inżynierem istnieje związek ($\chi^2(...) = \dots$; $p = \dots$; $V = \dots$) - mężczyźni częściej zostają inżynierami niż kobiety (Tab. 13).
