



Pogotowie[®]
Statystyczne

Jak wykonać w SPSS analizę korelacji Pearsona i Spearmana

TUTORIAL TEKSTOWY

TUTORIALE TEKSTOWE

T

Ten materiał jest jednym z serii tutoriali tekstowych, czyli poradników, w których w praktyczny sposób pokazujemy różne tajniki statystyki.

W TYM TUTORIALU

przedstawimy w jaki sposób wykonać analizę korelacji Pearsona i Spearmana w programie IBM SPSS Statistics oraz w jaki sposób zaraportować ich wyniki w standardzie APA 7.

Więcej tutoriali znajdziesz na www.pogotowiestatystyczne.pl w zakładce *Edukacja*.



Wprowadzenie i założenia

Analiza korelacji to metoda statystyczna służąca do oceny siły i kierunku związku między dwiema zmiennymi. Istnieją różne metody analiz korelacji, zarówno liniowych jak i nieliniowych.

Warto pamiętać, że analiza korelacji mierzy współwystępowanie wyników, a nie związek przyczynowo - skutkowy.

Istnieją dwie najczęściej wykorzystywane metody analizy korelacji:

- **Korelacja Pearsona** (oznaczana jako r) – analiza parametryczna¹ mierząca liniowy związek między zmiennymi. Jej stosowanie wymaga spełnienia kilku założeń, które zostaną opisane w dalszej części tego tekstu.
- **Korelacja Spearmana** (oznaczana jako ρ) – analiza nieparametryczna mierząca monotoniczny² związek między zmiennymi, w sytuacji gdy zmienne mają charakter porządkowy lub ilościowy, ale założenia dla korelacji Pearsona nie są spełnione.

Ten poradnik poświęcony zostanie właśnie korelacjom Pearsona oraz Spearmana. Główna różnica między nimi polega na tym, że korelacja Spearmana opiera się na porządku danych (rangach)³, a nie na ich rzeczywistych wartościach zmiennej, co ma miejsce w przypadku korelacji Pearsona.

DEFINICJA

Analiza korelacji to metoda statystyczna służąca do oceny siły i kierunku związku między dwiema zmiennymi. Dwie najbardziej popularne analizy korelacji to Pearsona i Spearmana, choć w rzeczywistości jest ich znacznie więcej (np. korelacja tau-b Kendalla).

¹ Test parametryczny to taki test, który opiera się na szacowaniu (estymacji) parametru – rzeczywistej wartości występującej w populacji, na podstawie danych zebranych w próbie pobranej z tej populacji.

² Związek monotoniczny występuje, gdy dwie zmienne mają tendencję do zmiany w tym samym kierunku (wznoszącym lub opadającym) ale niekoniecznie na stałym (liniowym) poziomie zmiany.

³ Pierwszym krokiem w analizie Spearmana jest rangowanie – zamiana wartości zmiennej poprzez ich uszeregowanie. Takie dwie przekształcone zmienne są następnie korelowane ze sobą w sposób analogiczny jak w analizie Pearsona.

Istnieją również inne metody analizy korelacji, przykładowo: korelacja tau-b Kendalla (alternatywa dla korelacji Spearmana), korelacja punktowa Pearsona (dla danych kodowanych jako 0 - brak występowania i 1 - występowanie); korelacja polichoryczna (dobry wybór w przypadku skali Likerta) czy korelacja poliseryjna (zoptymalizowana do korelacji zmiennych ilościowych z porządkowymi, dająca lepsze efekty niż Spearman). Jednak często to właśnie analizy Pearsona i Spearmana są używane powszechnie.

Kiedy stosujemy analizę korelacji Pearsona?

- kiedy obie korelowane zmienne są wyrażone na skali ilościowej,
- kiedy zależność między zmiennymi ma charakter liniowy.

Ponadto, wykonanie analizy korelacji Pearsona wymaga spełnienia założeń o:

- Rozkładzie normalnym⁴ ([tutaj](#) znajduje się materiał dotyczący zagadnienia testowania normalności rozkładu).
- Homoskedastyczności (równym rozkładzie wariancji reszt dla różnych wartości zmiennej niezależnej).
- Braku wartości odstających (*outlierów*).

Kiedy założenia te nie są spełnione, możliwym jest:

- Wykonanie testu, ryzykując że wynik będzie obciążony.
- Likwidacja przyczyny, np. usunięcie *outlierów*, transformacja rozkładu zmiennej (np. pierwiastkowanie, logarytmizacja)
- Wykonanie innej korelacji, np. nieliniowej albo nieparametrycznej (np. Spearmana).

PAMIĘTAJ

*Analizę korelacji Pearsona wykonujemy kiedy obie zmienne są wyrażone na skali ilościowej, zależność między nimi ma charakter liniowy oraz spełnione są określone założenia (np. brak *outlierów*). Gdy te nie są spełnione lub przynajmniej jedna ze zmiennych ma charakter porządkowy musimy zastosować inną analizę korelacji, np. Spearmana.*

⁴ Korelacja Pearsona jest względnie odporna na złamanie założenia o rozkładzie normalnym, dlatego przy niewielkim odchyleniu można obliczyć ten współczynnik. W sytuacji wystąpienia heteroskedastyczności, *outlierów* lub związku nieliniowego, odradzamy liczenie korelacji Pearsona.

Analiza korelacji opiera się na testowaniu poniższej hipotezy zerowej:

- H_0 : Wartość współczynnika korelacji jest równa 0 (brak korelacji).

Odpowiada temu następująca hipoteza alternatywna:

- H_1 : Wartość współczynnika korelacji istotnie statystycznie różni się od 0.

Oznacza to, że jeśli wartość $p < 0,05^5$, wynik taki traktujemy jako istotny statystycznie, co pozwala wnioskować o występowaniu związku pomiędzy zmiennymi. Wynik nieistotny statystycznie, tj. $p > 0,05$ pozwala wnioskować o braku związku⁶.

Wykonanie analizy korelacji Pearsona i Spearmana w SPSS

W tym tekście wykonywać będziemy analizę dla następującego przykładowego pytania badawczego:

Czy istnieje związek pomiędzy poziomem odczuwanego stresu, ekstrawersją i samooceną? ⁷

W celu udzielenia odpowiedzi na to pytanie badawcze zostanie wykonana analiza korelacji Pearsona. Jednocześnie, poniższy tekst można traktować jako instruktarz dotyczący również analizy Spearmana (sprawa jest o tyle prosta, że obie analizy wykonuje się i interpretuje w podobny sposób).

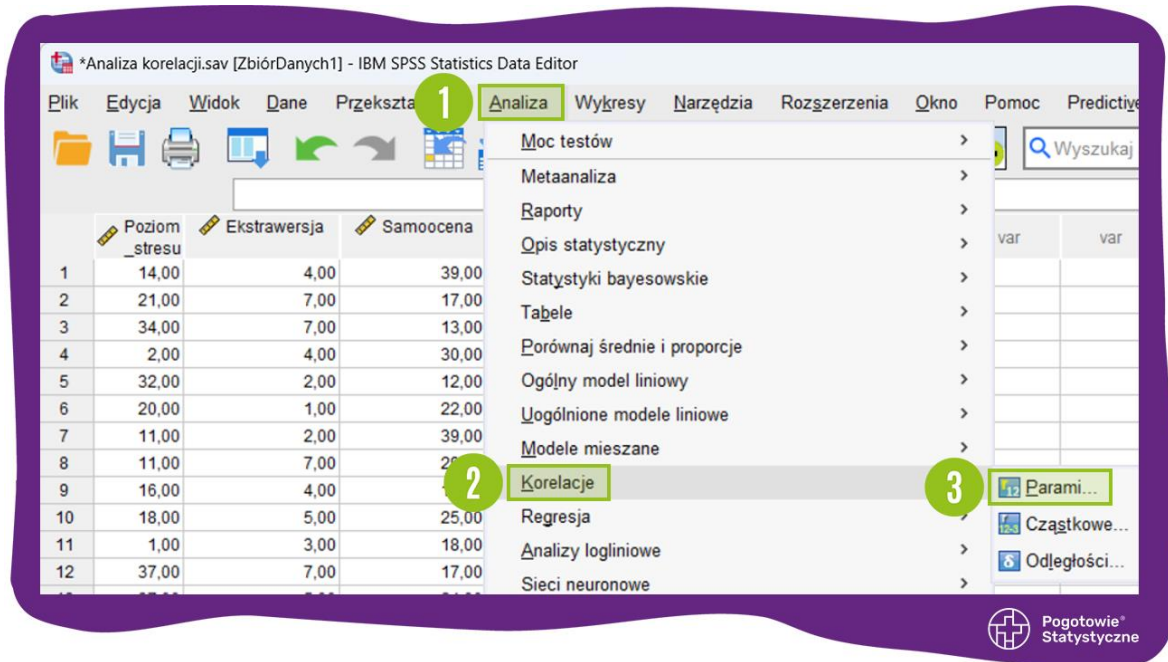
Aby wykonać analizę korelacji Pearsona lub Spearmana w SPSS, należy podjąć następujące kroki:

- (1) W oknie głównym programu klikamy w menu „Analiza”.
- (2) Najeżdżamy kursorem na opcję „Korelacje”, w efekcie czego rozwija nam się lista z wyborem analiz.
- (3) Z listy tej wybieramy opcję „Parami”.

⁵ Przy założeniu $\alpha = 0,05$. Taki próg przyjmuje się zazwyczaj, jednak może to być inna wartość np. 0,01 lub 0,1. W tym poradniku jako domyślną traktujemy $\alpha = 0,05$.

⁶ W praktyce jest to bardziej złożone. Taki wniosek możemy wyciągnąć jeśli moc testu jest wystarczająco wysoka ($1 - \beta \geq 0,80$). Jeśli moc testu jest niższa, wtedy wnioskujemy, że nie posiadamy wystarczających dowodów aby stwierdzić związek lub jego brak.

⁷ Dla uproszczenia sprawy, na potrzeby tego poradnika, formułujemy tylko ogólne pytanie badawcze, bez stawiania hipotez, w tym hipotez szczegółowych. Zakładamy przy tym, że domyślnie postawione zostaną tu (trzy) hipotezy niekierunkowe, a co za tym idzie dla wyników analizy korelacji interpretowana będzie dwustronna wartość p .



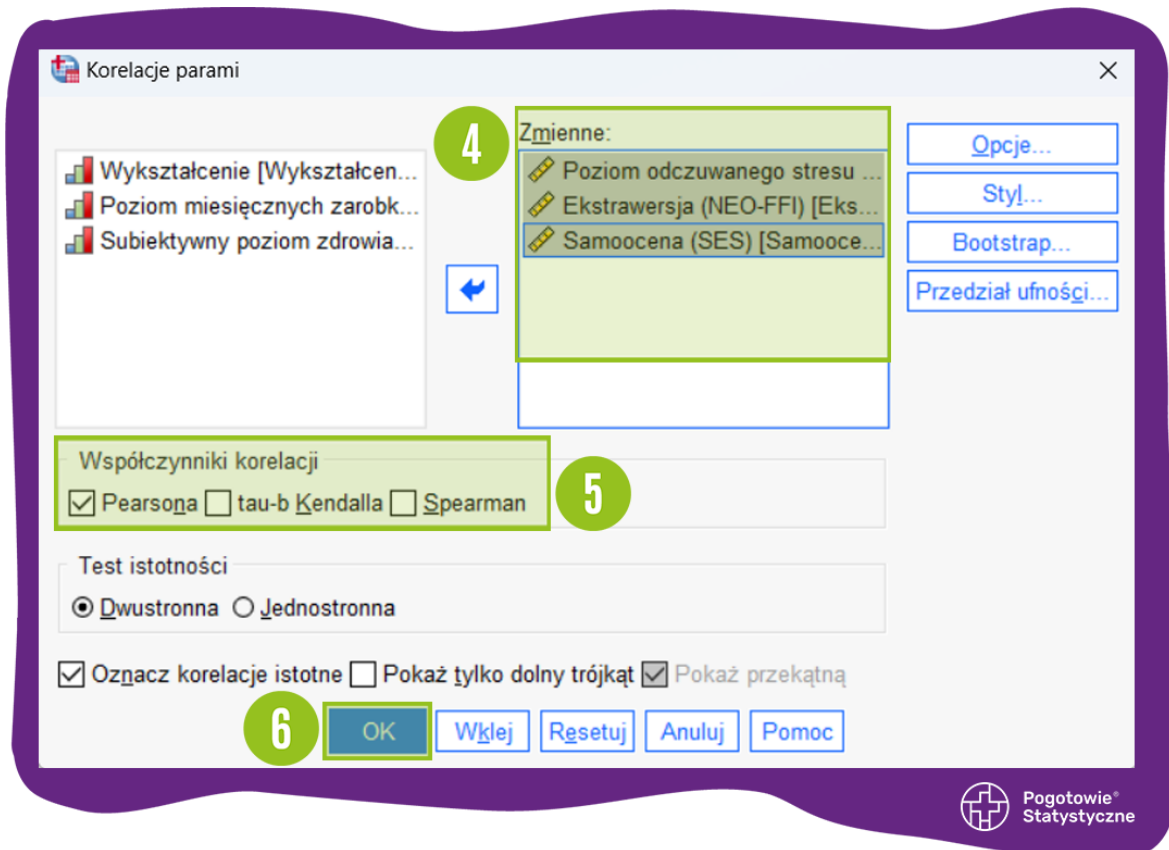
Rysunek 1. Zrzut ekranu z głównego okna programu SPSS przedstawiający sposób wyboru okna analizy korelacji z menu.

Kolejne kroki to:

- (4) W górnym oknie „Zmienne” umieszczamy zmienne, które chcemy ze sobą skorelować.
- (5) Poniżej, w sekcji “Współczynniki korelacji”, zaznaczamy opcję „Pearsona”⁸ lub „Spearmana” w zależności od tego którą analizę korelacji chcemy wykonać. W przypadku testowania hipotezy dwustronnej resztę opcji zostawiamy w ustawieniu domyślnym⁹.
- (6) Na końcu, zatwierdzamy wszystkie ustawienia analizy klikając przycisk „OK”.

⁸ W przypadku obu zmiennych ilościowych program liczy tutaj klasyczną korelację Pearsona. Kiedy natomiast jedna ze zmiennych jest dychotomiczna (0-1) a druga ilościowa, wtedy program automatycznie policzy punktową korelację Pearsona.

⁹ Jeżeli chcemy wybrać test jednostronny (w przypadku testowania hipotezy kierunkowej), to w opcji „Test istotności” zaznaczamy opcję „Jednostronna”. Na potrzeby tego tutorialu wykonujemy test dwustronny dla założonej hipotezy niekierunkowej.



Rysunek 2. Zrzut ekranu z okna menu analizy korelacji w programie SPSS przedstawiający sposób ustawiania opcji analizy korelacji Pearsona.

Uwaga! Choć dalsze instrukcje będą dotyczyć analizy korelacji Pearsona, to konstrukcja wydruku i interpretacja wyniku dla analizy korelacji Spearmana jest analogiczna.

Test został wykonany. Wyniki otrzymujemy w postaci wydruku w skład którego wchodzi jedna tabela.

Wartości na przecięciu zmiennych w pionie i poziomie tworzą wynik dla jednej analizy korelacji (korelacja pary zmiennych). Do wyników każdej analizy kolejno odnoszą się trzy wartości:

- (1) Wiersz „Korelacja Pearsona”: Współczynnik korelacji r Pearsona dla każdej skorelowanej pary zmiennych.
- (2) Wiersz „Istotność (dwustronna)”: Wynik testu istotności (wartość p) dla każdego współczynnika korelacji.
- (3) Wiersz „N”: Informacja o liczbie obserwacji uwzględnionych w analizie dla danej korelacji.

Korelacje

		Poziom odczuwanego stresu (PSS-10)	Ekstrawersja (NEO-FFI)	Samoocena (SES)
Poziom odczuwanego stresu (PSS-10)	Korelacja Pearsona	1	-0,279**	-0,339**
	Istotność (dwustronna)		0,005	<0,001
	N	100	100	100
Ekstrawersja (NEO-FFI)	Korelacja Pearsona	-0,279**	1	0,469**
	Istotność (dwustronna)	0,005		<0,001
	N	100	100	100
Samoocena (SES)	Korelacja Pearsona	-0,339**	0,469**	1
	Istotność (dwustronna)	<0,001	<0,001	
	N	100	100	100

** . Korelacja istotna na poziomie 0.01 (dwustronnie).

Pogotowie Statystyczne

Rysunek 3. Zrzut ekranu z wydruku wyników w programie SPSS przedstawiający zawartość tabeli z wynikami analizy korelacji Pearsona.

Wyniki analizy przedstawione są w postaci macierzy korelacji, to znaczy w układzie korelacji „wszystkiego ze wszystkim”. W naszym przykładzie mamy do czynienia z macrycą o wielkości 3x3, co oznacza że otrzymujemy łącznie 9 wyników¹⁰. W praktyce jednak, nie trzeba interpretować wszystkich tych wartości:

- Trzy wyniki po przekątnej tabeli to korelacja każdej zmiennej z sobą samą (np. samoocena x samoocena). Dla nich z definicji korelacja jest równa 1, wartość p w tym przypadku nie jest liczona.
- Pozostałe sześć wyników po obu stronach przekątnej stanowi swoje lustrzane odbicie. Oznacza to, że te sześć wyników to tak naprawdę trzy wyniki, ale podane dwukrotnie, raz jako korelacja zmiennej A i B oraz raz jako zmiennej B i A itd.

Interpretacja wyniku

Teraz czas na interpretację wyników analizy korelacji. Zaczynamy od sprawdzenia czy hipoteza zerowa dla tego testu została odrzucona czy nie. W tym celu dokonujemy interpretacji wartości p.

¹⁰ Możliwe jest w SPSS uzyskanie widoku, który uwzględnia jedynie „ważne” korelacje, bez wyników po przekątnej oraz lustrzanego odbicia. Taki widok uzyskamy zaznaczając w menu analizy korelacji opcję „Pokaż tylko dolny trójkąt”.

Zasada interpretacji wartości p dla analizy korelacji jest następująca:

- Jeżeli $p < 0,05$ (wynik istotny statystycznie): odrzucamy hipotezę zerową o braku związków – **stwierdzamy występowanie związku między zmiennymi**.
- Jeżeli $p > 0,05$ (wynik nieistotny statystycznie): przyjmujemy hipotezę zerową o braku związku – **stwierdzamy brak związku między zmiennymi**¹¹.

PAMIĘTAJ

Istotny statystycznie ($p < 0,05$) wynik analizy korelacji wskazuje na występowanie związku między zmiennymi. Wynik nieistotny statystycznie ($p > 0,05$)* wskazuje na brak związku.*

* Przy założeniu $\alpha = 0,05$

W naszym przykładzie przedstawia się to w następujący sposób:

		Poziom odczuwanego stresu (PSS-10)	Ekstrawersja (NEO-FFI)	Samoocena (SES)
Poziom odczuwanego stresu (PSS-10)	Korelacja Pearsona	1	-0,279**	-0,339**
	Istotność (dwustronna)		0,005	<0,001
	N	100	100	100
Ekstrawersja (NEO-FFI)	Korelacja Pearsona	-0,279**	1	0,469**
	Istotność (dwustronna)	0,005		<0,001
	N	100	100	100
Samoocena (SES)	Korelacja Pearsona	-0,339**	0,469**	1
	Istotność (dwustronna)	<0,001	<0,001	
	N	100	100	100

** Korelacja istotna na poziomie 0.01 (dwustronnie).

Wynik istotny statystycznie ($p < 0,005$)

Rysunek 4. Zrzut ekranu z wydruku wyników w programie IMB SPSS przedstawiający interpretację wartości p w tabeli z wynikami analizy korelacji Pearsona.

¹¹ Znów, jest to uproszczenie. W praktyce wniosek taki możemy wyciągnąć jeśli moc testu jest wystarczająco wysoka ($1 - \beta \geq 0,80$). Jeśli moc testu jest niższa, wtedy wnioskujemy, że nie posiadamy wystarczających dowodów aby stwierdzić korelację lub jej brak.

Wszystkie trzy testowane związki okazały się istotne statystycznie, ponieważ w każdym przypadku $p < 0,05$. Dokładne wartości p dla poszczególnych korelacji wynoszą kolejno:

- Związek poziomu stresu z ekstrawersją: $p = 0,005$.
- Związek poziomu stresu z samooceną: $p < 0,001$.
- Związek ekstrawersji z samooceną: $p < 0,001$.

Hipoteza została zweryfikowana, teraz czas na dwa kolejne etapy interpretacji wyniku:

- W przypadku gdy wynik jest istotny statystycznie, wiemy jedynie, że występuje korelacja między zmiennymi, ale nie wiemy jeszcze jaki jest jej kierunek¹². W tym celu oceniamy znak współczynnika korelacji, aby określić czy związek jest dodatni czy ujemny.
- Wartość p wskazuje na istotność statystyczną wyniku, nie zaś na siłę korelacji między zmiennymi. Oby ją określić należy zinterpretować wartość współczynnika korelacji i na tej podstawie oszacować wielkość efektu.

Współczynnik korelacji może mieć wartość od -1 (całkowity ujemny związek) do 1 (całkowity dodatni związek). Wartość 0 oznacza brak związku.

Interpretacja kierunku korelacji jest następująca:

- Dodatnia korelacja wskazuje na wzrost wartości jednej zmiennej wraz ze wzrostem wartości drugiej zmiennej.
- Ujemna korelacja wskazuje na spadek wartości jednej zmiennej wraz ze wzrostem wartości drugiej zmiennej, przy czym ta interpretacja obowiązuje w obie strony¹³.

Istnieją różne sposoby oceny siły efektu dla korelacji. Poniżej przedstawiono jedną z najbardziej popularnych interpretacji (Cohen, 1988)¹⁴:

- $r \leq 0,1$ – brak efektu
- $0,1 < r \leq 0,3$ – efekt słaby
- $0,3 < r \leq 0,5$ – efekt umiarkowany
- $r > 0,50$ – efekt silny

¹² Wszystko to przy założeniu, że wykonujemy test dwustronny. W teście jednostronnym wynik istotny statystycznie wystąpi jedynie gdy kierunek korelacji będzie zgodny z hipotezą.

¹³ Przykładowo, ujemny związek między poziomem stresu a ekstrawersją oznacza zarówno wzrost stresu wraz ze spadkiem ekstrawersji jak i spadek stresu wraz ze wzrostem ekstrawersji.

¹⁴ Interpretacja dla współczynnika ρ Spearmana jest analogiczna jak dla r Pearsona

DEFINICJA

Wartość współczynnika korelacji jest miarą siły efektu dla analizy korelacji. Dzięki niemu możemy określić jak silny jest związek między zmiennymi – słaby ($0,1 < r \leq 0,3$), umiarkowany ($0,3 < r \leq 0,5$) lub silny ($d > 0,5$). Raportowanie siły efektu jest jednym z wymagań stylu APA.

W naszym przykładzie interpretacja, uwzględniająca zarówno kierunek jak i siłę związku, będzie zatem następująca¹⁵:

- Związek poziomu stresu z ekstrawersją był ujemny i słaby ($r = -0,23$), wskazuje to na spadek stresu wraz ze wzrostem ekstrawersji.
- Związek poziomu stresu z samooceną był ujemny i umiarkowany ($r = -0,34$), wskazuje to na spadek stresu wraz ze wzrostem samooceny.
- Związek ekstrawersji z samooceną był dodatni i umiarkowany ($r = 0,40$), wskazuje to na wzrost samooceny wraz ze wzrostem ekstrawersji.

Raportowanie wyników analizy korelacji Pearsona w stylu APA 7

Ogólne wytyczne dotyczące raportowania wyników w APA 7 zostały przedstawione [w tym artykule](#). Zaprezentowany jest w nim również sposób raportowania wyników analizy korelacji w postaci tabelarycznej.

Poniżej przedstawiono przykładowy sposób raportowania wyników omawianej jako przykład w tym tekście analizy w formie opisu tekstowego wykonanego według stylu APA7.

PAMIĘTAJ

Pod [tym linkiem](#) znajdziesz artykuł w którym wyjaśniamy najważniejsze podstawowe zasady raportowania wyników analizy statystycznej w stylu APA 7 wraz z przykładami raportowania wyniku analiz korelacji w tekście i tabeli.

¹⁵ Warto wspomnieć, że współczynników korelacji nie można między sobą wprost porównywać. Choć wartość $r = 0,40$ jest wyższa niż $r = -0,34$ (uwzględniamy wartość współczynnika, bez kierunku korelacji) to jedynie odpowiedni test statystyczny porównujący współczynniki korelacji (test Z Fishera) będzie w stanie zweryfikować hipotezę dotyczącą tej różnicy.

PRZYKŁAD

Celem sprawdzenia czy poziom postrzeganego stresu, poziom ekstrawersji oraz poziom samooceny były ze sobą powiązane, wykonano analizę korelacji r Pearsona.

Przeprowadzona analiza wykazała, że odczuwany stres korelował istotnie statystycznie oraz ujemnie z poziomem ekstrawersji ($r = -0,23$; $p = 0,019$). Oznacza to, że wraz ze wzrostem ekstrawersji poziom odczuwanego stresu malał. Siła tego związku była słaba.

Ponadto, odnotowano istotny statystycznie i ujemny związek pomiędzy nasileniem odczuwanego stresu a poziomem samooceny ($r = -0,34$; $p < 0,001$). Wskazuje to na wzrost odczuwanego stresu wraz ze spadkiem samooceny. Siła tego związku była umiarkowana.

Również istotny statystycznie okazał się związek pomiędzy poziomem ekstrawersji i samooceny ($p < 0,001$). Dodatnia wartość współczynnika korelacji ($r = 0,40$) wskazuje na wzrost samooceny wraz ze wzrostem ekstrawersji. Związek ten był umiarkowany.

PRZYKŁAD

Literatura

American Psychological Association. (2020). *Publication manual of the American Psychological Association 2020: the official guide to APA style* (7th ed.). American Psychological Association.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Spis rysunków

Rysunek 1. Zrzut ekranu z głównego okna programu SPSS przedstawiający sposób wyboru okna analizy korelacji z menu. (**strona 6**)

Rysunek 2. Zrzut ekranu z okna menu analizy korelacji w programie SPSS przedstawiający sposób ustawiania opcji analizy korelacji Pearsona. (**strona 7**)

Rysunek 3. Zrzut ekranu z wydruku wyników w programie SPSS przedstawiający zawartość tabeli z wynikami analizy korelacji Pearsona. (**strona 8**)

Rysunek 4. Zrzut ekranu z wydruku wyników w programie IMB SPSS przedstawiający interpretację wartości p w tabeli z wynikami analizy korelacji Pearsona. (**strona 9**)

Pogotowie Statystyczne Paweł Iwankowski
ul. prof. Stefana Hausbrandta 34/88
80-126 Gdańsk
NIP: 7412032970,
REGON: 280490493

tel. 501 599 278
info@pogotowiestatystyczne.pl



Autorzy:

Andrzej Jankowski

ajankowski@pogotowiestatystyczne.pl

Kacper Sawicki

sawicki.kacper@gmail.com

Korekta:

Paweł Krasa

Paweł Iwankowski

Zapoznaj się z naszą ofertą:
www.pogotowiestatystyczne.pl