Statystyki opisowe i normalność rozkładu

Wyjaśnienia oraz przykładowy wzorzec tabeli w standardzie APA 7

Statystyki opisowe to metody służące do opisu i podsumowania danych. Obliczamy je w różnych celach, choć z praktycznego punktu widzenia dwa są podstawowe. Pierwszym z nich jest eksploracja danych, czyli wstępna, ogólna analiza, pozwalająca ocenić poprawność zakodowania bazy danych oraz wyciągnąć podstawowe wnioski dotyczące poszczególnych zmiennych. Drugim jest przetestowanie założenia dotyczącego rozkładu normalnego, będącego warunkiem użycia niektórych testów statystycznych. W takiej sytuacji analizę statystyk opisowych uzupełnia się często wykonaniem testów normalności rozkładu, np. Kołmogorowa-Smirnowa lub Shapiro-Wilka.

Poniżej przedstawiamy przykładowy wzorzec raportowania wyników analizy statystyk opisowych oraz testu normalności rozkładu w tabeli wykonanej według standardu APA 7. Pod tabelą znajduje się również opis zastosowanych skrótów.

Tabela 1

Podstawowe statystyki opisowe wraz z testem Shapiro-Wilka mierzonych cech osobowości

Cecha osobowości	*M*	*Mdn*	*SD*	*Sk.*	*Kurt.*	*Min.*	*Maks.*	*W*	*p*
Neurotyczność	22,02	21,00	9,03	0,19	-0,31	16,00	43,00	0,99	0,734
Ekstrawersja	26,79	27,00	6,97	0,04	-0,56	13,00	43,00	0,99	0,722
Otwartość na doświadczenia	26,47	26,00	6,38	0,12	-0,37	10,00	40,00	0,98	0,362
Ugodowość	28,76	29,00	7,23	-0,03	0,20	10,00	45,00	0,98	0,300
Sumienność	30,88	32,00	7,00	-0,12	-0,68	17,00	46,00	0,98	0,227

Skróty zastosowane w tabeli:

Miary tendencji centralnej

M – wartość średniej arytmetycznej

Mdn – mediana – wartość dzieląca uporządkowany zbiór danych na pół

Miary rozproszenia

SD – odchylenie standardowe, miara rozproszenia wyników wokół średniej

Min. – najniższa wartość zbioru

Maks. – najwyższa wartość zbioru

Miary symetrii rozkładu

Skośność (***Sk.***) – miara symetrii rozkładu. Dla rozkładu normalnego wynosi 0. Gdy skośność jest większa od 0 mówimy o rozkładzie dodatnio skośnym / prawoskośnym (częstość wyników niskich jest większa niż wyników wysokich; $M > Mdn > Mo$). Gdy skośność jest mniejsza od 0 – rozkład jest ujemno skośny / lewoskośny – częstość wyników niskich jest mniejsza niż wyników wysokich ($M < Mdn < Mo$)

Kurtoza (***Kurt.***) – miara występowania wartości odstających. Podobnie jak w przypadku skośności, wartość 0 kurtozy wskazuje na kształt zbliżony do normalnego (rozkład mezokurtyczny). Wartości dodatnie kurtozy wskazują na rozkład leptokurtyczny, a im wyższa wartość kurtozy tym więcej pojawia się wartości bliskich skrajnych lub skrajnych. Z kolei ujemne wartości kurtozy (rozkład platykurtyczny) wskazują na brak wartości odstających

W ostatnich dwóch wierszach raportowania według przedstawionego wzorca przedstawiamy wyniki testu normalności rozkładu. Dwa najpopularniejsze takie testy to test Kołmogorowa-Smirnowa (statystyka D) oraz test Shapiro-Wilka (statystyka W). Ich działanie opiera się na porównaniu rozkładu danej zmiennej do rozkładu teoretycznego, jakim jest rozkład normalny. Jeśli wartość p dla tego testu jest mniejsza niż próg alfa (zwykle 0,05), tzn. mamy do czynienia z wynikiem istotnym

statystycznie, oznacza to, że rozkład zmiennej odbiega od rozkładu normalnego. Jednak wbrew występującemu często przekonaniu, wcale nie oznacza to, że dla tych zmiennych należy zrezygnować z wykorzystania testów parametrycznych (i np. zamiast nich wybrać testy nieparametryczne). Po pierwsze, testy te są odporne na złamanie poszczególnych założeń, a po drugie, wyniki testów normalności są nieprecyzyjne w przypadku względnie małych (zbyt mała moc testu) lub dużych (zbyt duża moc testu) próbach.

Większość autorów wskazuje, że testy parametryczne są odporne na złamanie założenia o normalności rozkładu, jeśli nie jest ono znaczące. Jednym ze sposobów oceny tego jest analiza wartości skośności i kurtozy. Jako „reguły kciuka" przyjmuje się, że na nieznaczne odchylenie od rozkładu normalnego wskazują wartość skośności i kurtozy mieści się w granicach wartości bezwzględnej 1 (między -1 a 1) lub wartości bezwzględnej 2 (między -2, a 2). Informacje te znajdziemy m.in. w poniższych pozycjach literaturowych:

George, D., Mallery, P. (2021). *IBM SPSS statistics 27 step by step: A simple guide and reference*. Routledge.

Trochim, W. M., & Donnelly, J. P. (2006). *The research methods knowledge base* (3rd ed.). Atomic Dog.

Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th). SAGE Publications.

Gravetter, F., Wallnau, L. (2014). *Essentials of statistics for the behavioral sciences* (8th ed.). Wadsworth.