

Jakub Szczerbuk

Szkoła Wyższa Psychologii Społecznej

Sylwia Bedyńska

Instytut Psychologii Polskiej Akademii Nauk

Tabelaryczne, graficzne i liczbowe sposoby podsumowywania zmiennych

W tym rozdziale znajdziemy:

- Informacje o sposobach podsumowania zmiennych:
 - Tabelarycznych – tabele częstości
 - Graficznych – wykresy słupkowe, kołowe, histogramy, skrzynkowe i kwantylowe
 - Liczbowych – statystyki opisowe: miary tendencji centralnej, rozproszenia i kształtu rozkładu
- Informacje o sposobach przeliczania wyników na jednostki standardowe (Z)

WPROWADZENIE

W poprzednim rozdziale zostały opisane podstawy pracy z programem SPSS, w tym między innymi sposób, w jaki powinniśmy wprowadzać dane zebrane w trakcie badań. W poniższym tekście zostaną przedstawione kolejne kroki, jakie należy wykonać przygotowując dane do analiz służących weryfikacji hipotez badawczych. W niniejszym rozdziale Czytelnik będzie miał możliwość zaznajomienia się z tabelarycznymi, graficznymi i liczbowymi sposobami opisu danych. Zrozumienie powyższych zagadnień i umiejętność wykorzystania ich w praktyce jest niezwykle istotna co najmniej z kilku powodów. Zaczniemy może od tego najbardziej prozaicznego – w każdym niemal raporcie z badań znajduje się część zatytułowana „Statystyki opisowe”. Można więc powiedzieć, że dzięki temu rozdziałowi Czytelnikowi będzie łatwiej spełnić wymogi formalne związane z pisaniem raportu z badań – będzie po prostu wiedział, co wpisać do tego podrozdziału. Drugim i tak naprawdę znacznie ważniejszym powodem jest możliwość bliższego przyjrzenia się danym. Bez tego początkowego etapu przeprowadzenie właściwych analiz statystycznych i ich adekwatna interpretacja byłyby bardzo trudne. Ponadto wiele przedstawionych tu pojęć jest częścią podstawowego „języka” badaczy, a więc ich znajomość ułatwi również komunikację. Podsumowując, zaprezentowane tu zagadnienia będą niezbędne zarówno w pracy nad własnymi danymi empirycznymi, jak również podczas lektury niemalże każdego artykułu naukowego.

Zajmijmy się Inteligencją Emocjonalną. Będąc zwolennikami podejścia, w którym traktuje się ją jako zdolność¹, możemy zdefiniować ją jako „zdolność do odbierania i wyrażania emocji, asymilowania emocji w przebieg procesów poznawczych, rozumienia i wykorzystywania emocji w procesach myślowych oraz regulacji emocji (zarówno swoich, jak i innych)” (Mayer, Salovey i Caruso, 2000). Takie ujęcie Inteligencji Emocjonalnej determinuje rodzaj narzędzi wykorzystanych przez nas w badaniu, ale o tym słów kilka za chwilę.

Załóżmy, że chcemy sprawdzić, czy poziom Inteligencji Emocjonalnej jest powiązany z poziomem „wysportowania” osób badanych. Można wskazać przesłanki pozwalające sądzić, że powyższe zmienne są ze sobą powiązane. Nie od dziś wiadomo, że ćwiczenia fizyczne wpływają na polepszenie nastroju. Skoro osoby o wysokiej Inteligencji Emocjonalnej są bardziej świadome swoich emocji, to być może lepiej radzą sobie z regulacją własnego nastroju, wykorzystując ćwiczenia fizyczne jako element tej regulacji.

Aby móc zbadać opisaną wyżej relację, udaliśmy się w sobotni poranek do pobliskiego parku, w którym wielu ludzi uprawia, zyskując coraz większą popularność, jogging. Postanowiliśmy dokonać pomiaru dwóch interesujących nas zmiennych: (a) poziomu „wysportowania” deklarowanego przez osoby badane oraz (b) poziomu Inteligencji Emocjonalnej mierzonej za pomocą testu MSCEIT. Po analizie odpowiedzi udzielonych przez osoby badane udało nam się je podzielić na cztery grupy: (a) ludzi niećwiczących, przebywających w parku z powodu

¹ Wśród badaczy zajmujących się Inteligencją Emocjonalną można wyróżnić dwie główne „frakcje”: (a) grupę traktującą Inteligencję Emocjonalną jako zdolność (mierzoną testami) oraz (b) grupę, która uznaje ją za cechę (mierzoną inwentarzami).

PRZYGOTOWANIE DANYCH DO ANALIZY

dziecka lub też zwierzęcia domowego, (b) ludzi ćwiczących „weekendowo” tudzież „świętecznie”, (c) zapalonych sportowców amatorów oraz (d) zawodowych sportowców (okazało się bowiem, że w naszym parku akurat biegła drużyna rugby oraz kilku innych mniej lub bardziej znanych sportowców). Dane znajdują się w pliku *Rozdział3_.sav* na stronie książki.

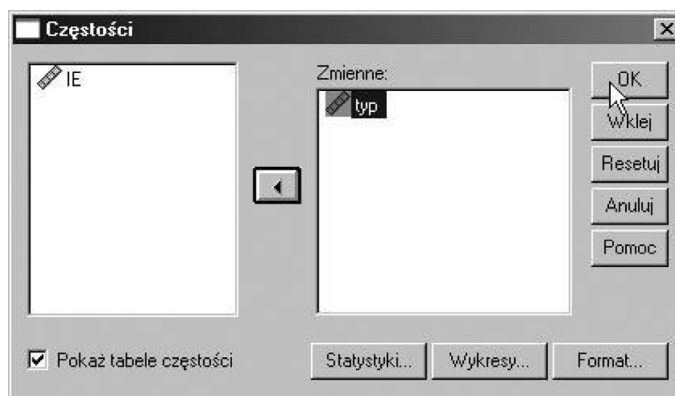
CZĘSTOŚCI

Tabela częstości

Pierwszym i najprostszym sposobem podsumowania danych jest **tabela CZĘSTOŚCI**. Dzięki niej będziemy mogli przyjrzeć się zmiennej, której wskaźnikiem jest deklarowany poziom „wysportowania”. Zmienna ta jest mierzona na skali porządkowej (informacje na temat skal pomiarowych znajdzie Czytelnik w rozdziale 1). Tabela CZĘSTOŚCI podsumowuje informacje o liczbie osób, które wybrały daną wartość zmiennej, a zatem o częstości wystąpienia tej wartości zmiennej w zbiorze danych. Aby ją uzyskać, klikamy kolejno: menu ANALIZA, OPIS STATYSTYCZNY i wreszcie CZĘSTOŚCI (rys. 3.1):



Rys. 3.1. Wybór opcji CZĘSTOŚCI w menu ANALIZA – Opis statystyczny.



Rys. 3.2. Definiowanie zmiennych poddawanych analizie częstości.

3 • TABELARYCZNE, GRAFICZNE I LICZBOWE SPOSOBY...

Kolejnym krokiem jest przeniesienie do prawego okna zmiennej TYP oraz kliknięcie przycisku OK (rys. 3.2). Raport, który otrzymujemy, przedstawiony jest w formie tabeli na rys. 3.3.

typ		
N	Ważne	245
	Braki danych	0

typ					
		Częstość	Procent	Procent ważnych	Procent skumulowany
Ważne	niećwiczący	71	29,0	29,0	29,0
	ćwiczący "weekendowo"	56	22,9	22,9	51,8
	sportowiec "amator"	61	24,9	24,9	76,7
	sportowiec zawodowiec	56	22,9	22,9	99,6
	11,00	1	,4	,4	100,0
	Ogółem	245	100,0	100,0	

Rys. 3.3. Tabela statystyki częstości dla zmiennej TYP.

Widzimy, że SPSS przygotował nam dwie tabele. Z mniejszej – STATYSTYKI – możemy dowiedzieć się, ile mamy osób badanych, jak również uzyskujemy informację o ilości braków danych (więcej o brakach danych w Rozdziale 2). Jak widać, udało nam się wprowadzić dane dotyczące 245 osób badanych i nie mamy żadnych braków danych. Druga, bardziej rozbudowana tabela (TYP), zawiera znacznie więcej informacji. Spójrzmy na pierwszą kolumnę – znajdują się w niej etykiety czterech wyróżnionych grup osób, ale oprócz tego cyfra „11”. Trzeba się zastanowić, czy możliwe jest uzyskanie wartości 11 w analizowanej zmiennej, a zatem trzeba określić możliwy zakres zmiennej (jej minimum i maksimum). Jeżeli jakaś wprowadzona do danych wartość przekracza możliwy zakres, to może to oznaczać, że popełniliśmy błąd wprowadzając dane. Skoro grupy kodowane były wartościami od 1 do 4, to możliwe, że przez pomyłkę, zamiast 1 wpisaliśmy 11. Zanim wrócimy do EDYTORA DANYCH, by skorygować tę pomyłkę, spójrzmy jeszcze na drugą kolumnę tabeli CZĘSTOŚCI – mamy w niej przedstawione liczebności poszczególnych wartości zmiennej. Ta informacja przyda nam się, by sprawdzić, czy tylko raz zdarzyła nam się taka pomyłka. Na podstawie tej tabeli możemy stwierdzić, że w zbiorze danych jest tylko jedna osoba z wartością 11. Spróbujmy znaleźć tę wartość w EDYTORZE DANYCH (rys. 3.4).

PRZYGOTOWANIE DANYCH DO ANALIZY

14 :

	typ	IE	var
1	1,00	65,00	
2	1,00	66,00	
3	11,00	66,00	
4	1,00	67,00	

Rys. 3.4. Błędnie wpisana wartość w EDYTORZE DANYCH.

Jak widać na rys. 3.4, osoba nr 3 w kolumnie TYP, zamiast 1 ma przypisaną wartość 11. Skorygujmy ten wynik i wpiszmy zamiast 11 wartość 1² i skorzystajmy jeszcze raz z opcji CZĘSTOŚCI. Poczynione poprawki przyniosły spodziewany efekt – błędna wartość zniknęła, a liczba osób, które wybrały wartość 1, zwiększyła się o jedną (tabela na rys. 3.5).

typ

		Częstość	Procent	Procent ważnych	Procent skumulowany
Ważne	niećwiczący	72	29,4	29,4	29,4
	ćwiczący „weekendowo”	56	22,9	22,9	52,2
	sportowiec „amator”	61	24,9	24,9	77,1
	sportowiec zawodowiec	56	22,9	22,9	100,0
	Ogółem	245	100,0	100,0	

Rys. 3.5. Tabela częstości dla zmiennej TYP po korekcie błędnie wpisanej wartości tej zmiennej.

Patrząc teraz na kolumnę CZĘSTOŚCI, możemy powiedzieć, że w badanej próbie znalazły się: 72 osoby niećwiczące, 56 osób ćwiczących „weekendowo”, 61 sportowców amatorów oraz 56 zawodowców. Znamy już więc liczebność poszczególnych grup. Dzięki kolumnie PROCENT będziemy mogli dodatkowo powiedzieć, jaki procent wszystkich badanych stanowią poszczególne grupy. Jak widać, osoby niećwiczące stanowią 29,4% wszystkich badanych, podczas gdy najmniej, bo 22,9% ćwiczący „weekendowo” oraz zawodowcy. Spójrzmy teraz na kolejną kolumnę – PROCENT WAŻNYCH. Aby ją opisać, posłużymy się przykładem: wyobraźmy sobie, że w naszym zbiorze mamy 10 braków danych – wówczas wydruk wyglądałby następująco (tabela na rys. 3.6).

² Jeśli nie wiemy jak skorygować błędna wartość, można ją zmienić na kod braków danych. W takim wypadku przyjęło się deklarować wartość 9, 99 lub inną kombinację dziewiątek. Ta konwencja nie jest powszechnie stosowana. Więcej na ten temat w Rozdziale 2.

Statystyki

typ		
N	Ważne	235
	Braki danych	10

typ

		Częstość	Procent	Procent ważnych	Procent skumulowany
Ważne	niećwiczący	62	25,3	26,4	26,4
	ćwiczący "weekendowo"	56	22,9	23,8	50,2
	sportowiec "amator"	61	24,9	26,0	76,2
	sportowiec zawodowiec	56	22,9	23,8	100,0
	Ogółem	235	95,9	100,0	
Braki danych	9,00	10	4,1		
Ogółem		245	100,0		

Rys. 3.6. Różnica między procentem i procentem ważnych w tabelach częstości.

W kolumnie PROCENT WAŻNYCH mamy podsumowane wartości procentowe dla poszczególnych grup, gdzie punktem odniesienia są jedynie obserwacje niebędące brakami danych. Innymi słowy braki danych nie są w ogóle brane pod uwagę. Spójrzmy na poprzednio omawianą kolumnę – PROCENT: widzimy, że wartości procentowe dla grupy niećwiczących zmniejszyły się w porównaniu z wartościami zaprezentowanymi w tabeli na rys. 3.5. Stało się tak, gdyż braki danych „odebrały” jej kilka punktów procentowych. W kolumnie PROCENT WAŻNYCH taka sytuacja nie ma miejsca – i jest to jedyna różnica pomiędzy tymi dwoma kolumnami. Wróćmy teraz do wydruku, aby omówić ostatnią już kolumnę PROCENT SKUMULOWANY (tabela na rys. 3.6).

Kolumna PROCENT SKUMULOWANY może być przydatna, gdy chcemy szybko sprawdzić, jaki procent całej próby stanowi grupa osób o określonej wartości zmiennej i wartościach niższych. Wyobraźmy sobie, że chcemy się dowiedzieć, jaki procent stanowią ćwiczące „weekendowo” lub rzadziej (wartości 1 i 2 łącznie). Aby znaleźć odpowiedź, patrzymy w kolumnie PROCENT SKUMULOWANY w wierszu, gdzie znajduje się grupa osób ćwiczących „weekendowo”. Wartość procenta skumulowanego wynosi 50,2%, co oznacza, że tyle właśnie osób ćwiczy „weekendowo” lub rzadziej (czyli nie ćwiczy wcale). Spróbujmy rozważyć inny przykład: chcemy się dowiedzieć, jaki procent próby stanowią osoby, które nie są zawodowcami (a więc wszyscy oprócz osób z czwartej grupy). Szukamy więc procenta skumulowanego dla wartości zmiennej oznaczającej „sportowców amatorów” i odczytujemy wartość: 76,2%. Tak więc kolumna procent skumulowany jest tak skonstruowana, że pokazuje dodane do siebie procenty ważnych danego wiersza i wszystkich znajdujących się wyżej w tabeli. Warto jeszcze wspomnieć, że procent skumulowany ma jakikolwiek sens pod warunkiem, że zmienna jest mierzona na skali co najmniej porządkowej.

Jak widać, **dzięki tabelom częstości uzyskujemy informacje na temat omawianej zmiennej** (w liczebnościach i szczególnie wartościach procentowych). **Możemy też odnaleźć drobne błędy popełnione w trakcie wpisywania danych, a następnie skorygować je w EDYTORZE DANYCH.**

Zawsze rozpoczynamy więc analizy od obejrzenia częstości, bo pomyłki to rzecz ludzka i w dużych zbiorach danych trudno ich uniknąć.

*Graficzny
opis zmiennych*

GRAFICZNY OPIS ZMIENNYCH

Wykresy

Oprócz suchych wartości liczbowych warto także podsumowywać wartości zmiennej w formie graficznej. Program SPSS umożliwi wykonanie kilku podręcznych wykresów w oknie CZĘSTOŚCI. Wracamy więc do tego okna klikając kolejno: ANALIZA, OPIS STATYSTYCZNY i wreszcie CZĘSTOŚCI. W dolnej części okna znajduje się przycisk WYKRESY. Klikamy go i otwiera się nowe okno dialogowe (rys. 3.7).



Rys. 3.7. Definicja wykresów w oknie CZĘSTOŚCI.

Do dyspozycji użytkownika jest kilka rodzajów wykresów. Te, z których możemy w tym momencie skorzystać, to wykres słupkowy oraz wykres kołowy (trzeci typ – histogram jest przeznaczony do wizualizacji zmiennych mierzonych na skalach ilościowych – zostanie więc omówiony w dalszej części rozdziału). Wspomniane wykresy to nic innego jak graficzna prezentacja tych wyników, które użyliśmy za pomocą tabel częstości. Zaczijmy od **wykresu słupkowego**.

SPSS daje możliwość wyboru, czy wartości przedstawione na wykresie będą przedstawiać liczebności każdej z wartości zmiennej (opcja CZĘSTOŚCI) czy też ich procentowy udział (opcja PROCENTY). Niektórzy z pewnością zauważą, że wybierając opcję CZĘSTOŚCI uzyskamy wizualizację wyników zaprezentowa-

Wykres słupkowy