

Barbara Kossakowska
Sylwia Bedyńska
Szkoła Wyższa Psychologii Społecznej

Wprowadzanie danych i podstawy pracy z programem IBM SPSS Statistics

W tym rozdziale znajdziemy:

- Podstawowe informacje o programie IBM SPSS Statistics
- Opis edytorów (danych, raportów i poleceń)
oraz ich podstawowych funkcji
- Sposoby definiowania i wpisywania różnego rodzaju danych

CZĘŚĆ I
PRZYGOTOWANIE
DANYCH DO ANALIZY
STATYSTYCZNEJ

ROZDZIAŁ

2

WPROWADZENIE

IBM SPSS Statistics jest programem działającym w dominującym obecnie środowisku Microsoft Windows oraz w rzadziej spotykanych systemach operacyjnych Linux i MAC OS. Każdy zaznajomiony z tymi systemami operacyjnymi użytkownik zauważy, że program SPSS wykazuje duże podobieństwo w podstawowych funkcjach i trybach pracy do innych programów pracujących w tych środowiskach. Główny człon nazwy – SPSS – to skrót od *Statistical Package for Social Sciences*. Pakiet ten wykorzystywany powszechnie w naukach społecznych: psychologii, socjologii czy pedagogice, ale służy także do analizy danych z systemów operacyjnych, transakcyjnych, danych sprzedażowych, wyszukiwania zagrożeń w transakcjach internetowych, poszukiwania nadużyć w systemach rejestracji usług medycznych. Widać więc, że jego zastosowanie znacznie wykracza poza zagadnienia sugerowane nazwą pakietu.

W skład pakietu IBM SPSS Statistics wchodzi kilkanaście modułów: część z nich to moduły uniwersalne, takie jak IBM SPSS Statistics Base czy IBM SPSS Regression, inne są bardziej specjalistyczne i grupują statystyki najczęściej używane w pewnym obszarze wiedzy, np. IBM SPSS Direct Marketing czy IBM SPSS Forecasting. Te ostatnie moduły są dość mocno zautomatyzowane, tak by odciążyć użytkownika z konieczności samodzielnego wyboru statystyk.

Właściwie co roku pojawia się nowa wersja programu IBM SPSS Statistics, więc warto wspomnieć, że książka jest dedykowana 20 wersji programu, choć prawdopodobnie większość operacji wykonuje się w podobny sposób także w wersjach niższych niż 20.0. Zmiany mają bowiem raczej charakter ewolucyjny niż rewolucyjny i najczęściej polegają na dobudowywaniu kolejnych modułów. Obsługa programu jest podobna jak w innych programach, z którymi użytkownik miał już styczność. Poszczególne opcje można wybierać za pomocą myszki bądź też korzystając z klawiatury komputerowej. Skróty klawiszowe CTRL+C (kopiuj) lub CTRL+V (Wklej), do których niektórzy użytkownicy Windowsa przywykli, działają także w SPSS-ie.

Omówmy najpierw główne elementy konstrukcji programu. Praca z pakietem IBM SPSS Statistics odbywa się w kilku rodzajach okien nazywanych edytorami. Są to edytory: danych, raportów, wykresów i poleceń. Część edytorów ma autonomiczną funkcję i może zostać zapisana jako samodzielny plik. Dotyczy to edytora danych, raportów i poleceń. Inne edytory: np. edytor wykresów otwierane są w raporcie a następnie zamykane po dokonaniu operacji na obiekcie, który został w nich wyedytowany. Edytor Danych służy do wpisywania surowych danych oraz przygotowania swego rodzaju „słowniczka” etykiet zmiennych i ich wartości. Edytor Danych ma rozszerzenie .sav, a w programie Windows jest oznaczony specyficzną ikoną (rysunek 2.1). Edytor Raportów to okno, w którym

pojawiają się efekty intensywnej pracy obliczeniowej, czyli wyniki zamówionych analiz w postaci tabel i wykresów. Informacje zawarte w tym edytorze mogą zostać także wzbogacone przez autora o szczegółowe komentarze i opisy, a tabele mogą być edytowane w bardzo elastyczny, szybki i wygodny sposób poprzez tabele przestawne. Ten edytor ma rozszerzenie .spo oraz charakterystyczną ikonę (rysunek 2.2). Edytor Poleceń to szczególnie skuteczne narzędzie dla bardziej zaawansowanych użytkowników, dla których korzystanie z programu za pomocą górnego menu przestało być efektywne ze względu na dużą liczbę przeprowadzanych analiz. Każda analiza, przekształcenie czy wykres może zostać wykonane dwojako: za pomocą wyboru odpowiedniej opcji w menu lub też poprzez wpisanie polecenia w Edytorze Poleceń. Treść poleceń może zostać zapisana w postaci pliku o rozszerzeniu .sps (ikona na rysunku 2.3).



Rysunek 2.1. Ikona pliku danych



Rysunek 2.2. Ikona pliku raportów



Rysunek 2.3. Ikona pliku poleceń

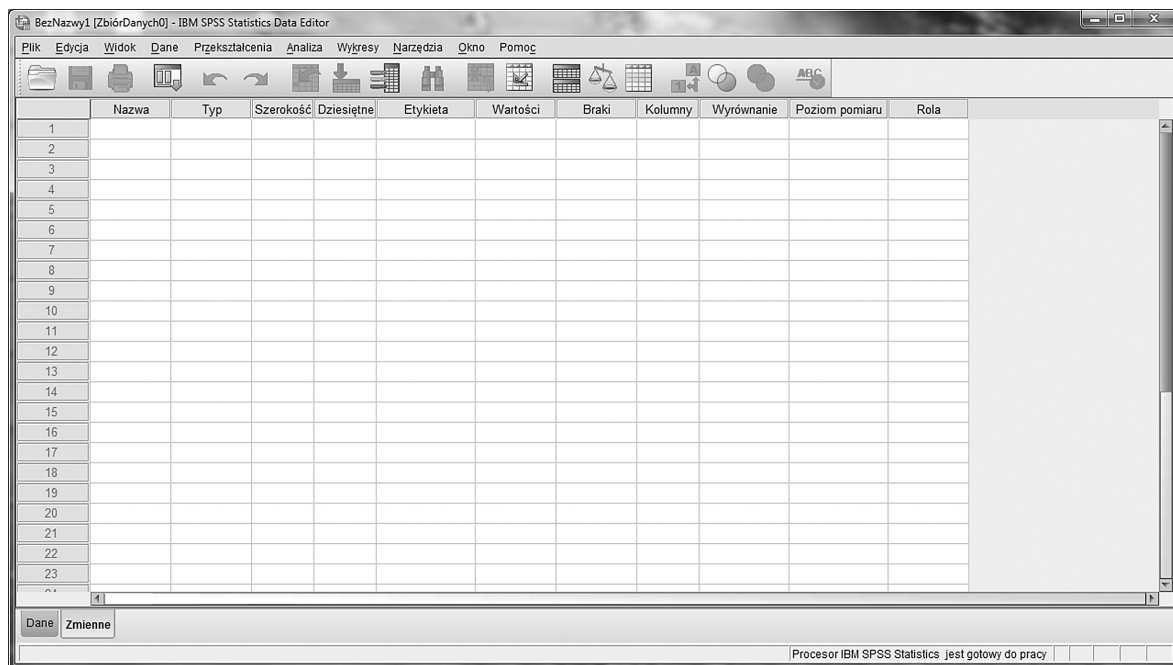
EDYTOR DANYCH

Jako pierwszy omówimy dokładnie Edytor Danych, gdyż zwykle analizy danych rozpoczynamy od ich wprowadzenia do programu statystycznego. Edytor Danych wygląda podobnie do arkusza kalkulacyjnego, choć istnieją też pewne różnice. Niezwykle istotne jest to, że wpisując dane, musimy pamiętać o kilku ogólnych zasadach. Zasady te są identyczne niezależnie od tego, do jakiego programu wprowadzamy dane, a więc dotyczą także takich sytuacji, gdy dane wprowadzamy do programu Excel czy innych arkuszy kalkulacyjnych. Zaczniemy od tych ogólnych zasad, by potem przejść do szczegółowych rozwiązań Edytora Danych.

Zastanówmy się najpierw, co to są dane. Otóż dane to informacje o osobie badanej, przy czym każda pojedyncza informacja stanowi osobny element danych. Wprowadzając dane, należy pamiętać o tym, że wyniki jednej osoby muszą się znaleźć w jednym (i tylko jednym) wierszu. Nawet jeżeli badamy tę samą osobę dwukrotnie tym samym testem, to jej odpowiedzi z obu testów muszą się znaleźć

w jednym wierszu w sąsiadujących kolumnach – najpierw wyniki pierwszego pomiaru, potem wyniki drugiego. Każdy kawałek informacji o tej osobie stanowi zmienną. Musimy więc się zastanowić, ile porcji informacji o osobie badanej mamy i dzięki temu określić liczbę zmiennych. Tak więc dane to zestaw zmiennych. Zmienne umieszczamy w kolumnach. Pamiętajmy, że ważne jest, aby wprowadzić do programu SPSS wszystkie dane jakie posiadamy – tak zwane dane surowe, na przykład odpowiedzi osoby na wszystkie pytania ankiety, a nie jedynie łączny wynik. Oczywiście wpisanie jedynie sumarycznego wyniku jest szybsze, ale niestety nie jest lepsze, gdyż uniemożliwia wykonanie niektórych analiz, a także zwiększa ryzyko popełnienia błędów w sumowaniu wyników. Trzeba także pamiętać, że program IBM SPSS Statistics, mimo iż pozwala wprowadzać dane tekstowe, to nie jest ich w stanie w takiej formie analizować. Do analizy jakościowej dedykowane są inne specjalistyczne programy. Jedynym językiem, który program rozumie są liczby. A zatem pojawia się konieczność „przełożenia” tekstowych informacji o osobie (na przykład jej płci) na informacje liczbowe, poprzez **kodowanie danych**. Sposób kodowania ustalamy zwykle sami, przypisując odpowiedziom osoby badanej wartości liczbowe, gdy mamy do czynienia ze zmiennymi jakościowymi. Zmienne ilościowe nie muszą być specjalnie kodowane, bo występują już w postaci liczbowej, którą wpisujemy do danych. Do problemu kodowania powrócimy w kolejnych częściach rozdziału.

KODOWANIE DANYCH



Rysunek 2.4. Okno Edytora Danych – widok ZMIENNE

Wróćmy teraz do Edytora Danych. Jego okienko ma dwie zakładki w lewym dolnym rogu. Te zakładki mają nazwy DANE oraz ZMIENNE. Kliknięcie zakładki powoduje otwarcie odpowiedniego okna. Zaczniemy od okna ZMIENNE (rysunek 2.4) stanowiącego rodzaj „słowniczka”, w którym użytkownik programu ma możliwość szczegółowego opisanie danych wprowadzanych do programu, nazw zmiennych i sposobu ich kodowania. Choć niezbędne jest tylko wprowadzenia nazw zmiennych, zachęcamy gorąco do możliwie dokładnego uzupełnienia informacji w tej zakładce, bo ułatwia to zarówno orientację w pliku z danymi, jak i późniejsze odczytywanie wyników analiz w Edytorze Raportów.

OKNO ZMIENNE

Zacznijmy zatem od wprowadzenia nazw zmiennych. W zakładce ZMIENNE w pierwszej kolumnie możemy umieścić nazwę zmiennej. IBM SPSS STATISTICS wymaga jednak, by nazwa zmiennej spełniała następujące reguły:

- ◆ **Nazwa zmiennej** nie może przekraczać pewnej liczby znaków – wprowadzone nazwy muszą mieć długość 64 bajtów (zwykle to 64 znaki). W starszych wersjach programu, na przykład 11,5 i niższych, ograniczenie jest bardzo restrykcyjne i nazwa zmiennej mogła zawierać jedynie 8 znaków. Nazwa zmiennej powinna być krótka, ale najlepiej, jeśli oddaje jej znaczenie.
- ◆ W nazwie nie wolno stosować znaków specjalnych: spacji, oraz znaków: !, %, ^, &, *, +, /, -, =, ?, ‘, ,, (), :, ;.
- ◆ Nazwa zmiennej zawsze musi zaczynać się literą i nie może kończyć się kropką. Kropka może się pojawić w środku nazwy zmiennej.
- ◆ Nie można wpisać dwóch zmiennych o tej samej nazwie.

NAZWA ZMIENNEJ

Jeśli wpisujemy nazwę niezgodną z powyższymi regułami, IBM SPSS STATISTICS uniemożliwi nam zapis i zasugeruje zmianę nazwy za pomocą specjalnego komunikatu.

Kolejną kolumną w oknie ZMIENNE jest **TYP**. Nie będziemy omawiać szczegółowo wszystkich możliwości w tym oknie, a jedynie zasygnalizujemy dwa różne typy i konsekwencje ich zastosowania dla dalszych analiz. Pierwszym typem zmiennych – najczęściej stosowanym – są zmienne liczbowe, które można opisać za pomocą **typu numerycznego**. Drugim typem są zmienne w postaci wpisanych do programu słów oznaczających płeć, kolor oczu czy inne zmienne jakościowe. Typ tekstowy, gdy wpisujemy dane zebrane w pytaniach otwartych, ma jedną poważną wadę. Jeśli typ zmiennej został zdefiniowany jako **typ tekstowy**, to program IBM SPSS STATISTICS nie jest w stanie przeprowadzić żadnych analiz, gdyż jak wspomnieliśmy wcześniej, program potrafi zrozumieć tylko informacje liczbowe. Trzeba też pamiętać, że typ ma znaczenie dla szerokości zmiennej. Jeśli chcemy wpisać dane tekstowe, takie jak chociażby imię i nazwisko osoby badanej, to musimy odpowiednio zdefiniować szerokość zmiennej, tak by zmieściły się w niej informacje o większej liczbie znaków. Można to zrobić w kolumnie **SZEROKOŚĆ**. Za pomocą strzałek z prawej strony okna możemy zwiększyć lub zmniejszyć szerokość danej zmiennej.

TYP NUMERYCZNY

TYP TEKSTOWY

Kolumna DZIESIĘTNE, znajdująca się w zakładce ZMIENNE, służy do ustalenia liczby miejsc dziesiętnych wyświetlanych w oknie DANE. Jak można się domyślać, ta kolumna ma znaczenie tylko w przypadku zmiennych numerycznych i jest istotna dla wygody użytkownika. Ta opcja ustawień zmienia bowiem tylko sposób wyświetlania wartości, ale nie jest w stanie ich zmodyfikować. Ważne: jeśli chcemy dane eksportować do Excela, to musimy ustawić odpowiednio miejsca dziesiętne, tak by w Excelu pojawiły się właściwe wartości. Jeśli ustawimy miejsca dziesiętne na jedno, a wartości liczbowe będą miały dokładność do dwóch miejsc, to w Excelu pojawią się zaokrąglone wartości, a nie te oryginalne, co może powodować błędy w obliczeniach. Na przykład, gdy przestawimy programie IBM SPSS Statistics wyświetlanie wartości dziesiętnych do jednego miejsca po przecinku, a wartość w danych wynosi 0,55, to po imporcie do Excela zostanie pokazana wartość 0,5.

ETYKIETA

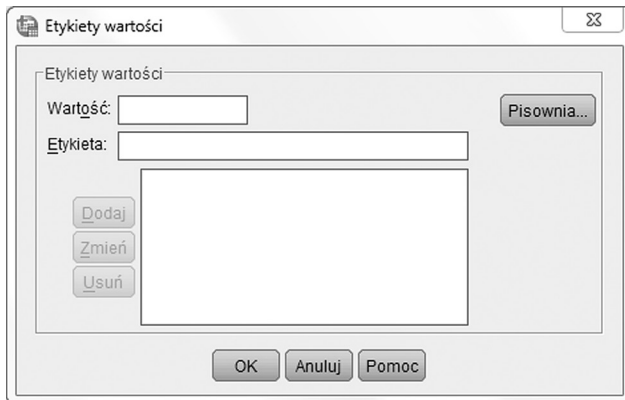
Kolejną bardzo ważną informacją o zmiennej jest **ETYKIETA**. Ponieważ nazwy zmiennych muszą być krótkie (choćby ze względu na wygodę posługiwania się nimi), to warto zapisać także bardziej szczegółową informację o tym, co oznaczają kolejne zmienne, na przykład pełne brzmienie pytań czy stwierdzeń ankiety, którą się posługujemy. Nie ma tu ograniczeń jeśli chodzi nie tylko o liczbę znaków, lecz także ich rodzaj, a więc można swobodnie korzystać ze znaków interpunkcyjnych. W etykietach warto też wpisać rodzaj jednostki pomiarowej, jaką się posługujemy, na przykład, czy wiek jest kodowany w latach czy miesiącach, gdyż w przypadku pewnych specyficznych grup osób (choćaby małych dzieci) może to nie być oczywiste.

WARTOŚCI

Kolejna kolumna – **WARTOŚCI** – ma istotne znaczenie dla czytelności wydruku, jaki pojawi się w Edytorze Raportów po wykonaniu analiz. Kolumna ta pozwala na zapisanie informacji o tym, jakimi wartościami są kodowane poszczególne opcje odpowiedzi. Załóżmy, że na pytanie „Jak często myjesz zęby?” możliwe są następujące odpowiedzi: bardzo często, często, przeciętnie często, rzadko, bardzo rzadko. Jak już wspominałyśmy wcześniej, aby IBM SPSS STATISTICS mógł wykonać jakiegokolwiek obliczenia, niezbędne jest przyporządkowanie każdej odpowiedzi wartości liczbowej. Ponieważ odpowiedzi te możemy uporządkować (zmienna jest mierzona na skali porządkowej), to wartości, którymi je zakodujemy, powinny odpowiadać temu porządkowi. Możemy więc zakodować te odpowiedzi następująco: 5 – bardzo często, 4 – często, 3 – przeciętnie często, 2 – rzadko, 1 – bardzo rzadko. A zatem każdej wartości liczbowej przypisana jest określona etykieta słowna. Jeśli klikniemy odpowiedni wiersz w kolumnie **WARTOŚCI**, w okienku pojawi się szary kwadracik (z prawej strony). Jego kliknięcie pozwala otworzyć okno, w którym w części **WARTOŚĆ** możemy wpisać liczbę, którą oznaczona jest dana odpowiedź, a poniżej, w oknie **ETYKIETA**, pełne brzmienie odpowiedzi. Aby wpisać kolejną parę – wartość i odpowiadającą jej etykietę – należy kliknąć na przycisk **DODAJ**, co spowoduje przeniesienie tej pary do dużego okna

ETYKIETA

poniżej. Jeśli popełnimy błąd, wystarczy kliknąć na błędną parę, która pojawi się ponownie w okienkach: WARTOŚĆ i ETYKIETA. Można ją wtedy skorygować, a korektę zaakceptować przyciskiem ZMIENŃ, który umieści odpowiednie informacje ponownie w dolnym okienku. Jeśli cała para składająca się z wartości i etykiety jest błędna, wystarczy na nią kliknąć i wybrać przycisk USUŃ.

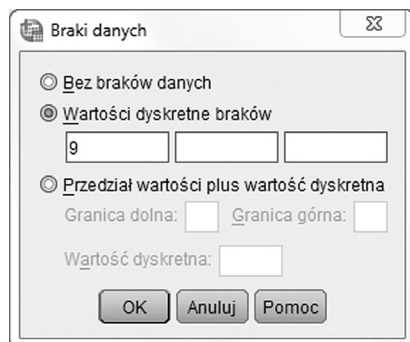


Rysunek 2.5. Okno definiowania Wartości zmiennych

Często podczas wprowadzania danych stykamy się z problemem braku odpowiedzi. Niekiedy badany przeoczy pytanie lub celowo je opuści. Dobrze jest „poinformować” IBM SPSS STATISTICS o braku danych. Można zostawić puste miejsce w DANYCH lub też nadać brakowi szczególną wartość. W niektórych środowiskach, zwłaszcza badaczy zajmujących się badaniami sondażowymi, przyjęło się oznaczać **braki danych** cyfrą 9 lub jej kombinacją, czyli 99, 999, 9999 itd. Trzeba pamiętać, że wartość oznaczająca brak danych musi wykraczać poza wartości zmiennej. Gdy zmienna przyjmuje wartości o 1 do 5, to brak danych najprościej zakodować jako 9. Gdy jednak zmienna przyjmuje wartości od 1 do 10, to trzeba zastosować 99. Jeśli już ustalimy, jaką wartością będzie kodowany brak danych, musimy o tym poinformować program IBM SPSS STATISTICS, aby tę wartość traktował specjalnie. Musimy zatem zadeklarować tę wartość (lub te wartości) w kolumnie BRAKI DANYCH w tym wierszu, który odnosi się do zmiennej zawierającej niepełne dane. Gdy klikniemy na pole w tej kolumnie, w odpowiednim wierszu z prawej strony pola pojawi się szary kwadracik z trzema kropkami. Kliknięcie na kwadracik spowoduje pojawienie się okna w kilkoma opcjami, z których najczęściej używana jest DYSKRETNE WARTOŚCI BRAKÓW DANYCH (rysunek 2.6). W znajdujących się w tej opcji trzech polach możemy wpisać trzy wartości liczbowe, których użyliśmy do oznaczenia braków danych.

BRAKI DANYCH

Kolejne dwie kolumny pozwalają definiować dwa istotne aspekty wyglądu okna z danymi, a mianowicie szerokość poszczególnych zmiennych i sposób

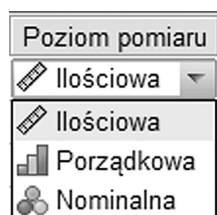


Rysunek 2.6. Okno definiowania braków danych

wyrównywania tekstu w widoku DANE (odpowiednio KOLUMNY i WYRÓWNANIE). Dobór odpowiednich opcji zostawiamy już czytelnikowi, gdyż jest to głównie kwestia estetyki.

W poprzednim rozdziale opisane zostały rodzaje skal pomiaru zmiennych. Przypominamy, że możemy mierzyć zmienne na skalach jakościowych i ilościowych. Skale jakościowe to nominalna i porządkowa, do ilościowych zaliczamy skalę przedziałową i stosunkową. Skale pomiarowe definiują **POZIOM pomiaru**. W przypadku programu IBM SPSS STATISTICS mamy możliwość zdefiniowania 3 poziomów: NOMINALNY, PORZĄDKOWY i ILOŚCIOWY (rysunek 2.7). W przypadku skali ILOŚCIOWEJ nie ma rozróżnienia między skalą przedziałową (interwałową) a stosunkową (ilorazową), gdyż obie skale umożliwiają stosowanie tych samych statystyk. Podział ten nie ma więc większego znaczenia dla analiz wykonywanych na danej zmiennej. Definiowanie odpowiedniej skali dla każdej zmiennej jest informacją wyłącznie do naszej wiadomości, gdyż IBM SPSS STATISTICS niestety nie potrafi rozpoznawać skal pomiarowych, niezależnie od tego, co wpisujemy w tej kolumnie.

Teraz jeszcze ostatnia opcja – ROLA. W tej części okna ZMIENNE możemy zdefiniować, jaką funkcję pełni określona zmienna w badaniu. Jak wspominałyśmy w poprzednim rozdziale, w zależności od schematu badania zmienna może



Rysunek 2.7. Definiowanie poziomu pomiaru zmiennej – dostępne opcje to poziom ilościowy, porządkowy i nominalny

odgrywać rolę niezależnej, zależnej, może określać sposób podziału osób badanych na podgrupy itd. Ze względu na automatyzację niektórych funkcji w programie zdefiniowanie roli pozwala później łatwiej tworzyć wykresy i wykonywać analizy, bo to program podpowiada odpowiednie opcje, widząc jaką funkcję pełni zmienna. Umieszcza też zmienne w odpowiednich oknach (np. zmiennych zależnych czy czynników) po otwarciu okna dialogowego analizy, która obsługuje role. Moduły BASE, REGRESSION i ADVANCED ról nie obsługują. Do dyspozycji mamy następujące role: WEJŚCIE (dla zmiennej niezależnej lub predyktora), WYJŚCIE (dla zmiennej zależnej czy wyjaśnianej), ŁĄCZNIE (gdy zmienna może odgrywać obie role), BRAK (zmienna nie ma określonej roli), PODZIAŁ (zmienne definiujące podział na podzbiory), SEPARACJA (zmienne definiujące wybrane obserwacje). Domyślnie wszystkie zmienne mają zdefiniowaną rolę WEJŚCIE.

W zakładce DANE wpisujemy surowe dane, najlepiej zakodowane w postaci liczb. Choć nie zawsze widzimy sens takiej operacji, lepiej wpisać wszystkie możliwe dane zebrane od osoby badanej, nawet jeśli na początku procesu analizy wydaje nam się, że część danych nie będzie potrzebna dla weryfikacji postawionych przez nas hipotez. Dane surowe powinny także zawierać odpowiedzi na każde pytanie ankiety, którą stosujemy, a nie tylko zbiorczy wynik. Trzeba także wspomnieć, że jeśli w zakładce ZMIENNE wprowadzimy nazwy zmiennych, to zostaną one automatycznie umieszczone w szarym nagłówku każdej kolumny w zakładce DANE, zamiast napisu VAR. Pomiędzy poszczególnymi komórkami okna DANE można się poruszać myszką, ale znacznie wygodniej jest używać do tego celu klawiszy strzałek, znajdujących się z prawej strony klawiatury.

Do wpisywania wartości liczbowych danych polecamy stosowanie klawiatury numerycznej, którą przy odpowiednim ułożeniu ręki można obsługiwać w wygodny sposób bezwzrokowo, jednocześnie przesuwając się między polami dzięki naciskaniu kciukiem strzałki. Ten specyficzny układ dłoni nie ma znaczenia, gdy wpisujemy niewielką liczbę ankiet, ale zaczyna być istotny przy dużej ilości danych do wprowadzenia.

WPROWADZANIE DANYCH – PRZYKŁAD

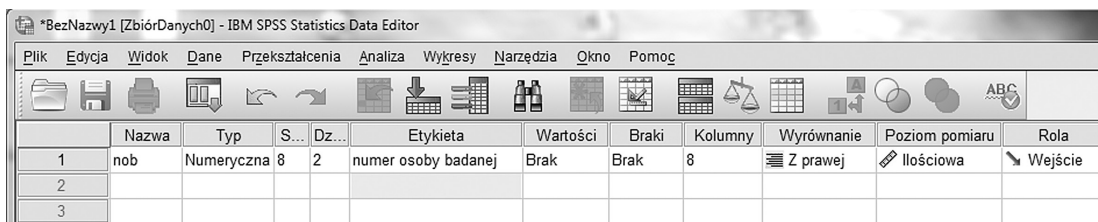
Aby omówić najważniejsze problemy, na które może natrafić czytelnik zmagający się z wprowadzaniem danych, zaprezentujemy przykład kilku wypełnionych ankiet, które będziemy wprowadzać do Edytora Danych. Ankiety pochodzą z badania zadowolenia z pracy i innych aspektów działalności zawodowej. Wypełniło je 6 osób. Pierwszą połowę osób przebadła Kasia, drugą połowę Michał. Kasia znalazła osoby badane na Pomorzu Zachodnim natomiast Michał na Pomorzu Wschodnim.

Oto ankiety:

Firma: państwowa/ prywatna		(1)
Liczba przepracowanych godzin (w tygodniu).....	39	
Ulubiona potrawa na stołówce.....	mielony	
Długość przerwy na lunch (godziny:minuty).....	1, 00	
Zadowolenie z pracy		
1- zdecydowanie nie, 2 – raczej nie, 3 – ani tak, ani nie, 4- raczej tak, 5- zdecydowanie tak		
Firma: państwowa/ prywatna		(2)
Liczba przepracowanych godzin (w tygodniu).....	27	
Ulubiona potrawa na stołówce.....	schabowy	
Długość przerwy na lunch (godziny:minuty).....	1, 00	
Zadowolenie z pracy		
1- zdecydowanie nie, 2 – raczej nie, 3 – ani tak, ani nie, 4- raczej tak, 5- zdecydowanie tak		
Firma: państwowa/ prywatna		(3)
Liczba przepracowanych godzin (w tygodniu).....	40	
Ulubiona potrawa na stołówce.....	ryba koperkowa	
Długość przerwy na lunch (godziny:minuty).....	0,30	
Zadowolenie z pracy		
1- zdecydowanie nie, 2 – raczej nie, 3 – ani tak, ani nie, 4- raczej tak, 5- zdecydowanie tak		
Firma: państwowa/ prywatna		(4)
Liczba przepracowanych godzin (w tygodniu).....	42	
Ulubiona potrawa na stołówce.....	mielony	
Długość przerwy na lunch (godziny:minuty).....	45 min	
Zadowolenie z pracy		
1- zdecydowanie nie, 2 – raczej nie, 3 – ani tak, ani nie, 4- raczej tak, 5- zdecydowanie tak		
Firma: państwowa/ prywatna		(5)
Liczba przepracowanych godzin (w tygodniu).....	43	
Ulubiona potrawa na stołówce.....	gotycki	
Długość przerwy na lunch (godziny:minuty).....	1, 10	
Zadowolenie z pracy		
1- zdecydowanie nie, 2 – raczej nie, 3 – ani tak, ani nie, 4- raczej tak, 5- zdecydowanie tak		
Firma: państwowa/ prywatna		(6)
Liczba przepracowanych godzin (w tygodniu).....	40	
Ulubiona potrawa na stołówce.....	kluski leniwe	
Długość przerwy na lunch (godziny:minuty).....	1, 00	
Zadowolenie z pracy		
1- zdecydowanie nie, 2 – raczej nie, 3 – ani tak, ani nie, 4- raczej tak, 5- zdecydowanie tak		

Rysunek 2.8. Ankiety wypełnione przez osoby badane

Wprowadzając dane ankietowe do programu IBM SPSS Statistics, należy pamiętać o istotnych regułach. Pierwsza zasada jest taka, że najpierw należy opisać zmienne, ich etykiety i wartości w zakładce ZMIENNE, a dopiero potem wypełnić okno DANE odpowiednimi wartościami. Zdefiniowanie nazw zmiennych ułatwia orientację w pliku z danymi i znacznie przyspiesza ich wprowadzanie. Z tego też powodu najlepiej, jeśli zmienne są wprowadzone dokładnie w takiej kolejności jak w kwestionariuszu. Druga zasada, wcale nie taka oczywista, dotyczy tego, że należy ponumerować ankietę. Dlaczego to takie ważne? Otóż nikt nie jest idealny i każdemu zdarzają się błędy. Numerowanie ankiet jest swego rodzaju ubezpieczeniem – jeśli popełnimy gdzieś błąd, zawsze możemy go skorygować, odszukując odpowiednią ankietę po zapisanym na jej rogu numerze. Jednak aby odszukać ankietę, musimy umieścić w danych zmienną, która będzie zawierała informację o numerze identyfikacyjnym osoby badanej. A zatem jako pierwszą zmienną wprowadzimy informację o numerze osoby badanej. Tak więc wchodzimy do zakładki ZMIENNE i w pierwszym wierszu kolumny NAZWA wpisujemy krótką nazwę tej zmiennej – proponujemy nazwę „nob”, a w kolumnie ETYKIETA wpisujemy długą nazwę „Numer osoby badanej” (rysunek 2.9).

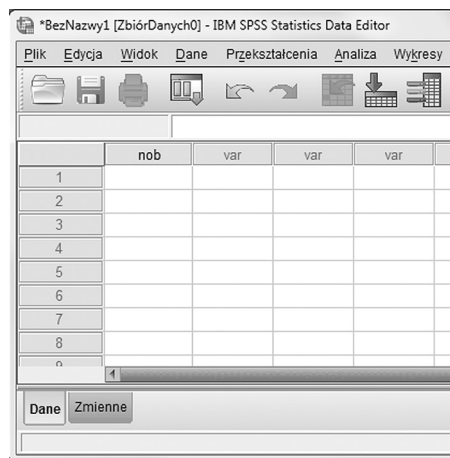


Rysunek 2.9. Definiowanie zmiennej „numer osoby badanej”

Po wpisaniu nazwy zmiennej w innych komórkach okna ZMIENNE pojawiają się informacje, których wcale nie wpisywaliśmy. IBM SPSS STATISTICS ma pewne domyślne ustawienia, które pojawiają się wtedy, gdy wpiszemy nazwę zmiennej. Oczywiście informacje te możemy w każdej chwili zmienić.

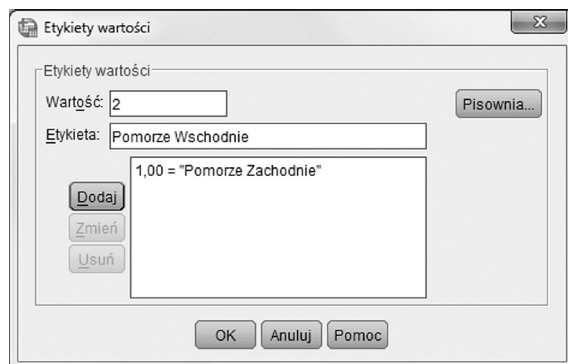
Zawsze numeruj ankietę i zapisuj zmienną „numer osoby badanej”. Numerowanie ankiet pozwala na szybkie odnalezienie ankiety w przypadku błędu przy wpisywaniu danych.

Zajrzyjmy teraz do zakładki DANE (rysunek 2.10). Czy coś się zmieniło? Owszem, pierwsza kolumna otrzymała nazwę „nob”. W tej kolumnie będziemy wpisywać kolejne numery osób badanych. Najpierw jednak zdefiniujemy wszystkie nazwy zmiennych, etykiety i wartości.



Rysunek 2.10. Wygląd okna Edytora Danych po nazwaniu pierwszej zmiennej

Wróćmy więc do zakładki ZMIENNE. Kolejną zmienną, którą chcemy zdefiniować, jest miejsce badania. W NAZWIE wpisujemy „miejsce”, a w etykiecie „Miejsce przeprowadzenia badania”. Zmienna ta przyjmuje dwie możliwe wartości, gdyż badanie było prowadzone w dwóch miejscach: na Pomorzu Zachodnim oraz na Pomorzu Wschodnim. Ponieważ IBM SPSS STATISTICS nie rozumie słów, dlatego poszczególnym miejscom musimy przyporządkować wartości liczbowe. Jakież? To już zależy tylko od naszej fantazji. Załóżmy, że cyfrą 1 będziemy oznaczać Pomorze Zachodnie, a cyfrą 2 Pomorze Wschodnie. Teraz jeszcze dla naszej wygody, wspomagając zawodną pamięć, warto wpisać ten sposób kodowania do odpowiedniej części kolumny WARTOŚCI. A więc umieszczamy kursor w pierwszej celce w kolumnie WARTOŚCI, a następnie klikamy przycisk z trzema kropkami. Otwiera się okno ETYKIETY WARTOŚCI, w którym wprowadzamy wartości liczbowe i odpowiadające im etykiety (zob. rysunek 2.11).



Rysunek 2.11. Definiowanie wartości liczbowych zmiennej oraz ich etykiet

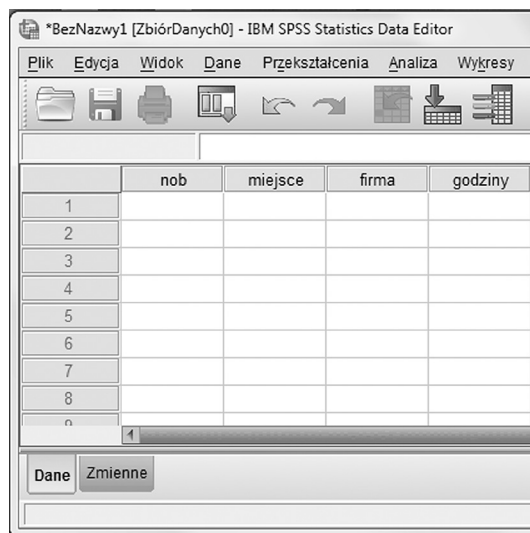
W polu WARTOŚĆ wpisujemy 1, a następnie w polu ETYKIETA umieszczamy słowa „Pomorze Zachodnie”. Następnie klikamy DODAJ. Analogicznie wpisujemy sposób kodowania Pomorza Wschodniego. Gdy już wpisujemy wszystkie wartości, klikamy OK. Pamiętajmy, żeby po wpisaniu etykiet kliknąć OK, jeśli zamkniemy okno krzyżykiem znajdującym się w prawym górnym rogu, to utracimy wpisywane w ostatnim kroku etykiety.

Pamiętajmy, że SPSS zapisuje zmiany w funkcjonalnych oknach (np. nadawanie etykiet wartościom zmiennej) tylko wówczas, gdy klikniemy przycisk OK. Jeśli zapomnimy go wcisnąć, wszelkie zmiany czy nowe dane wprowadzone przez nas zostaną utracone.

Kolejną informacją znajdującą się w ankiecie jest firma, w której pracuje osoba badana: państwowa lub prywatna. Zmienną tę możemy nazwać „firma”. W ETYKIECIE wpisujemy długie wyjaśnienie „firma, w której pracuje badany”. W WARTOŚCIACH zdefiniujemy, że 1 to „firma państwowa”, a 2 to „firma prywatna”.

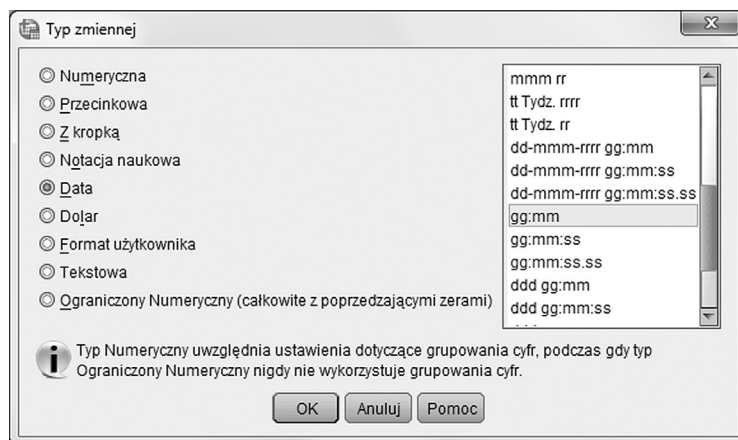
Kolejne pytanie, które zadano respondentom, dotyczyło liczby przepracowanych godzin w tygodniu. Utwórzmy więc zmienną o nazwie „godziny”, w ETYKIECIE której zapiszemy „liczba godzin przepracowanych w tygodniu”. Zapis jednostek w ETYKIECIE jest bardzo użyteczny. Wiele zmiennych możemy mierzyć w różnych jednostkach – przykładem może być waga, która bywa mierzona w kilogramach lub też tonach, w zależności od obiektu, który ważymy. I teraz ważna uwaga. Zmienne, które przyjmują wartości ciągłe, nie muszą być drobiazgowo opisane w kolumnie WARTOŚCI. Dlaczego? Jeśli w etykietach zmiennej mamy informację o jednostce pomiarowej, to wpisywanie informacji, że wartość 10 to „10 godzin” jest tylko niepotrzebnym mnożeniem pracy. A zatem program IBM SPSS STATISTICS daje jasną odpowiedź, ułatwiającą rozróżnienie zmiennych mierzonych na skalach jakościowych od skal ilościowych. **Zmienne jakościowe muszą zostać opisane za pomocą WARTOŚCI, ponieważ wartości liczbowe są przypisane poszczególnym odpowiedziom arbitralnie.** Zmienne ilościowe nie wymagają takiego opisu, gdyż ich wartości są zrozumiałe.

Wracamy ponownie do ankiet. Badanych pytaliśmy również o to, co najbardziej lubią z potraw serwowanych w firmowej stołówce. Nazwijmy tę zmienną „stołówka”, w ETYKIECIE zapiszemy „ulubiona potrawa w firmowej stołówce”. To pytanie było pytaniem otwartym i badani mogli wpisać dowolną odpowiedź. Takie pytania mają swoje zalety, ale także trochę wad. Zalety są takie, że nasza wyobraźnia może się okazać dość ograniczona i wymyślając potencjalne odpowiedzi, możemy nie uwzględnić ważnych z punktu widzenia badanego opcji. W pytaniu otwartym badani mają możliwość te odpowiedzi wyartykułować. Jednakże

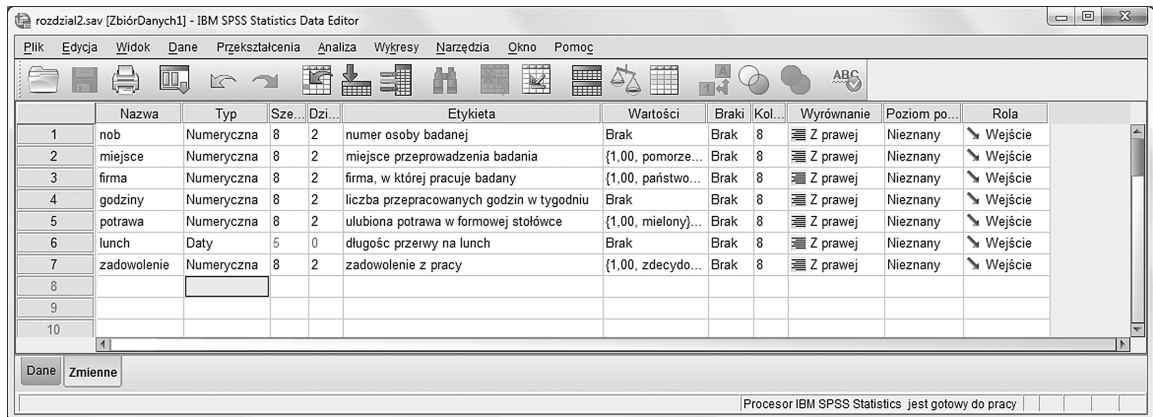


Rysunek 2.12. Wygląd Edytora Danych – widok DANE

ta swoboda bywa niekiedy kłopotliwa. Zwykle możliwych do wymyślenia opcji jest niezwykle dużo. Widzimy w odpowiedziach, że respondenci wpisali różne pokarmy. Na szczęście ankiet jest tylko kilka. Zwykle jednak nie sposób wpisać dosłownie odpowiedzi badanych, kodując każdą z nich w postaci cyfry, gdyż mielibyśmy zbyt dużą liczbę tych wartości. W takiej sytuacji konieczne staje się utworzenie kategorii, które będą grupowały większą grupę odpowiedzi. Sposób kategoryzacji odpowiedzi może jednak mieć wpływ na wyniki, więc trzeba go dokładnie przemyśleć, zanim wprowadzimy dane. Tutaj kodujemy każdą z odpowiedzi za pomocą innej cyfry, bo nie mamy zbyt wielu ankiet.



Rysunek 2.13. Definiowanie typu zmiennej



Rysunek 2.14. Edytor Danych – widok ZMIENNE ze zdefiniowanymi nazwami, etykietami i wartościami

Powróćmy do zakładki ZMIENNE. Kolejne pytanie dotyczyło długości przerwy na lunch (NAZWA „lunch”, ETYKIETA „długość przerwy na lunch (godziny:minuty)”). Kodowanie czasu nastęrcza trochę trudności – co ważne musimy w kolumnie TYP zaznaczyć opcję DATA, a następnie z listy wybrać zapis „GG:MM”, w którym przed dwukropkiem umieszczamy cyfry godzin, a po dwukropku cyfry minut (rysunek 2.13). Jak takie wartości wpisujemy? Po pierwsze, musimy pamiętać, żeby po wartości godzin wpisać dwukropek, bo inaczej program nie zrozumie danych. Po drugie, gdy wpisujemy liczbę 1,2, to IBM SPSS STATISTICS „zrozumie” to jako 1 godzinę i 2 minuty (a nie 1h i 20 minut). Aby program zorientował się, że chodzi nam o 1 godzinę i 20 minut, musimy wpisać na końcu tej liczby 0. Ponieważ format ten nie jest łatwy do obsługi, wydaje się, że lepiej ustalić jedną jednostkę i zapisywać czas, przeliczając wszystkie wyniki na minuty, lub też zaokrąglając do godzin.

Kolejne i ostatnie pytanie dotyczy zadowolenia z pracy, które jest oceniane na 5-stopniowej skali. Wpiszmy NAZWĘ zmiennej – „zadowolenie” oraz ETYKIETĘ „zadowolenie z pracy”. W WARTOŚCIACH kodujemy kolejno następujące pary wartości etykiet: 1 – zdecydowanie nie, 2 – raczej nie, 3 – ani tak, ani nie, 4 – raczej tak i 5 – zdecydowanie tak.

Po opisanu wszystkich zmiennych przechodzimy do okna DANE i wprowadzamy dane w postaci liczbowej do edytora. Jeśli wszystko poszło dobrze i nigdzie nie zrobiliśmy błędów we wpisywaniu danych, to dane powinny wyglądać tak jak na rysunku 2.15.

Na zakończenie naszej pracy zapisujemy plik. Zachęcamy do częstego zapisywania zmian dokonanych w pliku z danymi czy raportami, bo program IBM SPSS STATISTICS niestety sam nie archiwizuje pliku systematycznie, tak jak program Word. Warto przyzwyczać się do częstego używania ikonki z dyskieta na pasku

	nob	miejsce	firma	godziny	potrawa	lunch	zadowolenie	var	var	var
1	1,00	1,00	1,00	39,00	1,00	1:30	4,00			
2	2,00	1,00	1,00	27,00	2,00	1:00	2,00			
3	3,00	1,00	2,00	40,00	3,00	0:30	5,00			
4	4,00	2,00	2,00	42,00	1,00	0:45	3,00			
5	5,00	2,00	2,00	43,00	4,00	1:10	1,00			
6	6,00	2,00	1,00	40,00	5,00	1:00	5,00			
7										
8										
9										

Rysunek 2.15. Edytor Danych – widok DANE z wpisanymi wartościami liczbowymi

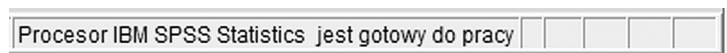
narzędzi czy skrótu CTRL+S, który umożliwia zapis aktualnej wersji pliku. Taki nawyk uchroni nas przed nieprzyjemnymi konsekwencjami nieoczekiwanych błędów w pracy programu powodujących jego zamknięcie. Niezapisywanie zmian powoduje bowiem w przypadku zamknięcia programu powrót do tej wersji pliku, którą zapisaliśmy, a więc pozbawia nas efektów ciężkiej pracy wykonanej już po zapisaniu zmian.

I jeszcze jedna uwaga techniczna. Podczas wpisywania danych może się zdarzyć, że zapędzimy się za daleko w naszych danych i na końcu zmiennych, które zdefiniowaliśmy, pojawi się zmienna o dziwnej nazwie na przykład VAR00001, lub też na końcu obserwacji pojawią się nowe, puste pola z przecinkami. Aby usunąć zbędne zmienne czy zbędne obserwacje, klikamy na szary pasek z numerem obserwacji lub nazwą zmiennej. Podświetla się wtedy cała kolumna, lub cały wiersz. Wtedy przyciskamy klawisz DEL i zbędne zmienne lub obserwacje zostają usunięte.

PASEK STANU

PASEK STANU

W Edytorze Danych oprócz dwóch zakładki DANE i ZMIENNE znajdują się także dwa istotne elementy. Pierwszy z nich jest umiejscowiony na dole z prawej strony okna Edytora Danych i nazywa się PASEK STANU (rysunek 2.16). Jest to szary pasek z sześcioma okienkami, w którym znajdują się istotne informacje o tym, co aktualnie się dzieje w programie.



Rysunek 2.16. Pasek stanu w dolnej części Edytora Danych

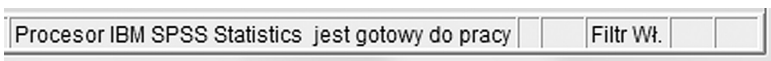
Pasek stanu jest bardzo użyteczny. Początkujący użytkownicy często zapominają wyłączyć filtr i otrzymują zaskakujące wyniki analiz. Dlatego, zanim w takiej sytuacji się zdenerwujemy i zaczniemy przeklinać SPSS i całą statystykę, warto najpierw sprawdzić pasek stanu.

Na rysunku widzimy na PASKU STANU w pierwszym okienku napis PROCESOR GOTOWY. Taka informacja oznacza, że IBM SPSS STATISTICS aktualnie nie wykonuje żadnych obliczeń i jest gotowy do pracy. Jeśli program jest w trakcie wykonywania analiz statystycznych, to na pasku stanu pojawi się napis WYKONUJE wraz z precyzyjną informacją o rodzaju aktualnie przetwarzanej operacji. Wtedy w oknie obok będzie można zobaczyć liczbę obserwacji, które już zostały poddane procesowi obliczeniowemu.

Komunikat o procesorze na Pasku stanu.

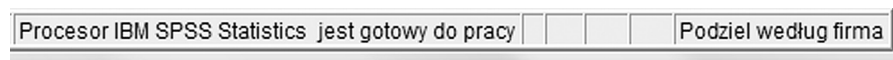
- ◆ Procesor gotowy' oznacza, że SPSS i nasz komputer nie jest obciążony żadnymi obliczeniami.
- ◆ Informacja 'Wykonuje...' oznacza, że SPSS aktualnie realizuje polecenie. Trzeba wtedy poczekać z kolejną analizą, aż do ponownego ukazania się komunikatu 'Procesor gotowy'.

Drugie okienko paska stanu STATUS OMS pokazuje informacje, czy został włączony system przekazywania wyników analiz do zewnętrznych plików, takich jak pliki tekstowe czy HTML. Trzecie okno paska stanu to OBSZAR LICZNIKA PRZETWARZANYCH OBSERWACJI, które pokazuje, ile obserwacji zostało już przeliczonych w danym momencie. Zwykle dzieje się to tak szybko, że informacja ta nie jest wyświetlana. Kolejne, czwarte okno – OBSZAR STATUSU FILTROWANIA/WYKORZYSTANIA OBSERWACJI – informuje, czy analiza, którą wykonamy, zostanie przeprowadzona na wszystkich osobach badanych, czy też wyselekcjonowana została pewna ich część. Niekiedy chcemy przeprowadzić analizy na wybranej grupie badanych i właśnie wtedy, po zdefiniowaniu, na jakiej podgrupie badanych będą wykonywane dalsze obliczenia, w czwartym okienku paska stanu znajdzie się napis FILTR WŁ. (rysunek 2.17).



Rysunek 2.17. Pasek stanu pokazujący włączenie filtrowania obserwacji

Kolejne okno może zwierać komunikat WAGA WŁ., gdy obserwacje są ważone według wybranej zmiennej (zob. rozdział 7). W ostatnim zaś umieszczana jest informacja PODZIEL WEDŁUG, gdy wykonujemy analizy na podgrupach wyodrębnionych z całego zbioru danych (rysunek 2.18).



Rysunek 2.18. Pasek stanu pokazujący włączenie podziału na podzbiory

PASEK NARZĘDZI

PASEK NARZĘDZI

Omówimy teraz poszczególne opcje paska narzędzi znajdującego się pod górnym menu. Umieszczone na nim przyciski umożliwiają szybki dostęp do podstawowych funkcji z górnego menu. Część z tych ikonek jest podobna do używanych w innych programach, inne są charakterystyczne dla programu IBM SPSS STATISTICS. Na rysunku 2.19 prezentujemy wszystkie ikonki. Ich opis w kolejności zgodnej z kolejnością na pasku narzędzi znajduje się w tekście poniżej.



Rysunek 2.19. Pasek narzędzi w Edytorze Danych

1. OTWÓRZ PLIK. Funkcja ta pozwala otworzyć dowolny typ pliku IBM SPSS STATISTICS, a także pliki programów, z których możliwy jest import danych. Warto zwrócić uwagę na nazwy programów, które znajdują się w rozwijanej liście w okienku PLIKI TYPU. Możemy tam zobaczyć, że program IBM SPSS STATISTICS pozwala na import danych z Excela, danych tekstowych oraz innych programów statystycznych, takich jak Statistica lub SAS-a.
2. ZAPISZ PLIK. Gdy zapisujemy po raz pierwszy plik, ikonka ta działa jako ZAPISZ JAKO, a więc umożliwia zapisanie nowego pliku pod określoną nazwą w wybranej lokalizacji. Gdy plik ma już nadaną nazwę, funkcja ta spowoduje zapisanie ostatnich zmian.

Zapisanie pliku niesie ze sobą kilka konsekwencji:

- ◆ Chroni nas przed utratą ostatnich zmian, np. gdy zawiesi się nam SPSS lub system operacyjny.
- ◆ Uniemożliwia powrót do poprzednich wersji pliku (plik stary zastępowany jest nowym) i przycisk COFNIJ staje się nieaktywny.

3. DRUKUJ. Funkcja ta umożliwia nam wydrukowanie zawartości aktualnie otwartego okna. Możemy zatem na wydruku uzyskać surowe dane w formie tabeli, jeśli jesteśmy w zakładce DANE lub tabelę z opisem danych, jeśli aktywna jest zakładka ZMIENNE. Jeśli otwarty jest Edytor Raportów, ta funkcja pozwoli wydrukować wykonane obliczenia i wykresy.
4. PRZYWOŁAJ OKNO. Ta ikona pozwala na otwarcie ostatnio używanych funkcji i analiz. Docenimy ten przycisk, gdy okaże się, że musimy wykonać ponownie analizę, do której dostęp wymaga wielu kliknięć.
5. POWTÓRZ I COFNIJ. Tak jak w każdym edytorze tekstu również w programie IBM SPSS STATISTICS możliwe jest cofnięcie zmiany lub powrót do niej.
6. PRZEJDŹ DO OBSERWACJI. Gdy klikniemy ten przycisk, pojawi się okienko, w którym możemy wpisać numer poszukiwanej obserwacji. IBM SPSS STATISTICS przeniesie nas bezpośrednio do niej, tzn. umieści wiersz danych, w którym znajduje się ta obserwacja w pierwszym wierszu edytora danych widocznym na ekranie. Funkcja ta jest szczególnie przydatna, gdy mamy bardzo liczną próbę. Uwaga! Warto pamiętać, że numer, który mamy wpisać, jest numerem nadawanym przez IBM SPSS STATISTICS a nie dodaną przez nas zmienną „numer osoby badanej”.
7. PRZEJDŹ DO ZMIENNEJ. Ta funkcja umożliwia łatwe znalezienie odpowiedniej zmiennej – przydatna wtedy, gdy dane zawierają dużą liczbę zmiennych.
8. INFORMACJA O ZMIENNYCH. Kliknięcie tej ikonki spowoduje wyświetlenie opisu zmiennych, znajdujących się w pliku danych.
9. ZNAJDŹ. Funkcja ta pozwala na znalezienie szukanej wartości/litery/słowa w obrębie jednej zaznaczonej zmiennej. Najpierw musimy zaznaczyć zmienną klikając na jej nazwę (jeden klik), by została zaczerniona, a szukana wartość zostanie pokazana na białym tle.
10. WSTAW OBSERWACJĘ. Dzięki tej opcji możemy wstawić nowy wiersz, czyli obserwację (osobę badaną). Jeśli zaznaczymy wiersz nr 9, to IBM SPSS STATISTICS wstawi tam nowy pusty wiersz, przesuwając poniższe obserwacje o jeden w dół. Przycisk jest nieaktywny, gdy znajdujemy się w zakładce ZMIENNE.
11. WSTAW ZMIENNĄ. Analogicznie do funkcji WSTAW OBSERWACJĘ przycisk ten dodaje nową zmienną. Aby wybrać miejsce pojawienia się nowej zmiennej, trzeba zaznaczyć odpowiednią kolumnę.
12. PODZIEL DANE NA PODZBIORY. Ikona ta otwiera okno umożliwiające podzielenie całej grupy osób badanych na mniejsze podgrupy na podstawie określonej zmiennej. Po wyborze tej funkcji na pasku stanu pojawi się informacja „Podziel według” i nazwa zmiennej określającej podział.
13. WAŻENIE OBSERWACJI. Funkcja umożliwia ważenie obserwacji (zob. rozdział 7).

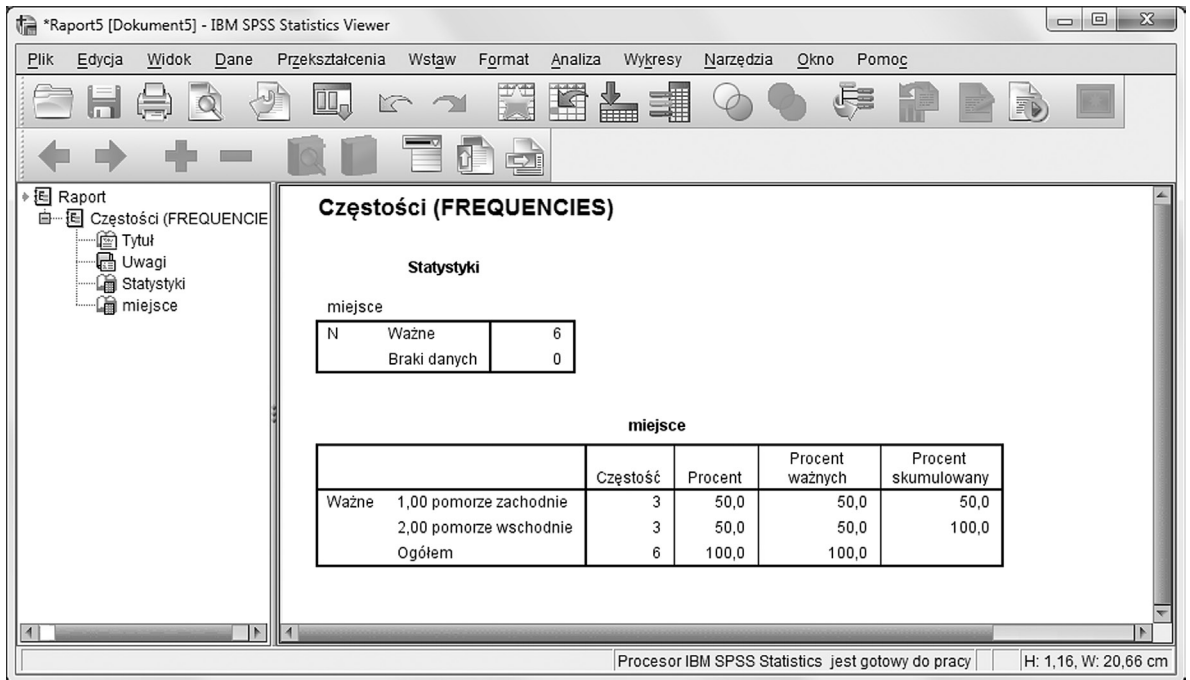
14. **WYBIERZ OBSERWACJE.** Ikona umożliwia szybki dostęp do okna dialogowego pozwalającego wybrać te obserwacje, które spełniają określone warunki, zdefiniowane na bazie zawartych w danych zmiennych.
15. **EYKIETY WARTOŚCI.** Kliknięcie na tę ikonę pozwala na wyświetlenie w Edytorze Danych w widoku Dane etykiet tekstowych zamiast wartości liczbowych.
16. **UŻYJ ZESTAWÓW.** Jeśli dane mają dużą liczbę zmiennych, można zdefiniować pewne ich części, nadając im określoną nazwę i tworząc zestaw. Dzięki temu można wygodniej orientować się w dużych plikach danych. Opcja UŻYJ ZESTAWÓW umożliwia przełączanie się między zdefiniowanymi wcześniej zestawami.
17. **POKAŻ WSZYSTKIE ZMIENNE.** Prezentuje listę wszystkich zmiennych
18. **SPRAWDZANIE PISOWNI.**

EDYTOR RAPORTÓW

Jak już wspomniano, Edytor Raportów przechowuje wyniki analiz wykonanych przez IBM SPSS STATISTICS w postaci tabel lub wykresów. A zatem wyniki wszelkich operacji obliczeniowych po ich potwierdzeniu przyciskiem OK zostaną umieszczone w specjalnym pliku, który można zapisać, a potem ponownie otworzyć. Okno Edytora Raportów składa się z dwóch elementów: spisu treści z lewej strony oraz okna z poszczególnymi tabelami i wykresami. Kliknięcie na pozycje spisu treści pozwala na szybkie odnalezienie odpowiedniej tabeli i jest bardzo wygodne przy dużej liczbie analiz.

Warto wspomnieć, że wszystkie tabele i wykresy można skopiować do programu Word za pomocą opcji KOPIUJ, lub KOPIUJ SPECJALNIE. Opcja KOPIUJ umożliwia takie skopiowanie tabel z programu, aby podlegały one normalnej edycji pozwalającej na zmianę czcionki i innych elementów tabeli. KOPIUJ SPECJALNIE powoduje pojawienie się okna w którym możemy wybrać format kopiowania tabeli. Wybór kopiowania jako ZWYKŁY TEKST, TEKST SFORMATOWANY lub ARKUSZ ROBOCZY EXCEL pozwala na późniejsze wklejenie tabeli w postaci edytowalnej, wybór formatu OBRAZ powoduje, że po wklejeniu tabela nie będzie mogła być edytowana. Modyfikacji tabel możemy dokonywać w Edytorze danych, gdy klikniemy dwukrotnie na tabelę, którą chcemy modyfikować.

Dwukrotne kliknięcie na dowolny obszar tabeli pozwala nam dowolnie go modyfikować – wpisać zupełnie nowe treści lub edytować istniejące. Dzięki temu możemy zobaczyć wartość statystyk znajdujących się w tabeli oraz zaokrąglenie tych wartości do dowolnej liczby miejsc po przecinku. Trzeba bowiem wspomnieć, że liczby w tabelach nie są zaokrąglane, a jedynie wyświetlane maksymalnie do trzeciego miejsca po przecinku. Taki sposób wyświetlania wartości liczbowych powoduje, że liczba 0,99199 w tabeli pojawia się jako 0,991 co jest niezgodne z zasadą zaokrąglania. Dlatego tę liczbę powinniśmy skorygować do 0,992.



Rysunek 2.20. Okno Edytora Raportów – spis treści oraz okno z tabelami

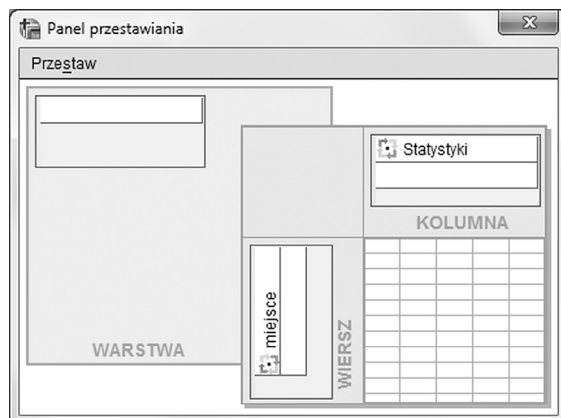
Bardzo ciekawym narzędziem w Edytorze Raportów jest **PANEL PRZESTAWIANIA**, dający możliwość swobodnej edycji tabeli wydruku, na przykład zamiany wierszy i kolumn czy ukrycia pewnej części wydruku. Panel ten prezentuje wygodne narzędzie graficzne, które ułatwia orientację, jakie informacje są umieszczone w wierszu, jakie w kolumnie, a jakie w warstwach (rysunek 2.22).

PANEL PRZEDSTAWIANIA

Każda informacja jest zaznaczona kolorowym kwadratem z opisem, który element tabeli zostanie zmodyfikowany po przeniesieniu określonego kwadratu. Każdy kwadracik można chwycić myszką i przenieść w inne miejsce. W ten sposób można zamienić wiersze, kolumny oraz umieścić określone elementy tabeli w **WARSTWACH** (czyli ukryć pewne informacje pod innymi). Warto niekiedy poeksperymentować, by uzyskać możliwie najbardziej czytelny wydruk.

miejsce		Częstość	Procent	Procent ważnych	Procent skumulowany
Ważne	1,00 pomorze zachodnie	3	50,0	50,0	50,0
	2,00 pomorze wschodnie	3	50,0	50,0	100,0
	Ogółem	6	100,0	100,0	

Rysunek 2.21. Aktywne okno edycji tabel w Edytorze Raportów



Rysunek 2.22. Panel przestawiania w Edytorze Raportów

Na rysunku 2.23 przedstawiamy wygląd tabeli częstości po przestawieniu statystyk na warstwy. Wtedy spośród wszystkich dostępnych statystyk (częstości, procentów) prezentowane są tylko jedne, wybrane przez użytkownika (tutaj częstości).

Statystyki		
miejsce		
N	Ważne	6
	Braki danych	0

miejsce		
Statystyki	Częstość	
Ważne	1,00 pomorze zachodnie	3
	2,00 pomorze wschodnie	3
	Ogółem	6

Rysunek 2.23. Wygląd tabeli częstości po przestawieniu statystyk na warstwy. Obok wygląd panelu przestawiania

Edytor Raportów posiada także pasek narzędzi, ale większość funkcji dostępnych na tym pasku jest zbieżna z tymi opisywanymi w Edytorze Danych (zob. rysunek 2.24). Nie będziemy ich zatem opisywać ponownie, pozostawiając czytelnikowi możliwość samodzielnego odkrycia ich możliwości.



Rysunek 2.24. Pasek narzędzi Edytora Raportów

EDYTOR POLECEŃ

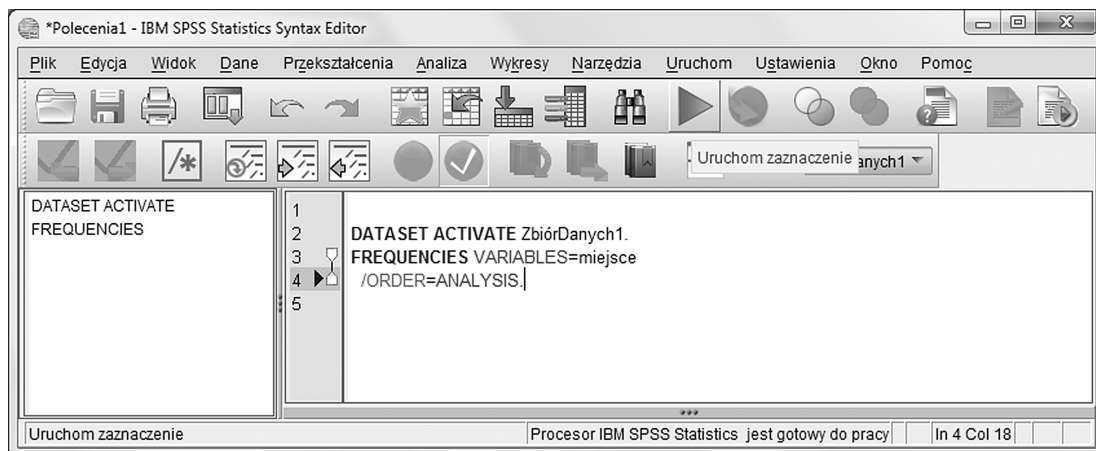
Każda operacja obliczeniowa w programie IBM SPSS STATISTICS może zostać wykonana różnymi sposobami. Jednym z nich, najprostszym, jest wybór odpowiednich opcji z górnego menu poprzez kolejne okna dialogowe. Jednak wtedy, gdy powtarzamy wielokrotnie ten sam typ analiz, zmieniając jedynie zestaw uwzględnianych zmiennych, wchodzenie za każdym razem do okienka dialogowego może być nie tylko pracochłonne, ale i czasochłonne. A można to zrobić inaczej. Narzędziem umożliwiającym niezwykle ekonomiczną pracę z programem IBM SPSS STATISTICS jest EDYTOR POLECEŃ. Aby przejść do Edytora Poleceń, wystarczy, gdy uruchamiamy analizę z górnego menu, zamiast potwierdzenia wykonania analizy przyciskiem OK, należy wybrać przycisk WKLEJ.

EDYTOR POLECEŃ

Otworzy się wtedy automatycznie okienko Edytora Poleceń z zapisanym w języku angielskim poleceniem (rysunek 2.25). Polecenia mają specyficzną składnię – zaczynają się zawsze nazwą analizy, która będzie wykonywana. Dość łatwo można w języku poleceń znaleźć nazwy zmiennych. Jeśli zmienimy zestaw analizowanych zmiennych, wpisując z klawiatury nazwy nowych zmiennych, to trzeba wydać programowi komendę – wykonaj to polecenie. Można to zrobić za pomocą ikony URUCHOM ZAZNACZENIE. Trzeba tylko pamiętać, aby zaznaczyć kursorem to polecenie, które chcemy wykonać. Można też wykorzystać skrót klawiaturowy CTRL+R.

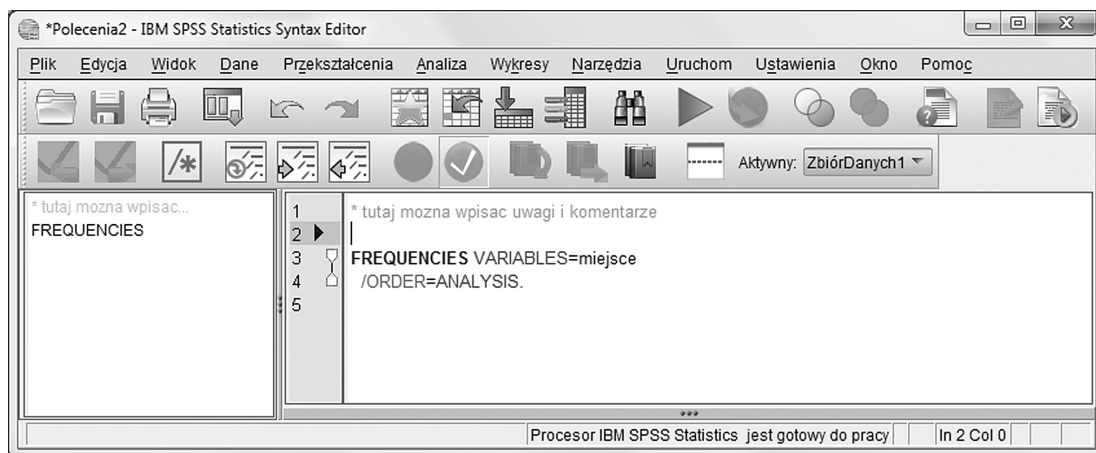
Efekty działania Edytora Poleceń zobaczymy w Edytorze Raportów, jeśli wykonywaliśmy analizy statystyczne, bądź też w Edytorze Danych, jeśli dokonywaliśmy przekształceń na zmiennych.

Jeśli znamy polecenia na pamięć, to możemy także otworzyć Edytor Poleceń wybierając PLIK w górnym menu, a następnie NOWY > POLECENIE. Wpisujemy wtedy odpowiednie komendy i wykonujemy analizy statystyczne, nie wchodząc w ogóle do górnego menu. Edytor Poleceń, jak każdy inny plik, może zostać zapisany i wykorzystany ponownie, jeśli wykonujemy serię analiz tego samego typu na kolejno zbieranych danych. Pozwala więc na znaczną automatyzację pracy i jej przyspieszenie. Pliki z zapisanymi poleceniami do pewnych standardowych procedur badawczych można znaleźć w Internecie. Przykładem są przygotowane przez A. Greenwalda



Rysunek 2.25. Zapis polecenia w oknie Edytora Poleceń

i współpracowników pliki poleceń do tworzenia wskaźników ukrytych postaw mierzonych IAT (*Implicit Association Test*). Dodatkowo język poleceń pozwala na wykonanie takich operacji obliczeniowych, które nie są dostępne z górnego menu. Tak więc język poleceń jest naprawdę językiem dla wtajemniczonych, choć dość łatwo jest przystąpić do ich grona, szukając odpowiednich informacji w Internecie (szukamy słowa SYNTAX, które po angielsku oznacza język poleceń) bądź w części POMOCY nazwanej MATERIAŁY O SKŁADNI POLECENI. Niestety podręcznik ten jest w języku angielskim, ale jego zawartość znacznie wykracza standardowe treści pomocy. W gruncie rzeczy to naprawdę rozbudowany podręcznik z przykładami zastosowań składni poleceń, a nie tylko ich prosty spis. Warto też pamiętać



Rysunek 2.26. Plik języka poleceń z wpisanym w nagłówku komentarzem

o tym, że w pliku tym możemy zapisać dodatkowe informacje o wykonywanych analizach i operacjach, swego rodzaju didaskalia. Aby program IBM SPSS Statistics nie próbował rozpoznać tych komentarzy jako swoje komendy jako pierwszy znak wpisujemy gwiazdkę, a następny wiersz zostawiamy pusty (rysunek 2.26).

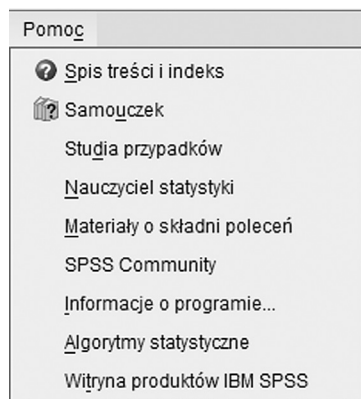
Bardziej zaawansowani użytkownicy programu IBM SPSS Statistics docenią także z pewnością możliwości automatyzacji pracy za pomocą języków programowania: Python, czy makra, które pozwalają uzyskać dodatkowe statystyki niedostępne w zwykłym programie. Przydatne makra zmieścił na przykład Andrew F. Hayes na swojej stronie www.afhayes.com. Na stronie można też znaleźć sposób wczytania makra do programu IBM SPSS Statistics.

POMOC

Skoro już jesteśmy przy temacie pomocy, warto powiedzieć kilka słów na jej temat. IBM SPSS STATISTICS ma dość rozbudowaną pomoc, w nowszych wersjach programu napisaną przeważnie w języku polskim, co ułatwia korzystanie z niej. Składniki POMOCY zajmują osobną pozycję górnego menu. Główne z nich to: SPIS TREŚCI I INDEKS, SAMOUCZEK, NAUCZYCIEL STATYSTYKI oraz STUDIA PRZYPADKÓW.

Jeśli wybierzemy SPIS TREŚCI I INDEKS, otworzy się duże okno ze spisem zagadnień z lewej strony, a ich opisem w prawej części okna. Okno spisu treści jest podzielone na zakładki. Zakładka SPIS TREŚCI daje dostęp do tematycznego układu treści pomocy, INDEKS umożliwia wyszukiwanie haseł z pomocą słowa kluczowego. Podstawowe informacje są w języku polskim, opis bardziej zaawansowanych funkcji w języku angielskim. Użytkownicy ciekawi, jak są obliczane poszczególne statystyki, mogą znaleźć te informacje w części pomocy ALGORYTMY STATYSTYCZNE. Algorytmy też są jednak prezentowane w języku angielskim.

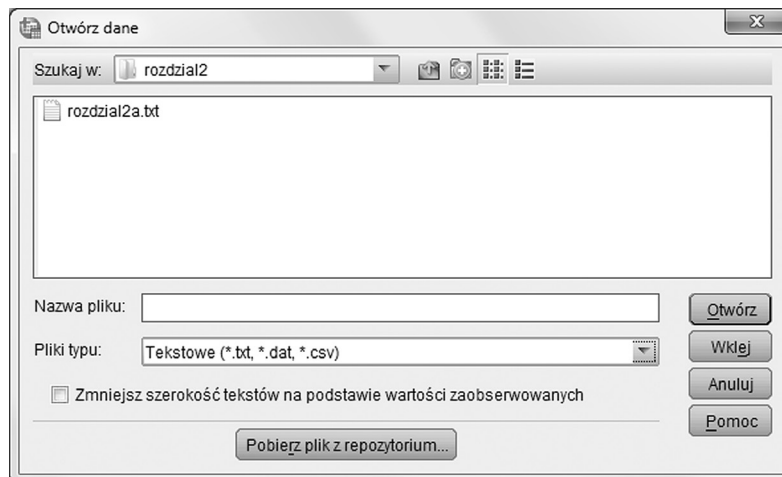
POMOC



Rysunek 2.27. Główne składowe Pomocy

IMPORT DANYCH **IMPORT DANYCH Z INNYCH PROGRAMÓW**

Dane można wpisywać w różnych plikach: tekstowych, html, Excela czy baz danych. Dane mogą bowiem pochodzić z różnych źródeł – gdy są zbierane przez Internet, mogą mieć format bazy danych albo pliku Excela, gdy zbieramy je za pomocą programów pozwalających projektować bardziej złożone procedury badawcze, mamy często do czynienia z plikami tekstowymi. Takie dane można spokojnie zaimportować do programu IBM SPSS Statistics za pomocą okna dialogowego PLIK – OTWÓRZ – DANE. Wybieramy plik, który chcemy zaimportować, a następnie z rozwijanej listy wybieramy typ pliku. Wczytajmy plik tekstowy *rozdzial2a.txt*, w którym znajdują się wyniki dwóch dodatkowych osób, które wypełniły ankietę dotyczącą zadowolenia z pracy. Wchodzimy więc do PLIK – OTWÓRZ – DANE. Wybieramy plik *rozdzial2a.txt* i z rozwijanej listy PLIKI TYPU określającej typ wybieramy plik tekstowy (rysunek 2.28).

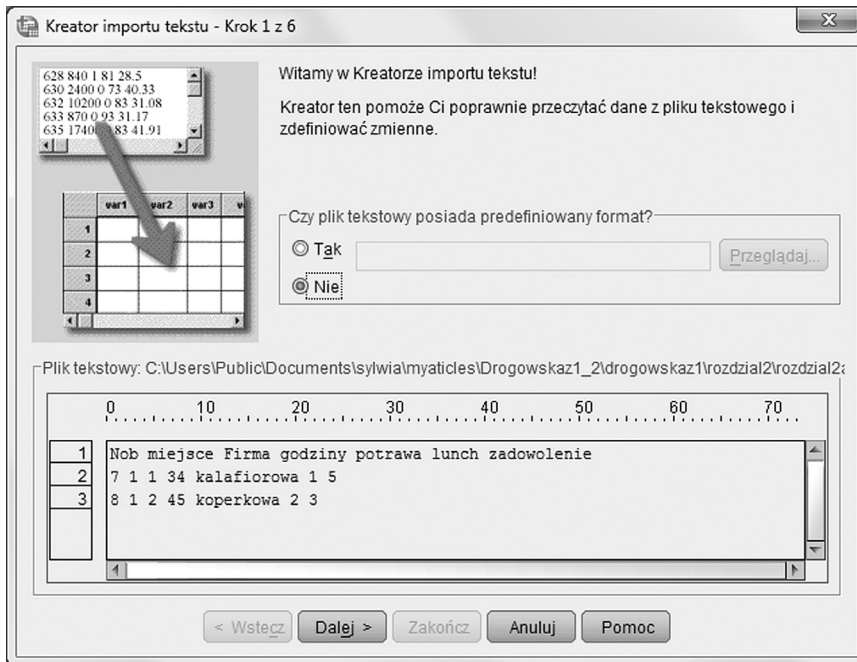


Rysunek 2.28. Wybór pliku do importu

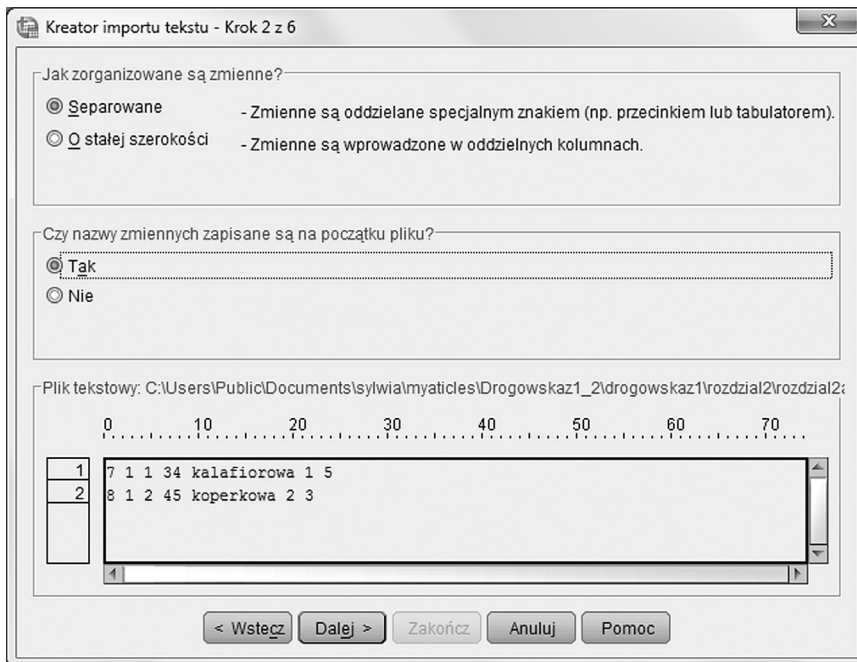
Następnie, gdy klikniemy OK, pojawi się okno KREATORA IMPORTU TEKSTU, który pomoże nam poprawnie zaimportować dane. W pierwszym oknie zapyta nas, czy dane mają predefiniowany format oraz zaprezentuje fragment danych, które odczytał z pliku tekstowego (rysunek 2.29).

Jeśli dane nie mają predefiniowanego przez nas wcześniej formatu, wybieramy opcję NIE i klikamy DALEJ. W następnym kroku musimy określić, czy zmienne są separowane, czyli oddzielane jakimś znakiem albo spacją, czy też są wprowadzone w kolumnach o stałej szerokości (rysunek 2.30).

W pliku, który chcemy zaimportować, zmienne są separowane, więc ta opcja powinna pozostać zaznaczona. Teraz druga kwestia: czy nazwy zmiennych są

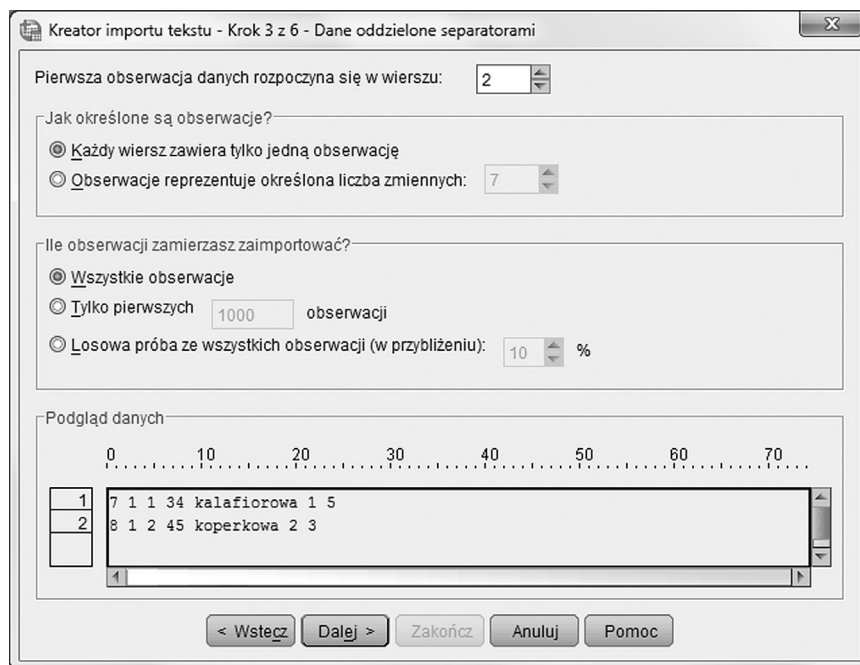


Rysunek 2.29. Określanie formatu importowanego pliku



Rysunek 2.30. Definiowanie położenia nazw zmiennych oraz sposobu oddzielania zmiennych

zapisane na początku pliku? W naszym zbiorze nazwy zmiennych są zapisane na początku, więc zaznaczamy opcję TAK i klikamy przycisk DALEJ. Wykonaliśmy już dwa kroki importu, teraz pora na trzeci. Kreator importu pyta nas, jak określona jest każda obserwacja – czy wyniki jednej osoby mieszczą się w jednym wierszu, czy w wielu (rysunek 2.31).

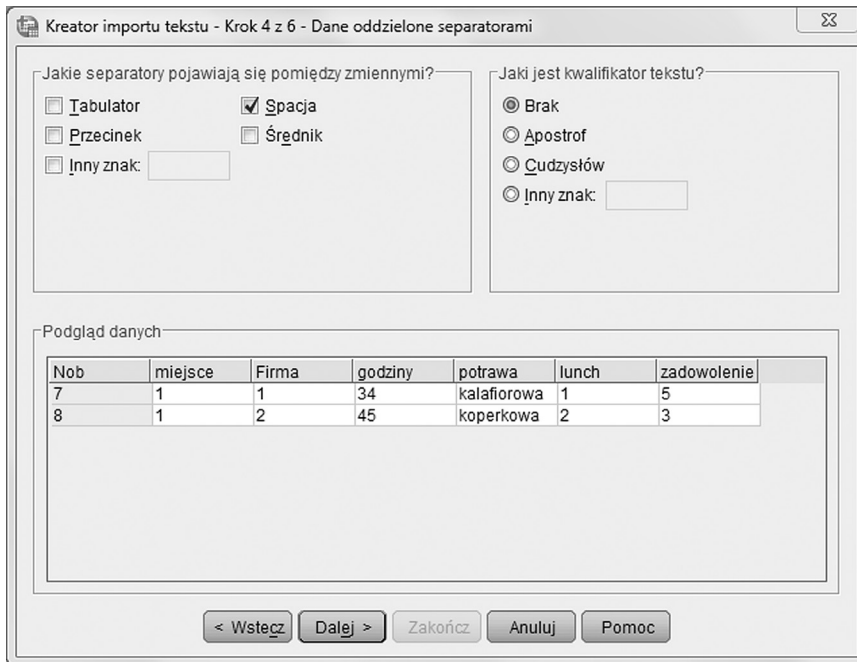


Rysunek 2.31. Określenie liczby importowanych obserwacji i wierszy zajmowanych przez jedną obserwację

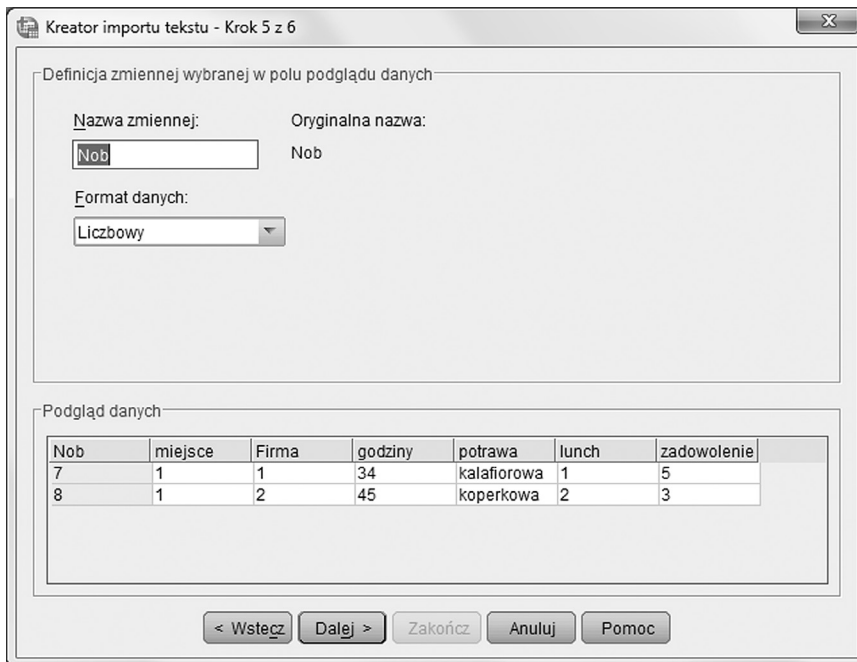
Wybieramy opcję KAŻDY WIERSZ ZAWIERA JEDNĄ OBSERWACJĘ. Kreator chce także wiedzieć, czy ma zaimportować wszystkie obserwacje, czy jakąś ich podgrupę. Chcemy zaimportować wszystkie obserwacje, więc zostawiamy domyślnie ustawioną opcję i klikamy dalej. W czwartym kroku musimy określić, co stanowi separator poszczególnych zmiennych. W naszym pliku jest to spacja, więc wybieramy tę opcję i klikamy przycisk DALEJ (rysunek 2.32).

W kolejnym kroku program przeprowadza nas przez proces definicji formatu niektórych zmiennych – zwykle tych, które znajdują się na początku zbioru danych i identyfikują osoby badane (rysunek 2.33).

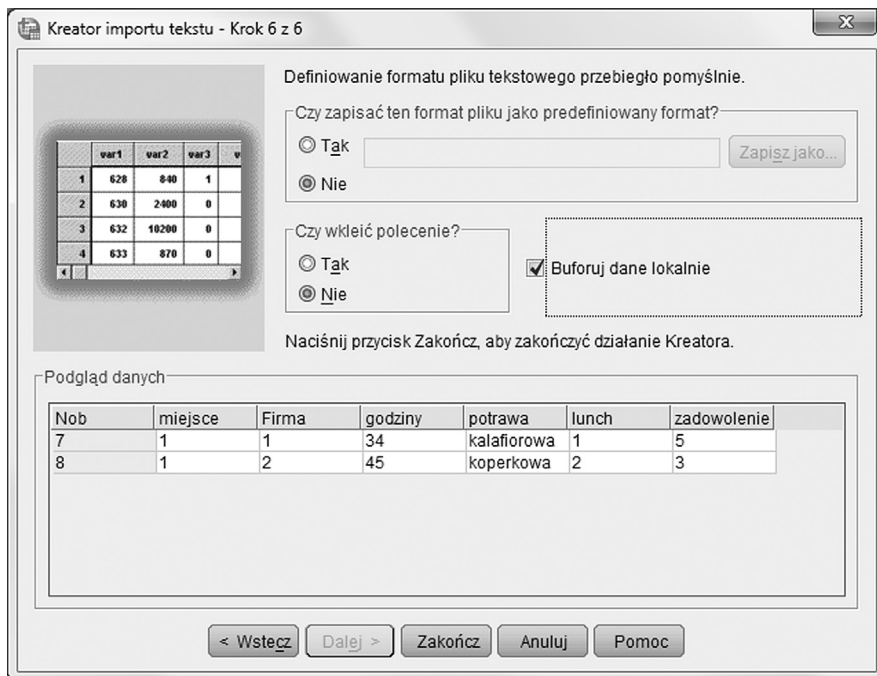
W ostatnim kroku mamy możliwość zapisania tego sposobu importu jako predefiniowany format lub wklejenie całej procedury do pliku Języka Poleceń (rysunek 2.34). Jeśli zamierzamy w ten sam sposób importować większą liczbę plików, warto wykorzystać którąś z tych opcji – ułatwi i przyspieszy to całą pracę.



Rysunek 2.32. Określenie separatora zmiennych



Rysunek 2.33. Definicja zmiennych



Rysunek 2.34. Możliwość zapisania sposobu importu w szablonie lub w pliku Języka Poleczeń

Teraz jeszcze musimy obejrzeć dane (rysunek 2.35) i zakodować zmienną *potrawa* odpowiednimi wartościami, zgodnie z ustalonym wcześniej sposobem kodowania. Zamiast „koperkowa” w zmiennej *potrawa* wstawiamy wartość 3, a zamiast „kalafiorowa” wstawiamy wartość 6. Zmieniamy też typ zmiennej z tekstowego na numeryczny. Wreszcie zapisujemy ten plik jako plik danych IBM SPSS Statistics pod nazwą *rozdzial2a*. Uwaga! Niekiedy w czasie importu błędnie jest rozpoznawany typ zmiennej – zamiast numerycznego program definiuje tekstowy, co uniemożliwia wykonanie analiz dla tak zaimportowanej zmiennej. Warto sprawdzić po imporcie, czy typ zmiennych jest prawidłowy.

	Nob	miejsce	Firma	godziny	potrawa	lunch	zadowolenie
1	7	1	1	34	kalafiorowa	1	5
2	8	1	2	45	koperkowa	2	3

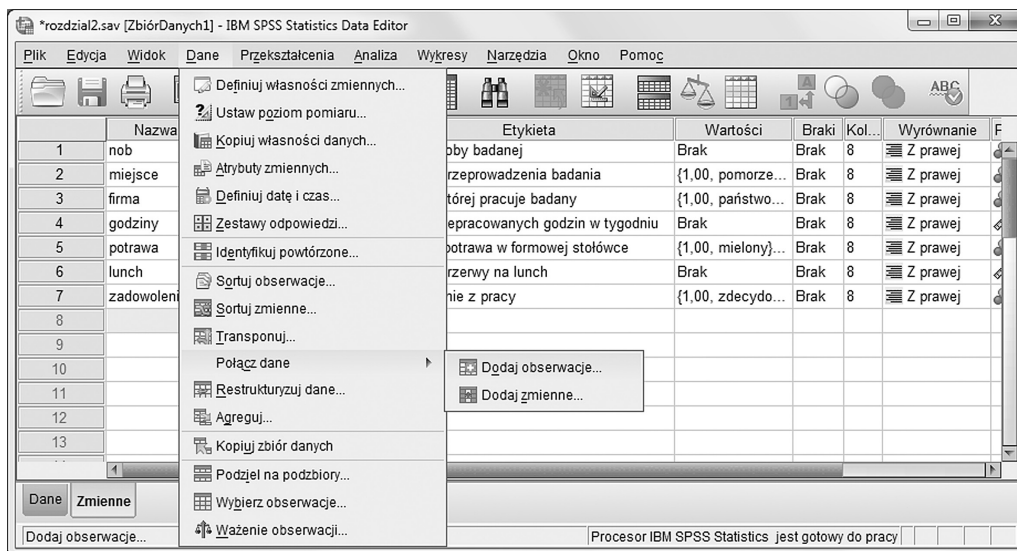
Rysunek 2.35. Wygląd zaimportowanych danych

ŁĄCZENIE DANYCH

DODAWANIE OBSERWACJI

DODAWANIE OBSERWACJI

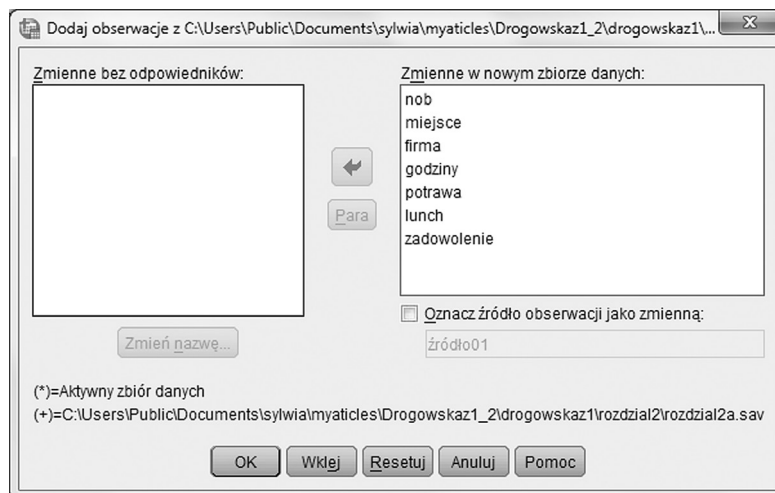
Teraz musimy połączyć te dane z wcześniej utworzonym zbiorem. Będziemy dodawać obserwacje. Program IBM SPSS Statistics w dość łatwy sposób pozwala połączyć dane, zarówno dodając obserwacje, jak i dołączając zmienne. Łączenie zbiorów ułatwia wpisanie do obu zbiorów zmiennej identyfikującej obserwacje. Ten unikalny identyfikator pozwala sprawdzić poprawność łączenia danych i automatyzację łączenia, gdy zbiory mają na przykład różną liczbę obserwacji. Przejdźmy do obliczeń. Otwieramy więc pierwszy zbiór danych *rozdzial2.sav* i w górnym menu wybieramy pozycję DANE, a następnie POŁĄCZ DANE (rysunek 2.36).



Rysunek 2.36. Wybór opcji DODAJ OBSERWACJE z górnego menu

Z dostępnej listy opcji wybieramy DODAJ OBSERWACJE i w oknie dialogowym, które pojawiło się w wyniku wyboru tej opcji, wybieramy plik, który zawiera te obserwacje, które chcemy dodać, a więc *rozdzial2a.sav*. Pod przyciskiem PRZEGLĄDAJ możemy wyszukać plik w odpowiednim folderze struktury plików na naszym komputerze. Po wybraniu pliku *rozdzial2a.sav* klikamy przycisk DALEJ i widzimy okno z dwiema listami zmiennych: ZMIENNE BEZ ODPOWIEDNIKÓW oraz ZMIENNE W NOWYM ZBIORZE DANYCH (rysunek 2.37).

W przypadku łączonych przez nas danych wszystkie zmienne pojawiły się na liście ZMIENNE W NOWYM ZBIORZE DANYCH. Zmienne mogą się pojawić też na liście ZMIENNE BEZ ODPOWIEDNIKÓW, jeśli mają różną nazwę lub też



Rysunek 2.37. Listy łączonych zmiennych

gdy mają odmienny typ. Tutaj nie mamy tego problemu, bo wcześniej dokonaliśmy już korekty zmiennej *potrawa*, więc łączenie zostaje wykonane bez trudu.

W przypadku, gdy zmienne różnią się jedynie nazwą i wiemy, że kodują to samo, możemy skorygować nazwę jednej ze zmiennych, by program mógł je połączyć w parę. Wybieramy wtedy z listy tę zmienną, której nazwę chcemy zmienić i klikamy przycisk ZMIENŃ NAZWĘ. Możemy także wybrać dwie zmienne z listy, a następnie włączyć przycisk PARA i przenieść je do okna z prawej strony.

	nob	miejsce	firma	godziny	potrawa	lunch	zadowolenie	var	var	var
1	1,00	1,00	1,00	39,00	1,00	1:30	4,00			
2	2,00	1,00	1,00	27,00	2,00	1:00	2,00			
3	6,00	2,00	1,00	40,00	5,00	1:00	5,00			
4	3,00	1,00	2,00	40,00	3,00	0:30	5,00			
5	4,00	2,00	2,00	42,00	1,00	0:45	3,00			
6	5,00	2,00	2,00	43,00	4,00	1:10	1,00			
7	7,00	1,00	1,00	34,00	6,00	0:00	5,00			
8	8,00	1,00	2,00	45,00	3,00	0:00	3,00			
9										
10										
11										
12										

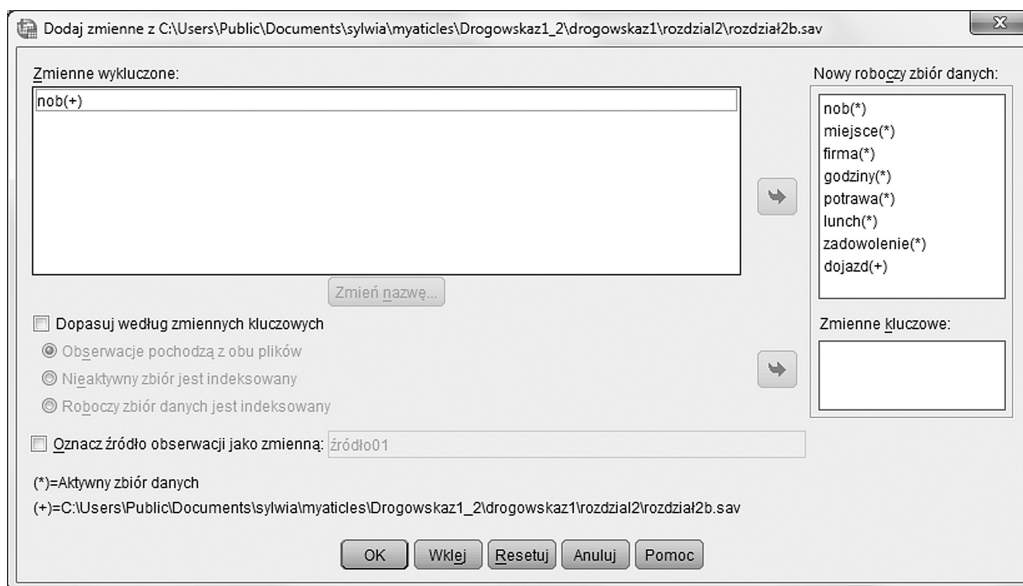
Rysunek 2.38. Plik danych z dołączonymi obserwacjami

DODAWANIE ZMIENNYCH

DODAWANIE ZMIENNYCH

W sytuacji, gdy chcemy dołączyć do zbioru dodatkowe zmienne, ważne jest, by oba pliki miały jakąś identyczną zmienną identyfikującą obserwacje. Tylko wtedy możemy bowiem być pewni, że dane jednej osoby zostaną zapisane we właściwym wierszu. Dołączmy więc do pliku, w którym mamy wyniki 8 osób (połączone wcześniej zbiory *rozdzial2* i *rozdzial2a*), dodatkową zmienną określającą sposób dojazdu do pracy (zmienna *dojazd*). Wchodzimy do okna dialogowego DANE – POŁACZ DANE – DODAJ ZMIENNE i ponownie wybieramy pod przyciskiem PRZEGLĄDAJ plik, w którym znajduje się dodatkowa zmienna (plik *rozdzial2b.sav*). W kolejnym oknie widzimy dwa zestawy zmiennych – ZMIENNE WYKLUCZONE – to te, które są zdublowane w obu plikach – tutaj jest to numer osoby badanej – zmienna *nob*, a pozostałe zmienne zostały zamieszczone w okienku NOWY ROBOCZY ZBIÓR DANYCH (rysunek 2.39).

Należy pamiętać, że kolejność osób badanych w obu zbiorach musi być identyczna. Tylko wtedy dane zostaną połączone prawidłowo. Dane z dodaną zmienną prezentuje rysunek 2.40.



Rysunek 2.39. Okno dialogowe DODAWANIA ZMIENNYCH

Podczas łączenia danych, zwłaszcza dodawania obserwacji, ważne jest, by wyniki osób badanych w obu plikach były ułożone w takiej samej kolejności. Jeśli tak nie jest, trzeba poustawiać dane.

	nob	miejsce	firma	godziny	potrawa	lunch	zadowolenie	dojazd	var	var
1	1,00	1,00	1,00	39,00	1,00	1:30	4,00	1,00		
2	2,00	1,00	1,00	27,00	2,00	1:00	2,00	1,00		
3	6,00	2,00	1,00	40,00	5,00	1:00	5,00	1,00		
4	3,00	1,00	2,00	40,00	3,00	0:30	5,00	1,00		
5	4,00	2,00	2,00	42,00	1,00	0:45	3,00	2,00		
6	5,00	2,00	2,00	43,00	4,00	1:10	1,00	2,00		
7	7,00	1,00	1,00	34,00	6,00	0:00	5,00	2,00		
8	8,00	1,00	2,00	45,00	3,00	0:00	3,00	2,00		
9										
10										
11										
12										

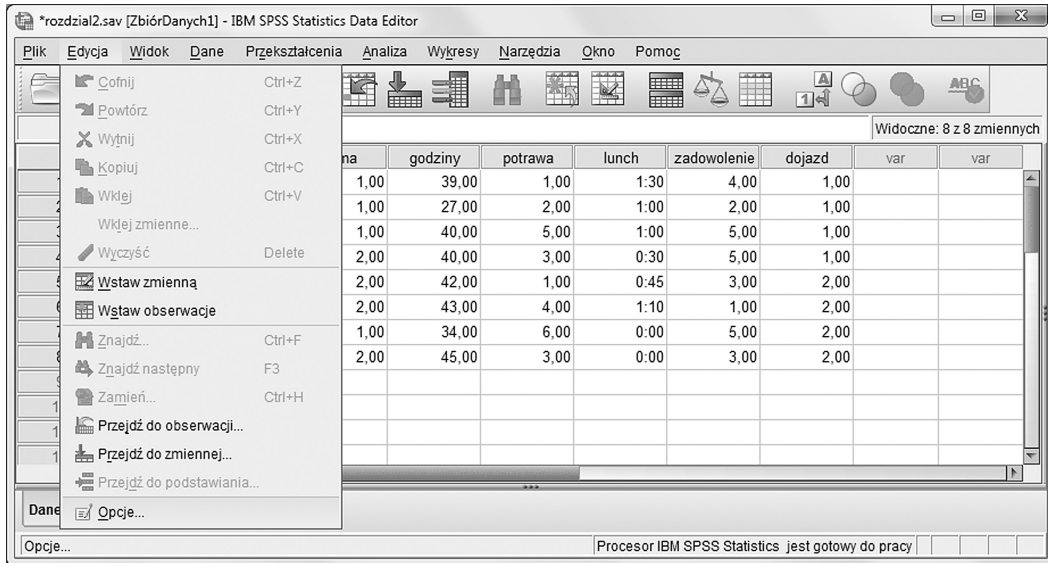
Rysunek 2.40. Okno danych z dodaną zmienną *dojazd*

DEFINIOWANIE OPCJI

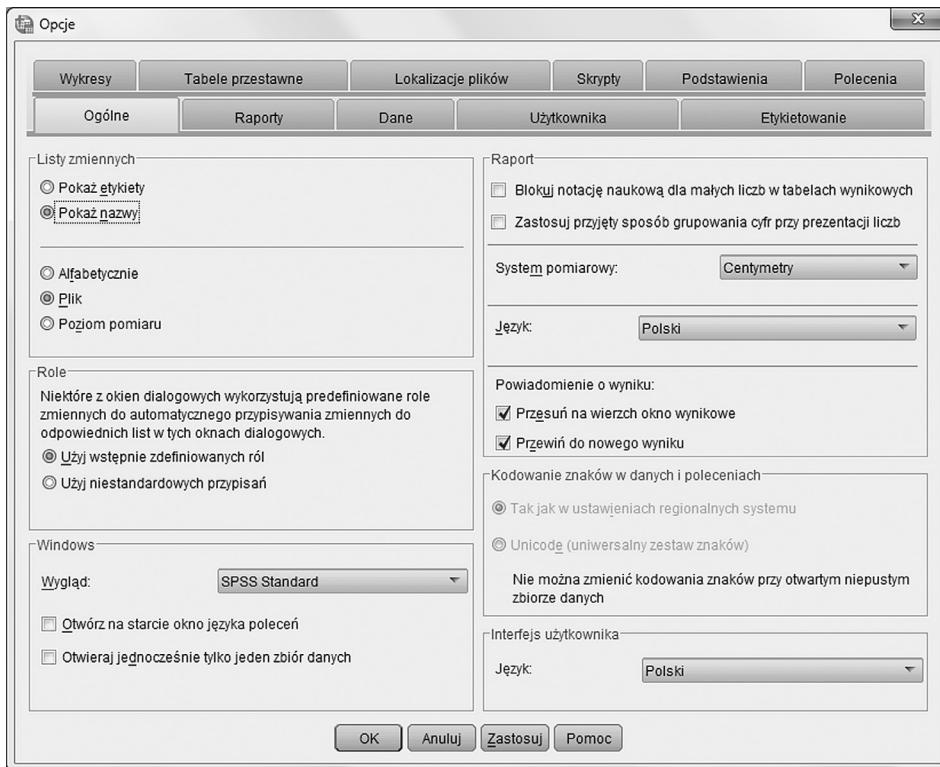
Program IBM SPSS Statistics pozwala na definiowanie ważnych opcji dotyczących języka, wyglądu tabel na wydruku i czcionek oraz sposobu wyświetlania zmiennych w listach w różnych oknach dialogowych za pomocą dość złożonego okienka OPCJE znajdującego się w górnym menu w zakładce EDYCJA (rysunek 2.41).

Po wejściu do tej pozycji górnego menu otwiera się okienko z kilkoma zakładkami, z których najciekawsze to: OGÓLNE, TABELE PRZESTAWNE, UŻYTKOWNIKA oraz ETYKIETOWANIE (rysunek 2.42).

W zakładce UŻYTKOWNIKA możemy zmienić separator miejsc dziesiętnych z przecinka na kropkę. To ważne dla osób, które posługują się językiem angielskim i chcą stosować notację angielską czy amerykańską. Pamiętajmy o zasadzie ogólnej, by wszystkie zmiany w tym oknie potwierdzić najpierw przyciskiem ZASTOSUJ, a dopiero potem OK. W zakładce TABELE PRZESTAWNE możemy wybrać styl wyświetlania tabel ze statystykami w Edytorze Raportów. Na uwagę zasługuje styl ACADEMIC, który jest zbliżony do standardu APA. W zakładce OGÓLNE mamy możliwość zdefiniowania języka, w którym program będzie się z nami komunikował. Możemy wybrać polski albo zmienić go, na przykład, na angielski. To jest także okno, w którym możemy określić sposób prezentacji zmiennych na listach: albo w postaci nazw albo etykiet. Zdecydowanie wygodniejsze wydaje się prezentowanie zmiennych w postaci nazw, ale wybór ten pozostawiamy czytelnikowi. Podobne informacje można zdefiniować



Rysunek 2.41. Wybór OPCJI z górnego menu



Rysunek 2.42. Okno dialogowe EDYCJI pozwalające definiować stałe ustawienia edytorów

w zakładce ETYKIETOWANIE, gdzie możliwe jest określenie sposobu prezentacji zmiennych (z etykietami zmiennych oraz etykietami wartości lub bez tych etykiet) zarówno w strukturze raportu, jak i w strukturze tabel przestawnych. W wielu przypadkach warto wybrać takie opcje, by zarówno zmienne, jak i ich wartości były opisywane jednocześnie przez wartości i etykiety czy nazwy i etykiety. Zwiększa to czytelność wydruku.

PODSUMOWANIE

IBM SPSS STATISTICS jest wygodnym narzędziem do przeprowadzania analiz statystycznych, zawierającym szereg rozwiązań technicznych, które pozwalają na automatyzację pracy (zob. Język Poleceń). Warto jednak pamiętać, że w rzeczywistości dokonuje on obliczeń za pomocą wprowadzonych wcześniej wzorów, a więc przypomina raczej kalkulator, a nie formę sztucznej inteligencji. Choć jest użyteczny w niektórych operacjach (wie, jak policzyć średnią czy inne statystyki), to w pewnych obszarach jest zupełnie bezradny. Po pierwsze nie wprowadzi sam za nas danych i nie jest w stanie sprawdzić, czy zrobiliśmy to poprawnie, jest też zupełnie ślepy na kwestię skal pomiarowych. Nie umie odróżniać, czy zmienna jest jakościowa, czy ilościowa. Pamiętajmy też, że użytkownik komunikuje się z programem za pomocą liczb poprzez wprowadzenie danych, a program przekazuje nam informację o rezultatach obliczeń także w formie liczbowej, pozostawiając interpretację wyników w naszej gestii. Aby wyciągnąć prawidłowe wnioski niezbędna jest znajomość statystyki. Nie musimy znać dokładnych wzorów obliczeniowych, ale musimy rozumieć sens poszczególnych analiz. Temu zagadnieniu będą poświęcone kolejne rozdziały.

Oczywiście nie sposób w tak krótkim rozdziale wyjaśnić wszystkich funkcji programu oraz tajników jego działania. Mamy nadzieję, że czytelnik zainteresowany tematem sięgnie do publikacji zwartych w bibliografii, do źródeł internetowych (www.IBM.SPSS.Statistics.pl) oraz zaawansowanej pomocy programu.