

Ilona Przybyłowska*, Maria Szymczak**

PRZYGOTOWANIE DANYCH Z WYWIADÓW KWESTIONARIUSZOWYCH
I ICH OPRACOWANIE NA EMC

I. Uwagi wstępne

W latach siedemdziesiątych zaczęto w polskich badaniach socjologicznych stosować techniki komputerowego opracowania danych w zakresie znacznie szerszym niż poprzednio. Wiązało się to przede wszystkim z włączeniem zagadnień socjologicznych do tzw. problemów centralnie sterowanych. Pracom badawczym podejmowanym w ramach tych problemów, wykorzystującym głównie standaryzowane techniki zbierania danych, postawiono krótkie terminy realizacji, co przy znacznym rozszerzeniu ich zakresu (duże próbki) eliminowało możliwość posłużenia się tradycyjnie stosowanymi sposobami opracowania ilościowego. Pojawiła się w związku z tym konieczność realizacji tej części zadań badawczych na elektronicznych maszynach cyfrowych.

Tylko nieliczni socjologowie mieli osobiste doświadczenia we współpracy z ośrodkami obliczeń elektronicznych i orientację co do reguł jakimi "rządzi się" komputerowe opracowanie danych. Brak wiedzy na ten temat, niezajomość języka programistów powodowały, iż w większości przypadków porozumienie między badaczami a pracownikami ośrodków obliczeniowych nie następowało szybko i było dla obu stron kłopotliwe. Trudności takie wystąpiły również w środowisku łódzkich socjologów. Nasunęło to pomysł opracowania tekstu instruktażowego, który dostarczałby podstawowych informacji o komputerowym opracowaniu danych z wywiadów kwestionariuszowych.

*Lr, adiunkt w Instytucie Socjologii UŁ.

**Mgr w Instytucie Socjologii UŁ.

Sytuacja badawcza, nazwana tu standardowymi badaniami kwestionariuszowymi, ma swoje cechy szczególne. Zawsze występuje w niej kwestionariusz, tzn. ustrukturalizowana lista pytań do respondentów. Wypowiedane podczas wywiadów i zapisane przez ankietera odpowiedzi na te pytania są źródłem informacji o faktach, postawach i opiniach występujących w realnych sytuacjach życiowych. Większość z tych pytań to różnego rodzaju pytania zamknięte. Uzyskuje się tu zatem dane ujednolicone. Nadają się one do opracowania ilościowego, które nie jest zazwyczaj zbyt skomplikowane (rozkłady zmiennych, krzyżówki).

Przygotowany tekst dotyczy tego rodzaju badań. Składa się z dwóch części. W pierwszej zapoznajemy czytelnika z budową i zasadami funkcjonowania elektronicznych maszyn cyfrowych, podając równocześnie podstawowe informacje o dostępnych w kraju komputerach i ich programowaniu. W części drugiej wyjaśniamy na czym polega i jak odbywa się zakładanie zbioru danych na emc. Omawiamy też niektóre zagadnienia związane z wykonywaniem obliczeń przez komputer.

II. Opis ogólny komputera

Komputer jest urządzeniem do przetwarzania informacji. Każdy problem, dla którego można podać algorytm postępowania (tj. sposób postępowania uwzględniający wszystkie możliwe przypadki) może być rozwiązany przez komputer dysponujący wystarczającymi możliwościami technicznymi.

Polecenia wykonania przez komputer operacji elementarnych nazywa się rozkazami. Wyróżnia się następujące grupy rozkazów: rozkazy wejścia, przesyłania, logiczne, arytmetyczne, sterujące i wyjścia.

Istnieje wiele typów komputerów. Wszystkie składają się jednak z pięciu podstawowych bloków funkcjonalnych: urządzeń wejściowych, pamięci, arytmometru, bloku sterowania i urządzeń wyjściowych.

A. U r z ą d z e n i a w e j ś c i o w e służą do wprowadzenia informacji (programów i danych) do komputera. Można je ogólnie podzielić na trzy podstawowe grupy, w zależności od tego, na jakim nośniku informacji wprowadza się informacje wejściowe:

- 1) czytniki kart perforowanych 80-kolumnowych, pracujące na zasadzie odczytu fotoelektrycznego,
- 2) czytniki taśmy perforowanej 5- 7- lub 8-ścieżkowe,
- 3) czytniki pisma magnetycznego, odczytujące i przenoszące do maszyny zawarte w dokumentach informacje wpisane ręcznie bądź w postaci umownych znaków naniesionych specjalnym atramentem, tuszem lub farbą o właściwościach magnetycznych; stosowane są najczęściej do obliczeń bankowych.

Karta perforowana jest obok taśmy perforowanej najpowszechniej stosowanym nośnikiem informacji. Karta 80-kolumnowa (zwana od nazwiska wynalazcy kartą Hollaritha) jest dosyć sztywnym, prostokątnym kartonikiem o wymiarach 189 x 83 mm. Ma ona 12 poziomych i 80 pionowych kolumn. Kombinacja dziurek w jednej kolumnie może reprezentować jeden znak (cyfrę, literę lub znak redakcyjny np. kropkę, przecinek) i wtedy na jednej karcie może być przedstawiona informacja składająca się maksymalnie z 80 znaków. Dziurka reprezentuje logicznie 1, a brak dziurki logicznie 0. Znaki numeryczne od 0 do 9 reprezentowane są przez dziurki w miejscach oznaczonych 0, 1, 2, ... 9. Pozostałe znaki reprezentowane są jako kombinacje kilku dziurek w jednej kolumnie wykorzystując nienumerowany wiersz 11, 12. Większość urządzeń do dziurkowania kart wyposażona jest w układy opisujące, które równocześnie z dziurkowaniem kodu danego znaku w kolumnie drukują ten znak nad kolumną. Jeden róg karty jest ścięty, co umożliwi jednoznaczne układanie kart w pliku. Czasem stosowane są karty 90- lub 96-kolumnowe.

Plik kart przeznaczonych do odczytania umieszczony jest w zasobniku czytnika kart perforowanych. W zasobniku tym mieści się od kilkuset do kilku tysięcy kart. Karty z zasobnika wybierane są kolejno jedna po drugiej przez mechanizm podający i transportowane za pomocą układu rolek napędowych poprzez system odczytu. Najczęściej stosowany jest fotoelektryczny system odczytu. Odczytywanie odbywa się kolumnami, czyli w jednej chwili odczytywany jest kod jednego znaku. Odczytane karty kierowane są do pojemnika odbiorczego.

Perforowana taśma papierowa jest stosowana jako nośnik informacji w urządzeniach wejściowych i wyjściowych przede wszystkim w małych maszynach. Kombinacja dziurek w jednym rzędku w poprzek taśmy reprezentuje kod jednego znaku. Najczęściej na taśmach perforowanych przechowywane są programy i pośrednie wyniki obliczeń, ponieważ bardzo szybko mogą one być wprowadzane do maszyny.

B. P a m i ę ć jest urządzeniem służącym do przechowywania informacji. Umożliwia ona "zapamiętanie" programu, danych potrzebnych do jego wykonania, wyników pośrednich i wyników końcowych. Pamięć składa się z ponumerowanych komórek, przy czym numer przypisany komórce nazywa się adresem. Urządzenia pamięci dzielimy na dwie grupy:

- 1) pamięć wewnętrzną zwaną pamięcią operacyjną,
- 2) pamięć zewnętrzną.

Pamięć wewnętrzną, czyli operacyjną najczęściej jest jedna w danym zestawie maszynowym, może być rdzeniowa lub bębnowa. Pamięć zewnętrzną dzieli się na taśmową i dyskową. Pamięć taśmowa należy do najbardziej rozpowszechnionego rodzaju pamięci zewnętrznej komputera. Informacja rejestrowana jest w niej na taśmie magnetycznej nawijanej na wymienne szpule. Jedna szpula zawiera zwykle 360 lub 720 m taśmy o znormalizowanej szerokości 12,7 mm. Informacje zapisywane są na taśmie magnetycznej blokami. Każdy blok posiada na początku znaki identyfikacyjne. Między blokami jest pusty kawałek taśmy nazywany przerwą międzyblokową. Szukanie danej informacji na taśmie odbywa się w ten sposób, że w pierwszej kolejności odczytywany jest znak identyfikacyjny bloku. Jeżeli w danym bloku nie ma poszukiwanej informacji, to jego zawartość nie jest czytana, a taśma zostaje automatycznie przewinięta do następnego bloku. Czas dostępu do żadanego bloku informacji zależy od jego położenia na taśmie i może sięgać kilku minut, jeśli zachodzi potrzeba przewinięcia całej szpuli. Pamięć taśmowa posiada następujące zalety:

- dużą pojemność szpuli taśmy,
- dużą szybkość przesyłania informacji,
- łatwą i szybką wymienialność szpul,
- tworzenie nieograniczonej ilości zbiorów danych oraz kilku kopii dla każdego z nich.

Pamięć dyskowa stanowi zespół wirujących dysków umieszczonych jeden pod drugim na wspólnej osi. Każdy dysk (cienki aluminiowy krążek) pokryty jest obustronnie warstwą nośnika magnetycznego. Zespół dysków zawiera zwykle 4-25 dysków i najczęściej stanowi wymienny pakiet. Informacja zapisywana (odczytywana) jest na obu powierzchniach każdego dysku, na ścieżkach tworzących współśrodkowe okręgi, przez głowicę zapisująco-odczytującą. Wymienne pakiety dyskowe dają możliwość tworzenia magazynów informacji o do-

wolnie wielkich pojemnościach. Czas dostępu do danej informacji jest o wiele krótszy niż w pamięci taśmowej, dlatego istnieje ogólna tendencja przechodzenia na pamięć dyskową.

Ponieważ pamięć wewnętrzna komputera jest ograniczona, dlatego pamięci zewnętrzne są niezbędne. W pamięciach zewnętrznych przechowywane są informacje do czasu zakończenia wszystkich obliczeń i wymazywane wtedy, gdy już nigdy więcej nie będą potrzebne; może to trwać dzień, miesiąc, itd. Natomiast pamięć wewnętrzna jest podzielona na dwie części. Jedna część jest stale zajęta na programy automatycznego sterowania całym komputerem. Pozostała część jest przeznaczona na dane i programy aktualnie przetwarzane. Po zakończeniu obliczeń i przesłaniu wyników jest ona wymazywana i wprowadzane są na jej miejsce nowe informacje.

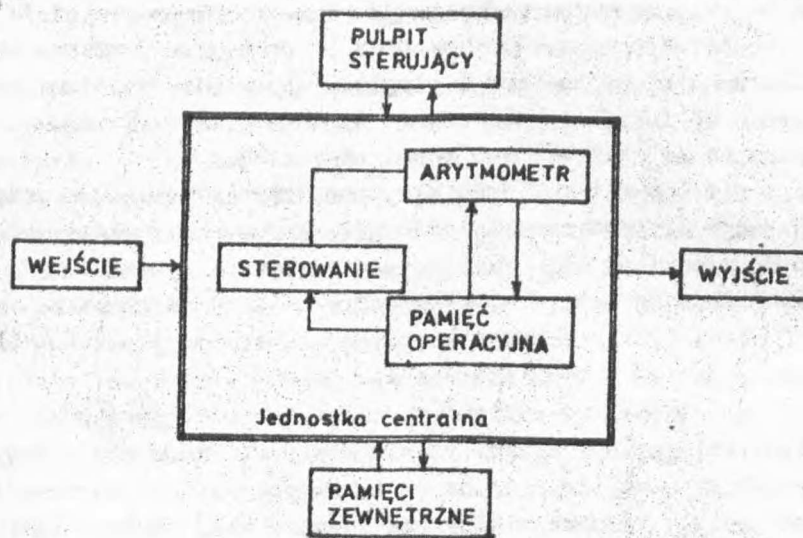
C. **A r y t m o m e t r** jest elektronicznym urządzeniem wykonującym operacje arytmetyczne i logiczne. Argumenty tych operacji przysyłane są z pamięci lub urządzeń wejściowych. Wyniki odsyłane są albo również do pamięci, albo bezpośrednio do urządzeń wyjściowych. O tym, jakiego rodzaju operacja ma być przez arytmometr wykonana, decyduje blok sterowania.

D. **B l o k s t e r o w a n i a**. Jego zadaniem jest pobieranie z pamięci kolejnych rozkazów programu i sterowanie pozostałymi blokami komputera w czasie realizacji tych rozkazów. Do bloku sterowania zalicza się również urządzenie służące do obsługi komputera przez operatora. Jest to tak zwany pulpit sterujący. Umożliwia on człowiekowi uruchamianie i zatrzymywanie komputera, wprowadzanie specjalnych kluczy do realizacji programów, zawiera sygnalizację nieprawidłowych stanów poszczególnych urządzeń. Przez pulpit sterujący człowiek komunikuje się z komputerem. Może to być podłączona do komputera maszyna do pisania lub monitor świetlny z klawiaturą maszyny do pisania.

E. **U r z ą d z e n i a w y j ś c i o w e** służą do wyprowadzania z komputera wyników programu. Podstawowym urządzeniem wyjściowym jest drukarka wierszowa. Informacje wychodzące na drukarkę wierszową to najczęściej końcowe wyniki obliczeń, ponieważ w tej postaci nie mogą być użyte jako dane wyjściowe do dalszych obliczeń. Wyniki wydrukowane na drukarce wierszowej nazywamy tabulogramem. Drukarki wierszowe mogą drukować 120-160 znaków w wierszu repertuar znaków wynosi 48-64, liczba egzemplarzy od 1 do 5. Do

wyprowadzenia wyników pośrednich z komputera służą dziurkarki kart lub dziurkarki taśmy papierowej podłączonej do komputera.

Urządzenia wejścia i wyjścia nazywane są urządzeniami peryferyjnymi. Blok sterowania, pamięć operacyjna (czyli wewnętrzna) i arytmometr umieszczone są w jednej obudowie i stanowią zasadniczy trzon komputera, dlatego określa się je jako jednostkę centralną. Kiedy wymieniamy nazwę maszyny, np. ODRA 1305, to oznacza ona nazwę jednostki centralnej w danej konfiguracji maszynowej.



Rys. 1. Schemat funkcjonowania EMC

Operator korzystając z pulpitu sterującego uruchamia maszynę. Kasuje poprzednią zawartość pamięci operacyjnej. Odpowiednim rozkazem wprowadza z urządzenia wejścia lub pamięci zewnętrznej program do pamięci operacyjnej i uruchamia jego realizację. Poszczególne rozkazy programu są odczytywane i zamieniane na sygnały sterujące. Jednym z pierwszych rozkazów w programie jest rozkaz czytania danych z konkretnego urządzenia. Rozkaz ten jest zamieniany na sygnał sterujący i następuje wczytywanie danych z urządzenia wejścia lub pamięci zewnętrznej do pamięci operacyjnej maszyny. Kolejnymi rozkazami programu dane z pamięci operacyjnej przenoszone są do arytmometru, gdzie dokonywane są obliczenia. Wyniki obliczeń, pośrednie lub końcowe, przenoszone są do pamięci operacyjnej i stamtąd do urządzenia wyjścia lub urządzeń pamięci zewnętrznych.

III. Programowanie i języki programowania komputera

Program zawiera ciąg rozkazów określających, jakie czynności, na jakich informacjach i w jakiej kolejności ma wykonać komputer oraz opisy obszarów pamięci, w których będą się znajdowały informacje w trakcie pracy komputera. Czynności związane z opracowywaniem tego rodzaju programów nazywamy programowaniem. Język, w którym odbywa się programowanie nazywa się językiem programowania.

Pierwsze komputery programowane były w językach komputera. Wszystkie komputery działają w systemie binarnym, dlatego rozkazy muszą być również przedstawione w systemie binarnym. Napisanie takiego programu wymaga przede wszystkim doskonałej znajomości listy rozkazów oraz struktury komputera. Programowanie w języku komputera jest więc żmudne i pracochłonne.

Obecnie nie programuje się w języku wewnętrznym maszyny, ale w języku zewnętrznym, który jest mniej skomplikowany. Komputery wyposażone są w specjalne programy zwane translatorami, które tłumaczą program napisany w danym języku zewnętrznym na język wewnętrzny maszyny. Mając translatorów wielu języków zewnętrznych, można rozwiązać dany problem za pomocą języka najbardziej dogodnego z punktu widzenia rozwiązywanego problemu. Daje to skrócenie czasu pisania programu, a także skrócenie czasu realizacji tego programu, czyli czasu komputera.

Spośród wielu języków programowania na szczególną uwagę zasługuje grupa języków problemowych. Do grupy tych języków należą: FORTRAN, ALGOL, COBOL, PL/1. Języki ALGOL i FORTRAN przeznaczone są do rozwiązywania problemów numerycznych - czyli do obliczeń naukowo-technicznych. COBOL przeznaczony jest do przetwarzania danych - do operacji i obliczeń dokonywanych na dużych zbiorach danych, przy czym obliczenia te nie są matematycznie skomplikowane. Język PL/1 jest połączeniem języków FORTRAN i COBOL.

Język FORTRAN został opracowany w 1957 r. dla komputera typu IBM 704, a obecnie większość współczesnych komputerów wyposażona jest w translator tego języka. Wszystkie sprzedawane maszyny typu IBM posiadają ten translator, wliczony w cenę maszyny, dostarczany przez producenta razem z maszyną. Język FORTRAN od czasu powstania jest modyfikowany i udoskonalany, obecnie staje się językiem najszerzej stosowanym na wszystkich typach maszyn.

IV. Polskie i eksploatowane w Polsce EMC

Poszczególne fazy rozwoju komputerów przyjęto nazywać generacjami, gdyż co 5-8 lat zmienia się zasadniczo wewnętrzna konstrukcja maszyn.

Komputery pierwszej generacji z połowy lat czterdziestych budowane były w technice lampowej. Polskie komputery pierwszej generacji to UMC-1, ZAM-1 i ZAM-2.

Komputery drugiej generacji pojawiły się pod koniec lat pięćdziesiątych. Realizowane były w technice tranzystorowej. Znaczna część maszyn cyfrowych eksploatowanych w Polsce to maszyny drugiej generacji: ZAM-41, ODRA-1204, ODRA-1304, ICT-1900.

Komputery trzeciej generacji budowane są przy użyciu układów scalonych. Powstały w połowie lat sześćdziesiątych. Do eksploatowanych w Polsce komputerów trzeciej generacji należą: ODRA-1305, ODRA-1325, IBM-360/370 oraz komputery serii RIAD.

Komputery czwartej generacji pojawiły się na przełomie lat sześćdziesiątych i siedemdziesiątych. Są to udoskonalone maszyny trzeciej generacji. Ponadto urządzenia zewnętrzne są dużo szybsze od poprzednich, co zwiększa szybkość pracy całego komputera.

Rozwój technologii układów scalonych umożliwił szerokie rozpowszechnienie się małych komputerów zwanych minikomputerami. Minikomputery posiadają niewielką pamięć i kilka urządzeń zewnętrznych - są szybkie, tanie i łatwe w obsłudze, a możliwości obliczeniowe dorównują możliwościom wielu maszyn drugiej generacji. Minikomputery pojawiły się w latach siedemdziesiątych i znalazły zastosowanie w biurach konstrukcyjnych, w pracowniach laboratoryjnych.

Podobne konstrukcje i podobne zastosowanie posiadają mikrokomputery. Mikrokomputer posiada urządzenia zewnętrzne i pamięci zewnętrzne, a sam przypomina maszynę do pisania. Zdaniem wielu rzeczoznawców w najbliższym czasie wielkie i średnie komputery będą stanowiły 25% wszystkich komputerów, pozostałe 75% to mini i mikrokomputery.

Istotą rozwoju komputerów nie jest ich szybkość, miniaturyzacja i pojemność pamięci, lecz możliwość zaprogramowania coraz większej ilości procesów z zakresu nauki, techniki i kultury.

O jakości poszczególnych komputerów informują następujące parametry:

- 1) szybkość wykonania operacji;
- 2) pojemność pamięci wewnętrznej maszyny - nie zawsze duża pojemność dla danego użytkownika jest konieczna;
- 3) wieloprogramowość - w tym samym czasie realizowanie kilku programów: ODRA-1204 - 2-programowa, ODRA-1304 - 4-programowa, ODRA-1305 - 16-programowa. Przy wieloprogramowości potrzeba bardziej rozbudowanej konfiguracji maszynowej o elementy wejścia i wyjścia, a szczególnie o urządzenie wyjścia;
- 4) wielodostępność - wielu użytkowników równocześnie komunikuje się z jednostką centralną za pomocą własnych urządzeń zwanych końcówkami lub terminalami. Terminale połączone są z jednostką centralną liniami transmisji danych, wszyscy użytkownicy obsługiwani są jednocześnie, czyli tak jak gdyby każdy dysponował własną maszyną. Zwykle terminalem jest dalekopis z ekranem służący do wyświetlania rysunków technicznych, wykresów i tekstów wprowadzanych do jednostki centralnej lub uzyskanych z jednostki centralnej.

V. Charakterystyka systemu OSIRIS

OSIRIS - to jest nazwa pakietu programów, zaprojektowanego jako narzędzie pracy dla ośrodków i instytucji badawczych w dziedzinie nauk społeczno-politycznych.

Pakiet powstał w wyniku współpracy Instytutu Socjologii z Centrum Obliczeniowym Uniwersytetu Michigan.

System OSIRIS może być realizowany na elektronicznej maszynie cyfrowej typu: IBM-360, IBM-370 oraz na EMC Jednolitego Systemu serii RIAD pod nadzorem systemu operacyjnego - OS.

Aktualna wersja pakietu OSIRIS obejmuje 60 programów, z których większość została napisana w języku FORTRAN-IV, a pozostała część w języku PL/1.

Za pomocą pakietu OSIRIS można wykonać wszystkie te obliczenia, które są niezbędne do przeprowadzenia badania wymagającego posłużenia się współczesnymi metodami statystycznymi. W rezultacie wykonania określonych programów uzyskuje się:

- a - zbiory danych
- b - zbiory danych pośrednich
- c - wyniki w postaci finalnej.

Dane pośrednie stanowią dane wejściowe do innego programu lub grupy programów. Dane pośrednie mogą być wyprowadzane w postaci zbiorów na różne nośniki informacji, którymi są: taśma magnetyczna, dysk magnetyczny, karta perforowana. Przykładem danych pośrednich może być tablica współczynników korelacji.

Wyniki obliczeń końcowych uzyskiwane są głównie w formie zre-dagowanych tablic statystycznych jedno lub dwuwymiarowych. Tablice jednowymiarowe zawierają rozkłady liczebności jednej cechy, a tablice dwuwymiarowe podają rozkłady warunkowe dla dwóch zmiennych, jak również rozkłady brzegowe dla tych cech.

Liczebności podawane są w liczbach bezwzględnych, a także istnieje możliwość opcjonalnego obliczenia indywidualnych i kumulowanych rozkładów procentowych w wierszach i kolumnie. Oprócz tablic uzyskuje się również wyniki w innej formie, np. w postaci wykresu.

Programy pakietu OSIRIS można podzielić w zależności od wykonywanej funkcji na dwie grupy:

- 1) programy organizacyjno-manipulacyjne (zarządzania danymi),
- 2) programy analizy danych.

Programy grupy pierwszej (zarządzania danymi) służą do wykonywania takich operacji, jak:

- zakładanie zbioru słownika i zbioru danych¹,
- tworzenie nowych zmiennych,
- listowanie danych czyli wydruk danych,
- korekty danych,
- sprawdzenie poprawności kodów,
- agregacja danych,
- łączenie zbiorów.

Przy pomocy programów grupy drugiej wykonuje się operacje analizy danych, które polegają na obliczeniu różnych mierników statystycznych, jak: średnia arytmetyczna, mediana, dominanta, od-

¹ Przygotowane na podstawie informacji źródłowych dane wyjściowe opisywane są za pomocą innego zbioru, zwanego słownikiem. Słownik ten jest pomocniczym zbiorem rekordów identyfikujących w sposób jednoznaczny każdą zmienną i jednocześnie opisujących format i położenie danych.

chylenia standardowe, współczynnik korelacji itd. oraz na zestawieniu odpowiednich tablic, a także prezentowaniu wyników w formie wykresów na osi współrzędnych.

W połowie lat siedemdziesiątych system OSIRIS został sprowadzony do Warszawy i uruchomiony na maszynie IBM w Zakładzie Elektronicznej Techniki Obliczeniowej (ZETO). Użytkownikami tego systemu są głównie socjologowie środowiska warszawskiego. Należy podkreślić, że nadal system ten nie jest w pełni poznany i wykorzystany. Trwają prace nad uruchomieniem nowych programów z tego pakietu, szukaniem nowych rozwiązań wewnątrz tych programów. System OSIRIS posiada trzy cechy:

- 1) jest nowoczesnym systemem,
- 2) język, w którym pisane są programy, należy do najefektywniejszych w rozwiązaniach problemów numerycznych,
- 3) pracuje na najnowszym sprzęcie.

Wszystkie te cechy dają duże możliwości doskonalenia efektywności działania systemu.

Ważnym elementem ułatwiającym współpracę i podnoszącym efekty pracy jest właściwe przygotowanie użytkownika. Użytkownik systemu (np. socjolog) powinien nie tylko posiadać wiedzę fachową w zakresie problemu badawczego, specyficzną koncepcję, narzędzie badania, być zawsze najbliższym procesowi badawczego itp., ale również powinien być przygotowany do współpracy z informatykami (programistami). Powinien przygotować swoje badania mając na uwadze późniejsze opracowanie komputerowe materiału. Jeśli informatyk posiada podstawowe informacje w zakresie wiedzy fachowej reprezentowanej przez użytkownika i użytkownik przygotowany jest do opracowania maszynowego swoich badań, wtedy praca obu jest efektywniejsza i przyjemniejsza.

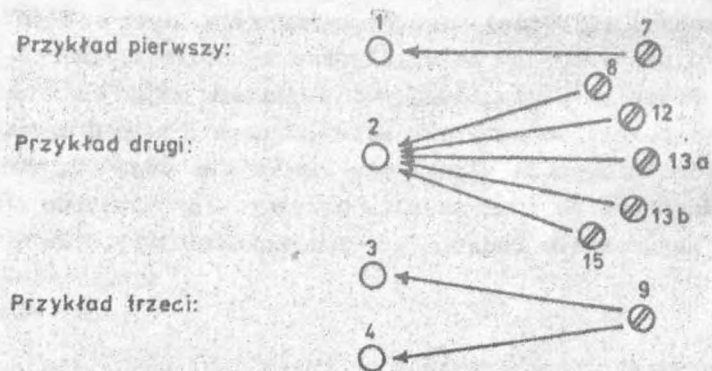
VI. Czynności poprzedzające budowę instrukcji kodowej

Przed przystąpieniem do budowy instrukcji kodowej badacz powinien dysponować listą problemów badawczych, które chce rozstrzygnąć poprzez ilościową analizę danych kwestionariuszowych. W badaniach starannie przygotowanych od strony koncepcyjnej lista taka jest już w zasadzie gotowa i stanowi element koncepcji badawczej. Oczywiście w omawianej tu fazie badania lista ta może

być zmodyfikowana poprzez wyłączenie z analizy niektórych problemów bądź też uwzględnienie nowych.

Następną czynnością jest przygotowanie listy zmiennych ważnych dla rozwiązania tych problemów. Zgodnie ze znaną niektórym socjologom terminologią J. Lutyńskiego poszczególne zmienne są równoważne poszczególnym badanym zjawiskom. Wyróżniając wartości danej zmiennej określamy stany rzeczy składające się na badane zjawisko. Rozstrzygnięcie tzw. jednostkowego pytania badacza czyli ustalenie jednostkowej informacji polega na ustaleniu wartości jaką przyjmuje dana zmienna w przypadku danej jednostki badania. Uzyskanie odpowiedzi na to pytanie w stosunku do kolejnych jednostek badania stanowi podstawę do rozwiązania problemu badawczego.

Dysponując listą zmiennych należy ustalić w stosunku do każdej z nich, z którym pytaniem z kwestionariusza jest ona tak powiązana, że odpowiedzi na to pytanie stanowiąc będą podstawę do ustalenia wartości, jakie ta zmienna może przyjmować. Oczywiście może być tak, że jedna zmienna - podobnie jak poszukiwana informacja - jest powiązana z kilkoma pytaniami kwestionariusza lub też jedno pytanie z kwestionariusza jest powiązane z kilkoma zmiennymi (por. rys. 2).



Legenda:

- - pojedyncza zmienna użyta przy opisie danego problemu
- ◐ - pojedyncze pytanie kwestionariusza

Rys. 2. Schemat graficzny powiązań zmiennych z pytaniami kwestionariusza

Mając pełną orientację w omówionej poprzednio kwestii przystępujemy do sporządzenia dwóch odrębnych list: 1) listy tzw. zmiennych pierwotnych oraz 2) listy tzw. zmiennych transformowanych.

Zmiennymi pierwotnymi będą te, które zostaną zapisane w kartach kodowych, natomiast zmiennymi transformowanymi te, które zbuduje komputer odpowiednio przekształcając jedną lub kilka zmiennych pierwotnych. Teoretycznie rzecz biorąc zmienne transformowane nie są konieczne, ponieważ koder przy wpisywaniu symbolu kodowego, odpowiadającego także bardzo złożonej zmiennej i informacji, może wykorzystywać łącznie odpowiedzi nawet na kilka pytań z kwestionariusza i kilka związanych z nimi pierwotnych informacji. Taki sposób postępowania nie jest jednak racjonalny ze względu na częste wówczas omyłki koderskie. Stąd też należy kierować się regułą, aby droga od zapisanych w kwestionariuszach odpowiedzi do symboli zapisanych w kartach kodowych nie była zbyt złożona i nie wymagała zbyt skomplikowanych operacji koderskich. Przy wykonywaniu każdej z tych operacji można bowiem popełnić błędy. Wyeliminowanie tych błędów jest wprawdzie możliwe w systemie komputerowego opracowania danych, ale jest bardzo pracochłonne. W sumie bardziej racjonalne jest maksymalne uproszczenie czynności koderskich i powierzenie wszelkich bardziej skomplikowanych operacji związanych z budowaniem zmiennych komputerowi.

Nie opłaca się np. kodować zmiennej podanej w przykładzie drugim ze schematu graficznego, gdyż ustalenie jej wartości będzie wymagało od koderów równoczesnego śledzenia odpowiedzi na pięć pytań z kwestionariusza. Bezpieczniej jest odpowiedź na każde z tych pytań kodować niezależnie. Wiąże się to z koniecznością jej rozbicia na 5 zmiennych cząstkowych o charakterze zmiennych pierwotnych, które zostaną później odpowiednio zagregowane przez komputer w ramach operacji transformowania zmiennych.

Niepotrzebne może się również okazać podwójne kodowanie odpowiedzi na pytania przedstawione w trzecim przykładzie: raz z punktu widzenia zmiennej nr 3, drugi raz z punktu widzenia zmiennej nr 4, jeśli tylko któraś z tych zmiennych będzie mogła powstać poprzez odpowiednie przekształcanie przez komputer jednej z nich.

Podsumowując, można by powiedzieć, iż dążąc do redukcji błędów, które mogą powstać podczas kodowania odpowiedzi z jednej strony, oraz maksymalnego wykorzystania możliwości operacyjnych

komputera z drugiej, należy przestrzegać zasady by kodowane zmienne, a więc zmienne pierwotne - były związane z pojedynczymi pytaniami z kwestionariusza i odpowiadały najprostszym poszukiwanym informacjom. Wyjaśnić tu należy, iż pojęciem "pytania z kwestionariusza" posługujemy się tu w sensie, jaki nadał mu J. Lutyński². Zdaniem tego autora pytanie z kwestionariusza to "wypowiedź badacza najczęściej w formie zdania pytajnego, skierowana do respondenta i żądająca od niego wypowiedzi określonej treściowo bardziej lub mniej dokładnie. Pytaniu towarzyszy zawsze dyrektywa zapisu reakcji (odpowiedzi) respondenta we wskazanym miejscu". Zgodnie z tym określeniem tzw. pytania tabelki składają się zawsze z kilku pytań, niezależnie od tego, że w kwestionariuszu bywają oznaczone tym samym numerem. Z wielu pytań w istocie składają się też takie pytania, w których po wymienieniu każdego elementu z listy zapisuje się osobno odpowiedź (np.: że respondent posiada dany przedmiot lub go nie posiada).

Przy sporządzaniu listy zmiennych pierwotnych korzystne jest również przestrzeganie innej jeszcze zasady, którą stosuje się do niektórych pytań zamkniętych.

Założmy, iż w kwestionariuszu umieszczono następujące pytanie:

"Prosiłbym, aby przez chwilę pomyślał P. o młodych ludziach pragnących w przyszłości zdobyć wysokie miejsce w społeczeństwie. Które z podanych czynników mają - P. zdaniem - wpływ na zdobycie wysokiego miejsca w naszym społeczeństwie:

1. wykształcenie
2. pochodzenie społeczne
3. płeć
4. przynależność partyjna
5. kwalifikacje zawodowe
6. zamożność rodziców
7. miejsce zamieszkania (wieś - miasto)
8. dobrze wykonywana praca
9. znajomości
10."

² Por. J. L u t y ń s k i, Pytanie jako narzędzie w surveyowych badaniach socjologicznych, "Studia Socjologiczne" 1979, z. 2.

W odpowiedzi na to pytanie poszczególni respondenci mogą wybrać od jednego do dziesięciu czynników. W tej sytuacji wygodniej jest uznać za zmienne pierwotne poszczególne punkty kafeterii i przyjmując, iż każda z nich przyjmuje dwie wartości: 1) respondent wybrał ten czynnik, 2) respondent nie wybrał tego czynnika. Zwiększy to wprawdzie ilość kolumn, które należy zarezerwować na zakodowanie odpowiedzi na to pytanie, ale znacznie uprości przygotowanie instrukcji kodowej i kodowanie danych. Ma to również tę zaletę, iż niezależnie od tego, jak badacz zamierza opracować odpowiedzi na to pytanie (obliczyć jak często wybierane są poszczególne czynniki, czy też wyróżnić i obliczyć, jak często występują określone kombinacje czynników), zawsze będzie mógł to osiągnąć, przy czym wszelkie bardziej skomplikowane operacje (w tym przypadku tworzenie kombinacji) wykona komputer.

Podobne racje przemawiają za tym, by identycznie postępować w przypadku pytań, w których żąda się wyboru kilku punktów kafeterii i wybór ten jest liczbowo określony. Załóżmy teraz, że w poprzednim pytaniu inaczej sformułowano polecenie skierowane do respondenta:

"Proszę wybrać najwyżej trzy czynniki mające P. zdaniem wpływ na zdobycie wysokiego miejsca w naszym społeczeństwie, a następnie uszeregować je od najważniejszego".

W tym przypadku - z identycznych jak poprzednio powodów - należałoby wyróżnić 3 następujące zmienne pierwotne: 1) czynniki preferowane na pierwszym miejscu, 2) czynniki preferowane na drugim miejscu, 3) czynniki preferowane na trzecim miejscu, traktując punkty kafeterii jako wartości każdej z tych zmiennych.

VII. Budowa instrukcji kodowej

Instrukcja kodowa jest podstawowym narzędziem badawczym wykorzystywanym w fazie przygotowania danych do analizy na EMC. Badacz powinien sporządzić ją osobiście. Punktem wyjścia dla jej opracowania jest lista zmiennych pierwotnych poszerzona o zmienne charakteryzujące próbę, które będą w analizie uwzględniane, a nie są związane z jakimś pytaniem kwestionariusza.

Instrukcja kodowa powinna zawierać następujące pozycje:

- 1) numer i nazwa kodowanej zmiennej:
- 2) numer pytania w kwestionariuszu, ewentualnie określenie miejsca w kwestionariuszu, w którym zapisano kodowany parametr próby:
- 3) klasyfikację bądź typologizację badanej zmiennej (mówiąc inaczej, wyróżnienie wartości danej zmiennej):
- 4) kryteria przyporządkowania zapisanych w kwestionariuszu informacji wyróżnionym klasom zmiennej (np. podanie wskaźnikowych odpowiedzi zaliczonych do danej klasy):
- 5) symbolizację numeryczną wyróżnionych klas:
- 6) określenie miejsca zapisu symboli numerycznych w karcie kodowej.

Klasy danej zmiennej wyróżnia się mając na uwadze rodzaj informacji potrzebnych dla rozwiązania danego problemu badawczego. Należy się również starać o to, by odzwierciedlały one "naturalną" strukturę kodowanego materiału. Spełnienie tego drugiego warunku jest szczególnie ważne, gdy materiał ten stanowią swobodne odpowiedzi na pytanie otwarte. Należy się z tymi odpowiedziami zapoznać wypisując je na oddzielnych kartkach z kwestionariuszy. Wypisami obejmuje się ok. 10% kwestionariuszy dobranych losowo.

Proponowana w instrukcji klasyfikacja zmiennej musi spełniać wymogi rozłączności, wyczerpywalności i jednorodności wewnętrznej klas. Niekiedy dla spełnienia wymogów rozłączności i wyczerpywalności zmuszeni jesteśmy tworzyć takie klasy jak "inne", "pozostałe" itp. W pewnych sytuacjach (analiza wielowątkowych odpowiedzi na pytania otwarte) możemy zrezygnować z przestrzegania kryterium wewnętrznej jednorodności klas i stosować typologizację zmiennej. Wówczas należy pamiętać o tym, że kodowana odpowiedź może należeć tylko do jednej z wyróżnionych klas i musi być wiadomo do której.

Dla symbolizacji wyróżnionych w instrukcji klas dozwolone są przy opracowaniu komputerowym znaki numeryczne (cyfry). Tzw. znaki dokładane (np. X, Y, Z) nie są tolerowane przez wszystkie programy i dlatego należy ustalić z programistami, czy i które z nich można w danym badaniu wykorzystać. Warto starać się o ujednoczenie tej symbolizacji. Wygodniej jest, gdy pewne stałe powtarzające się wartości zmiennych oznaczane są tymi samymi znakami. Ułatwia to zarówno kodowanie danych, jak i analizę tabulogramów. Znaki dokładane warto wykorzystywać dla symbolizacji klas, które nie są

nośnikami treści merytorycznie znaczących (np.: brak danych, nie dotyczy itp.). Przyspiesza to znacznie uzyskanie "czystych" tablic, tzn. takich, w których pominięto obserwacje nie mające merytorycznego znaczenia.

Wyjaśnić należy, iż kodowanie zmiennych ilościowych (np.: wysokość dochodów, liczba osób w gospodarstwie domowym itp.) polega na ich przepisywaniu z kwestionariusza na karty kodowe. Nie ma zatem potrzeby ustalania z góry jakichkolwiek przedziałów (klas danej zmiennej), ponieważ może to wykonać komputer.

Gdy kodowana zmienna ma charakter porządkowy, wtedy symbole numeryczne muszą być uporządkowane w kolejności rosnącej. Ułatwi to tworzenie innych zmiennych, bardziej złożonych, których elementami są te zmienne porządkowe.

Przygotowaną instrukcję należy sprawdzić kodując pewną ilość losowo dobranych kwestionariuszy. Ten swoisty pilotaż instrukcji kodowej przeprowadza się po to, aby ustalić, czy podane w instrukcji zasady przyporządkowywania odpowiedzi do poszczególnych klas są jednoznaczne. Jeśli koderzy nie są pewni, do której z wyróżnionych klas powinni zaliczyć daną odpowiedź, to jest to sygnał, że podane w instrukcji kryteria nie są zbyt precyzyjne i należałoby je zmienić. Gdy jest to niemożliwe, trzeba liczyć się z koniecznością podwójnego kodowania, jeśli nie całego kwestionariusza, to przynajmniej pytań, przy których takie wątpliwości koderskie wystąpiły (szersze uwagi na temat podwójnego kodowania znajdują się w dalszej części tego opracowania). Sprawdza się tu ponadto, czy w klasie "inne" mieści się nie więcej niż 5% odpowiedzi, oraz czy są one rzeczywiście wewnątrznie niejednorodne tzn. nie wiążą się z którąś z wyróżnionych klas.

VIII. Projektowanie karty kodowej

Karta kodowa stanowi pośrednik między kwestionariuszem a kartą perforowaną. Służy do zapisu zasymbolizowanych informacji. Jest dokumentem opracowywanym na użytek danego badania. Najczęściej zawiera 80 kolumn, co wynika z faktu, że tyleż kolumn zawiera powszechnie stosowana karta perforowana. Ponieważ nie jest to regułą, badacz musi porozumieć się z ośrodkiem obliczeniowym i

ustalić, ile kolumn będą zawierały karty perforowane użyte w jego badaniu.

Dane w karcie kodowej mogą być przedstawione w układzie pozycyjnym (tzn. kartę tworzy się dla kwestionariusza a numer kolumny jest zawsze związany ze zmienną) lub w układzie liczbowym (tzn. kartę tworzy się dla zmiennej). Dla omawianego tu typu badań charakterystyczny jest pozycyjny układ karty kodowej. W dalszej części rozważań zajmować się będziemy kartą kodową o układzie pozycyjnym.

Rzadko udaje się zapisać wszystkie dane dotyczące jednostki badania w obrębie jednej karty kodowej. Zazwyczaj dane te zapisuje się na kilku kartach. W przypadku gdy przewiduje się znacznie większą ilość kart kodowych dla zapisu danych pochodzących z jednego kwestionariusza, a zbiór kwestionariuszy jest duży, należy upewnić się, czy pamięć operacyjna komputera, który mamy do dyspozycji może przyjąć tak dużą ilość danych. Jeśli pamięć operacyjna nie jest dostatecznie pojemna, należy uzgodnić z programistami podział danych na bloki lub segmenty, na których będą dokonywane obliczenia.

Każda karta kodowa w układzie pozycyjnym zawiera część identyfikacyjną i merytoryczną. W części identyfikacyjnej zapisuje się najczęściej następujące informacje: 1) numer jednostki badania, 2) numer karty kodowej (w przypadku większej ilości kart niż jedna), 3) numer kodingu (przy podwójnym kodowaniu). Zmienne identyfikacyjne muszą być zapisane na początku każdej karty i należy na ich zapis zarezerwować odpowiednią ilość kolumn.

Ogólne zasady rezerwacji kolumn dla zapisu wartości danej zmiennej są proste. W pojedynczej kolumnie karty kodowej mogą być zapisane znaki numeryczne od 0 do 9 oraz znaki dokładane np. X, Y, Z. Jeśli zmienna przyjmuje nie więcej niż 10 wartości wystarczy dla jej zapisu zarezerwować jedną kolumnę. Jeśli ma ona więcej wartości niż 10, należy dla jej zapisu zarezerwować tyle kolumn, ile miejsc dziesiętnych zajmuje liczba równoważna liczbie wartości danej zmiennej.

W przypadku, gdy rezerwujemy dwie kolumny dla zapisu jakiejś zmiennej, np. kol. 12 i 13, a kodowana wartość tej zmiennej jest np. 7, wówczas kol. 12 zostawiamy pustą lub uzupełniamy ją ustalonym z programistami znakiem (np. 0), natomiast w kol. 13 wpisujemy znak 7.

Przygotowując kartę kodową należy pamiętać o tym, iż nie wolno zapisywać jednej zmiennej wielokolumnowej "okrakiem" na dwóch kartach. Nie jest natomiast błędem pozostawianie między zmiennymi pustych kolumn.

Warto włożyć nieco wysiłku w opracowanie kart kodowych od strony graficznej. Koduje się znacznie łatwiej i mniej popełnia się błędów, gdy karta jest czytelna. Zazwyczaj rysuje się prostokąty, wpisując u dołu nr kolumny. Wskazane jest zachowanie refleksów świetlnych między kolumnami zarezerwowanymi dla kodowania poszczególnych zmiennych. Identyfikację zmiennych znacznie ułatwia wprowadzenie do karty ich numeracji, bądź wpisanie ich nazwy. W tym miejscu chcemy wyjaśnić, iż numeracja zmiennych może, ale nie musi być ciągła. Inny konkurencyjny sposób numerowania nawiązuje do miejsca zapisu danej zmiennej. Numer zmiennej otrzyma się stosując następujący wzór:

nr karty kodowej, w której nr kolumny, w której zapisuje
zapisuje się daną zmienną $\times 100 +$ się daną zmienną³

Przypuśćmy, że jakaś zmienna zapisywana jest w karcie nr 9 w kolumnie nr 42. Zgodnie z tym wzorem należy jej przypisać nr 942. Inna zmienna zapisywana jest w karcie nr 11 w kol. nr 43. Przypisujemy jej w związku z tym nr 1143. Stosując ten wzór łatwo możemy zidentyfikować, w której karcie i w której kolumnie zapisana jest dana zmienna. Dwie ostatnie cyfry oznaczają zawsze nr kolumny, natomiast cyfry je poprzedzające nr karty kodowej. Stosując numerację zmiennych wg powyższego wzoru nie musimy jej wprowadzać do kart kodowych.

Czy karta kodowa jako ogniwo pośredniczące między kwestionariuszem a kartą perforowaną jest zawsze potrzebna? Wydaje się ona zbędna w sytuacji, gdy analizą obejmujemy niezbyt dużo zmiennych i jesteśmy w stanie przygotować instrukcję kodową bez posługiwania się wypełnionymi kwestionariuszami. Jeśli rezygnujemy z karty kodowej, musimy przewidzieć w kwestionariuszu miejsce na zapis symboli kodowych. Przed powieleniem kwestionariusza należy narysować i ponumerować kolumny, pamiętając o tym, że gdy dojdziemy

³ Gdy zmienna zapisywana jest w dwu lub większej ilości kolumn, dodaje się liczbę odpowiadającą numerowi pierwszej kolumny.

do kolumny nr 80, dalej numerujemy kolumny od 1 do 80 itd. Należy też zarezerwować miejsce na zapis danych identyfikacyjnych.

Kodowanie w kwestionariuszu ma swoje zalety: oszczędza czas, obniża koszty badań, zmniejsza ilość dokumentów a tym samym możliwość popełnienia błędów przy przenoszeniu informacji z kwestionariusza na karty kodowe, łatwiejszy jest powrót do "źródeł", gdy trzeba coś poprawić. Wyklucza to w zasadzie podwójny niezależny kodowanie, chyba że zastosuje się taki wariant postępowania, w którym pierwsze kodowanie wykona się na kartach kodowych, a drugie w kwestionariuszu.

Przy kodowaniu jednokrotnym w kwestionariuszu w oparciu o instrukcję przygotowaną "z góry", kodowanie można włączyć w zakres obowiązków ankietera.

IX. Kodowanie danych

Kodowanie obejmuje następujące czynności:

- 1) odczytanie zapisanej w kwestionariuszu odpowiedzi respondenta na dane pytanie,
- 2) zaliczenie jej do jednej i tylko jednej kategorii klasyfikacyjnej wyróżnionej w instrukcji kodowej,
- 3) zapisanie w odpowiednim miejscu karty kodowej znaku numerycznego symbolizującego przynależność danej odpowiedzi do danej kategorii klasyfikacyjnej.

Przypomnieć należy, iż zmienne wielokolumnowe zapełnia się od końca, np. zmienna nr 326 przyjmuje wartości od 0 do 14. Jest w związku z tym zapisywana w dwóch kolumnach karty kodowej: w kol. 26 i w kol. 27. Załóżmy, iż przyjmuje ona wartość 7. Zgodnie z podaną regułą powinna być zapisana w kol. 27, kol. 26 może pozostać pusta, może być uzupełniona "0" lub jakimś innym znakiem ustalonym z programistami. Mniej błędów popełnia się, jeśli wszystkie kolumny są zapełnione.

Kodowanie danych badacz z reguły zleca innym osobom, które do wykonywania tej czynności należy specjalnie przygotować. Każdej z nich badacz powinien dostarczyć kilka wypełnionych kwestionariuszy, instrukcję i karty kodowe. Materiały te wykorzystane będą w trakcie szkolenia, które może mieć charakter zarówno teoretyczny jak i praktyczny. Szkolenie teoretyczne polega na zapoznaniu ko-

derów z kwestionariuszem, instrukcją i kartą kodową. Jeśli instrukcja kodowa jest tak przygotowana, że działa "samoczynnie", szkolenie teoretyczne można pominąć. Szkolenie praktyczne polega na próbnym kodowaniu kwestionariuszy. Należy nim objąć wszystkich koderów. Ważne jest dokładne sprawdzenie efektów tego kodowania, a w przypadku stwierdzenia błędów - wyjaśnienie koderom, na czym one polegają.

Kodowanie danych może odbywać się w trybie grupowym (taśmowym), bądź indywidualnym. W pierwszym przypadku koderów gromadzi się w jednym miejscu i tak organizuje pracę, by jedna osoba kodowała zawsze odpowiedzi na te same pytania. W drugim przypadku jedna osoba koduje cały kwestionariusz.

Materiał może być kodowany jedno- lub dwukrotnie. W pierwszym przypadku każdy kwestionariusz kodowany jest tylko raz, natomiast w drugim przypadku jest on kodowany dwa razy. Przy systemie dwukrotnego kodowania dla każdego kwestionariusza należy przygotować dwa komplety kart kodowych. Warunkiem sensowności podwójnego kodowania jest zapewnienie mu niezależności. Oznacza to takie zorganizowanie procesu kodowania, by osoby klasyfikujące ten sam materiał nie mogły uzgadniać podejmowanych decyzji.

Podwójne kodowanie materiału jest szczególnie wskazane w badaniach, w których klasyfikowanie odpowiedzi wiąże się z ich interpretacją (a więc wtedy, gdy w kwestionariuszu jest dużo pytań otwartych z zapisem rejestrującym).

X. Kontrola zbioru danych pierwotnych

Z kart kodowych dane przenoszone są na karty perforowane, a następnie zapisywane są na taśmie magnetycznej lub dysku magnetycznym. Zostaje w ten sposób założony zbiór danych pierwotnych. Przy kodowaniu, perforacji i wprowadzaniu danych do pamięci komputera powstają zazwyczaj liczne błędy, zarówno przypadkowe, jak i systematyczne. Błędy te mogą polegać na wystąpieniu: 1) złego pojedynczego znaku mieszczącego się w zakresie danej zmiennej, 2) złego pojedynczego znaku spoza zakresu danej zmiennej, 3) zespołu złych znaków mieszczących się w zakresach danych zmiennych, 4) zespołu złych znaków spoza zakresów danych zmiennych (dwa ostatnie rodzaje to tzw. błędy przesunięć).

Źródłem tych błędów może być:

- różny sposób kodowania tego samego materiału (błędy koder-
skie),
- złe wyperforowanie danych (błędy perforacji),
- złe wczytanie przez czytnik kart dobrze wyperforowanych da-
nych (błędy przekłamań).

Ponieważ błędów może być bardzo dużo, istnieje konieczność przeprowadzenia kontroli założonego zbioru danych. Jeśli kodowa-
liśmy materiał podwójnie, zakładamy dwa zbiory danych, a czynno-
ści kontrolne realizujemy w trzech etapach.

W pierwszym etapie porównujemy analogiczne elementy każdego z
tych zbiorów. Zadanie to realizuje komputer zaprogramowany na
sygnalizowanie wszelkich rozbieżności między poszczególnymi ele-
mentami zbiorów. Gdy błędów jest dużo, poprawki wprowadza się do
komputera poprzez wymianę całych kart perforowanych (błędy koder-
skie poprawia się również na kartach kodowych). Jeśli jest ich ma-
ło, bardziej opłacalne jest sporządzenie listy poprawek, gdyż
nie wiąże się to z perforowaniem całych poprawionych kart. Wzór
takiej listy zamieszczamy w tab. 1.

T a b e l a 1

Lista poprawek nr ... dotyczy zbioru danych nr ...

Nr zmiennej pierwotnej	Nr badanego	Symbole kodu na taśmie magnetycznej	
		jest	powinno być
104	1	5	3
104	8	3	1
175	9	puste miejsce	x
180	9	7	8

Przy uruchomieniu odpowiednich programów można dokonać korekty
bezpośrednio w pamięci komputera, wymieniając błędne dane na
prawdziwe, co więcej, można skontrolować czy pamięć komputera
przyjęła te poprawki. Poprawia się tak długo, aż uzyska się peł-
ną zgodność między poszczególnymi elementami w zbiorze I i II.
Następnie eliminuje się z pamięci komputera jeden z tych zbiorów.

W drugim etapie eliminuje się te błędy, które nie mogły być
poprawione poprzednio. Kontrola jest tu ukierunkowana na wykrycie

znaków nie mieszczących się w zakresie poszczególnych zmiennych. Kontrolę tę wykona komputer pod warunkiem, iż programistom poda się informacje o zakresach zmiennych. Najwygodniej jest sporządzić odpowiednią listę (tab. 2).

T a b e l a 2

Lista zmiennych pierwotnych

Nr zmiennej pierwotnej	Nr karty kodowej	Nr kolumny (kolumn)	Zakres zmiennej		Występujące znaki dokładane	Uwagi
			min.	max.		
101	1	1-4	1	2000		Nr jednostki badania
105	1	5	1	1		Nr karty kodowej
106	1	6	1	2		
107	1	7	0	1	x	
108	1	8-9	0	12	x	
110	1	10	0	2	x	
111	1	11	0	5	xy	
112	1	12	0	5	xy	nie wystąpi 3

Na tym etapie kontroli wykrywa się głównie, choć nie wyłącznie, błędy koderskie. Ich poprawienie wymaga powrotu do materiałów źródłowych. Poprawki wprowadzać można analogicznie jak w etapie I poprzez wymiany kart perforowanych bądź poprzez listę. Zawsze jednak należy sprawdzić, czy zostały one dobrze zarejestrowane w pamięci komputera.

Etap trzeci to tzw. logiczna kontrola poprawności kodowania. W wyniku tej kontroli eliminuje się błędy, które nie mogły być poprawione w etapie pierwszym. Jest ona szczególnie ważna, gdyż umożliwia poprawienie systematycznych błędów koderskich, jakie powstały wewnątrz zakresów poszczególnych zmiennych, głównie w efekcie złej interpretacji zapisanych odpowiedzi. Kontrola logiczna jest możliwa jeśli między zakodowanymi zmiennymi występują określone związki logiczne: mogą to być relacje implikacji bądź tożsamości. Rozważmy to zagadnienie na przykładzie.

Założmy, iż w badaniu uwzględniono 5 następujących zmiennych:

Zp 130 - liczba dzieci respondenta	- zakres zmiennej 0-5
Zp 131 - wiek najstarszego dziecka	- " " 1-40, Y-ND
Zp 132 - wiek drugiego z kolei dziecka	- jw.
Zp 133 - wiek trzeciego z kolei dziecka	- jw.
Zp 134 - wiek czwartego z kolei dziecka	- jw.

Nietrudno spostrzec, iż pomiędzy przedstawionymi zmiennymi występują dwie następujące relacje:

t o ż s a m o ś c i :

$$(Z_{p130} = 0) \equiv (Z_{p131} = Y \wedge Z_{p132} = Y \wedge Z_{p133} = Y \wedge Z_{p134} = Y)$$

- oznacza to, że jeśli respondent nie posiada dzieci to wszystkie zmienne informujące o wieku tych dzieci powinny - zgodnie z przyjętą w kodowaniu symbolizacją - przyjmować wartość "Y" i odwrotnie, jeśli wszystkie zmienne informujące o wieku dzieci przyjmują wartość "Y" to zmienna określająca ich ilość powinna przyjmować wartość "0";

i m p l i k a c j i :

$(Z_{p131} > Z_{p134}) \rightarrow Z_{p130} = 4$ - oznacza to, że jeśli zmienna informująca o wieku najstarszego dziecka ma wartość wyższą od zmiennej informującej o wieku dziecka czwartego z kolei, to zmienna informująca o liczbie posiadanych dzieci powinna mieć wartość 4, ale nie odwrotnie. Może się bowiem zdarzyć tak, że w rodzinie są czworaczki.

Badacz powinien wychwycić możliwie dużo, a najlepiej wszystkie tego rodzaju związki logiczne między zmiennymi. Następnie powinien je opisać i przedstawić programistom, którzy uruchomią odpowiednie programy sprawdzające logiczną poprawność kodowania. Stwierdzone błędy będą następnie eliminowane w taki sam sposób jak w etapie pierwszym i drugim.

Przy pojedynczym kodowaniu kontrola założonego zbioru danych pierwotnych sprowadza się do sprawdzenia zakresów zmiennych i logicznej kontroli kodowania. Jest zatem słabsza, niż w przypadku kodowania podwójnego. Nie wykrywa się tu błędów koderskich wynikających z wątpliwości interpretacyjnych koderów. Zmniejsza się też przy tym trybie kontroli szansa na wychwycenie wszystkich błędów perforacji i błędów technicznych komputera.

XI. Transformowanie zmiennych

Gdy zbiór danych pierwotnych jest już "czysty", komputer wykonuje obliczenia celem uzyskania rozkładów zmiennych pierwotnych. Następnie przystępuje się do przekształcenia zmiennych pierwotnych w oparciu o reguły logiczne lub arytmetyczne, tam gdzie zostało to przewidziane przez badacza. W wyniku tych przekształceń powstaną nowe zmienne, nazywane zmiennymi transformowanymi. Dla zmiennych tych trzeba przewidzieć numery.

Należy w tym momencie odwołać się do koncepcji badań i sprawdzić, czy wszystkie zmienne potrzebne dla rozstrzygnięcia postawionych problemów znajdują się w zbiorze danych pierwotnych. Zazwyczaj jest tak - i badacz wie o tym od początku, sporządził wszak wcześniej listę zmiennych transformowanych - że część z tych zmiennych będzie dopiero teraz budowana przez komputer. Mogło się bowiem zdarzyć tak, że badacz chcąc uniknąć błędów w kodowaniu zrezygnował z łącznego kodowania odpowiedzi na kilka pytań kwestionariusza, zakodował je oddzielnie zdając sobie sprawę z tego, że dopiero po ich agregacji uzyska zmienną ważną dla rozstrzygnięcia problemu badawczego. Mogła wystąpić również taka sytuacja, że przy kodowaniu zmiennych ilościowych nie określił z góry przedziałów, dał polecenie aby przepisywać odpowiedzi z kwestionariusza, gdyż nie chciał podejmować decyzji arbitralnych. Teraz, znając rozkłady tych zmiennych będzie mógł podjąć bardziej trafne decyzje co do wielkości przedziałów.

Z rozkładami wszystkich zmiennych pierwotnych badacz powinien się bezwzględnie zapoznać, ponieważ ich analiza może go skłonić do innego, bardziej uzasadnionego ze statystycznego punktu widzenia, sposobu klasyfikowania zmiennej. Może mu też nasunąć niebrane dotąd pod uwagę sposoby opracowania materiału np. stworzenie skali, czy typologii.

Gdy badacz wie już dokładnie, jakie zmienne mają być budowane przez komputer, powinien przystąpić do opisu reguł rządzących przekształcaniem zmiennych pierwotnych w zmienne transformowane. Opis ten potrzebny jest programistom dla uruchomienia operacji transformowania zmiennych.

Aby bliżej zorientować czytelnika w tym zagadnieniu przedstawmy teraz, posługując się przykładami, zasadnicze typy transforma-

cji, zwracając uwagę na zapis reguł, które rządzą przekształcaniem zmiennych pierwotnych w odpowiednie zmienne transformowane.

A. Transformacja polegająca na zmniejszaniu liczby klas danej zmiennej pierwotnej

Założmy, iż zmienna pierwotna nr 540 opisująca poziom wykształcenia respondentów przyjmowała następujące wartości:

nie ukończone podstawowe	1
ukończone podstawowe	2
ukończone zasadnicze zawodowe	3
nie ukończone średnie	4
ukończone średnie	5
ukończona szkoła pomaturalna	6
nie ukończone wyższe	7
ukończone wyższe	8
brak danych	X

Analiza rozkładu tej zmiennej ujawniła, że pewne jej wartości w ogóle nie występują, inne zaś występują w niewielkiej liczbie. Skłoniło to badacza do takiego jej pogrupowania, by możliwe było wyróżnienie następujących klas wykształcenia:

nie wyższe niż podstawowe	1
wyższe niż podstawowe, ale co najwyżej średnie	2
wyższe niż średnie, ale co najwyżej nie ukończone wyższe	3
wyższe ukończone	4

Można to było osiągnąć przekształcając zmienną pierwotną nr 540 w nową zmienną transformowaną nr 581. Opis tej operacji polegającej na zmniejszeniu liczby klas zmiennej pierwotnej badacz przedstawił w formie tabeli (tab. 3).

T a b e l a 3

Nr zmiennych	Wartości zmiennych									
	1	2	3	4	5	6	7	8	X	
Zp 540	1	2	3	4	5	6	7	8	X	
Zt 581	1	1	2	2	2	3	3	4	X	

**B. Transformacja polegająca na tworzeniu klas
jakiejsz zmiennej mierzalnej
przy przyjęciu równych przedziałów**

Badano młodych robotników, zmienna nr 260 informowała o ich stażu pracy (z dokładnością do 1 miesiąca). Kodowanie tej informacji polegało na przepisaniu z kwestionariusza odpowiedniej liczby. Brak danych kodowano wpisując X. Badacz chciał następnie stworzyć zmienną transformowaną nr 290, która miała przyjmować następujące wartości: 1 - 1-6 miesięcy, 2 - 7-12 miesięcy, 3 - 13-18 miesięcy, 4 - 19-24 miesięcy, 5 - 25-30 miesięcy, 6 - 31-36 miesięcy itd. Przekazał programistom następującą informację:

Nr zmiennej transformowanej	Argument i reguła transformacji
Zt 290	Zp 260 grupowanie wartości \neq X co 6

**C. Transformacja polegająca na tworzeniu klas
jakiejsz zmiennej mierzalnej
przy przyjęciu nierównych przedziałów**

Badano średnie miesięczne dochody, jakimi dysponują młode małżeństwa. Zmienna informacja o wartości tych dochodów miała nr 261. Kodowanie danych polegało na przepisywaniu z kwestionariusza zadeklarowanych w wywiadzie kwot. Brak danych oznaczono symbolem "X". Po zapoznaniu się z rozkładem tej zmiennej badacz uznał za najbardziej sensowne przyjęcie następujących klas: 1 - do 4000, 2 - 4001-6000, 3 - 6001-7500, 4 - 7501-10000, 5 - 10001 i więcej. Wymagało to zbudowania zmiennej transformowanej nr 292. Przekazał programistom następujące polecenie

Nr zmiennej transformowanej	Argument i reguła transformacji
Zt 292	Zp 261 wartości \neq x pogrupować następująco: do 4000 - 1 4001-6000 - 2 6001-7500 - 3 7501-10000 - 4 10001 i więcej - 5

D. Transformacje polegające na agregowaniu
kilkun zmiennych pierwotnych

Np. b u d o w a t y p o l o g i i. W badaniach prowadzonych na wsi starano się opisać różne formy zaangażowania społecznego w sprawach lokalnych. W zbiorze danych pierwotnych znajdowały się 4 zmienne (nr 440, 441, 442, 443), które mogły posłużyć do stworzenia odpowiedniej typologii:

Zp 440 opisywała postrzeganie problemów występujących w środowisku lokalnym i przyjmowała następujące wartości:

R. nie widzi żadnego problemu	- 0
R. widzi 1 problem	- 1
R. widzi 2 problemy	- 2
R. widzi 3 problemy	- 3
R. widzi więcej niż 3 problemy	- 4
Brak danych	- X

Zp 441 określała działanie badanych w rozwiązywaniu postrzeganych przez siebie problemów i posiadała następujące wartości:

R. działał w rozwiązywaniu co najmniej jednego problemu	- 1
R. nie działał w rozwiązywaniu żadnego problemu	- 0
Brak danych	- X
Nie dotyczy (R. nie postrzega żadnych problemów)	- Y

Zp 442 określała udział badanych w czynach społecznych podejmowanych na rzecz środowiska lokalnego. Badacze wyróżnili następujące wartości tej zmiennej:

R. brał udział w czynach społecznych	- 1
R. nie brał udziału w czynach społecznych	- 2
Brak danych	- X

Zp 443 dotyczyła udziału badanych w zbiórkach pieniężnych organizowanych dla załatwienia konkretnych spraw ważnych dla środowiska lokalnego. Wartości tej zmiennej były następujące:

R. dawał pieniądze	- 1
R. nie dawał pieniędzy, ponieważ nie było takiej okazji	- 2
R. nie dawał pieniędzy, choć była taka okazja	- 3
Brak danych	- X

Zmienne te wykorzystano do budowy nowej zmiennej nr 499 opisującej 12 typów zaangażowania społecznego badanych w sprawach lokalnych.

W tabeli 4 przedstawiamy zarówno reguły przekształcenia tych 4 zmiennych pierwotnych w zmienną transformowaną, jak i merytoryczny opis wyróżnionych typów.

T a b e l a 4

Zmienne pierwotne				Zmienna transformowana
Zp 440	Zp 441	Zp 442	Zp 443	Zt 449
1v2v3v4	1	1	1	1 (R. postrzega sprawy, działa w sprawach, bierze udział w czynach społecznych i zbiórkach na cele społeczne)
1v2v3v4	1	1	3	2 (R. postrzega sprawy, działa w sprawach, bierze udział w czynach społecznych, ale nie bierze udziału w zbiórkach na cele społeczne)
1v2v3v4	1	2	1	3 (R. postrzega sprawy, działa w sprawach, bierze udział w zbiórkach, ale nie uczestniczy w czynach społecznych)
1v2v3v4	0	1	1	4 (R. postrzega sprawy, ale nie działa w sprawach przy równoczesnym uczestniczeniu w czynach społecznych i zbiórkach na cele społeczne)
1v2v3v4	1	2	3	5 (R. postrzega sprawy, działa w sprawach, ale nie uczestniczy ani w czynach społecznych ani w zbiórkach na cele społeczne)
1v2v3v4	0	2	1	6 (R. postrzega sprawy, ale w tych sprawach nie działa, równocześnie nie uczestniczy w czynach społecznych, choć daje pieniądze na cele społeczne)
1v2v3v4	0	1	3	7 (R. postrzega sprawy, ale w tych sprawach nie działa i równocześnie nie daje pieniędzy na cele społeczne choć uczestniczy w czynach społecznych)
0	Y	1	1	8 (R. nie widzi spraw i w związku z tym nie działa, ale bierze udział w czynach społecznych i składkach pieniężnych)
1v2v3v4	0	2	3	9 (R. widzi sprawy, ale w tych sprawach nie działa, nie bierze udziału w czynach ani w składkach pieniężnych na cele społeczne)
0	Y	1	3	10 (R. nie widzi spraw i dlatego nie działa, nie daje pieniędzy na cele społeczne ale bierze udział w czynach społecznych)

Tabela 4 (cd.)

Zmienne pierwotne				Zmienna transformowana
Zp 440	Zp 441	Zp 442	Zp 443	Zt 449
0	Y	2	1	11 (R. nie widzi spraw i dlatego nie działa, nie bierze udziału w czynach społecznych, ale daje pieniądze na cele społeczne)
0	Y	2	3	12 (R. nie widzi spraw i dlatego nie działa, nie bierze udziału w czynach społecznych, ani nie daje pieniędzy na cele społeczne)
Wszelkie inne układy				X (nie mają merytorycznego znaczenia)

E. Transformacje tworzone poprzez wzory arytmetyczne

Kilka zamieszczonych w kwestionariuszu pytań dotyczyło dochodów respondenta w głównym miejscu pracy. Odpowiedzi na każde z nich kodowano oddzielnie, przepisując do kart kodowych kwoty podawane przez respondentów. Brak danych kodowano jako "X" natomiast "nie dotyczy" jako "Y". Oto zmienne pierwotne oraz związane z nimi pytania:

Zmienna pierwotna nr 708: Jaka była przeciętna miesięczna wysokość P. zarobków podstawowych w ciągu ostatnich 12 miesięcy, tzn. od ... do ... Do otrzymanych na rękę kwot proszę dodać wysokość składek, spłacanych pożyczek oraz innych tego typu potrąceń przy wypłacie.

Zmienna pierwotna nr 709: Jeśli w ciągu ostatnich 12 miesięcy otrzymywał P. premię regulaminową, to jaka była w sumie wysokość tej premii? ...

Zmienna pierwotna nr 710: Jeśli w ciągu ostatnich 12 miesięcy otrzymywał P. deputaty, to jaka była łączna wysokość tych deputatów? ...

Zmienna pierwotna nr 711: Jaka była wysokość ostatnio otrzymanej 13-tej pensji? ...

Zmienna pierwotna nr 712: Jeśli w ciągu ostatnich 12 miesięcy otrzymywał P. nagrody (proszę tu wziąć pod uwagę również premie uznaniowe), to jaka była łączna wysokość tych nagród? ...

Badacza interesowały średnie miesięczne dochody respondenta w

głównym miejscu pracy. Sprawdził też staż pracy respondentów i okazało się, że wszyscy mieli dłuższy staż niż 1 rok. Regułę, w oparciu o którą zbudowano zmienną transformowaną "średnie miesięczne dochody respondenta" można zapisać w następujący sposób. Obliczeń nie wykonujemy, jeśli w którejkolwiek ze zmiennych od 708 do 712 wystąpi X, dla pozostałych przypadków obliczenia wykonuje się wg wzoru

$$\text{wartość zmiennnej 708} + \frac{\text{suma wartości } \neq Y \text{ z p 709, 710, 711, 712}}{12}$$

pierwotnej

Dotąd rozważaliśmy takie sytuacje, w których zmienne transformowane tworzyło się poprzez odpowiednie przekształcenia jednej lub kilku zmiennych pierwotnych. Nie są to sytuacje jedyne. Można budować zmienną transformowaną ze zmiennej (zmiennych) pierwotnej i zmiennej (zmiennych) uprzednio transformowanej. Można zatem mówić o trzech piętrach transformacji. Jest to ważne przy nadawaniu numerów zmiennym transformowanym, gdyż zmienna z wyższego piętra musi mieć zawsze numer wyższy. Programiści powinni przewidzieć jakiś system kontrolowania operacji transformowania zmiennych. Niezależnie od tego, badacz może przynajmniej wrywkowo sprawdzić, czy operacje te wykonano bezbłędnie. Prosi się programistów o listę (w ujęciu dla poszczególnych osób), w której były zapisane zarówno wartości zmiennych transformowanych jak i wartości wykorzystywanych przy tych transformacjach zmiennych pierwotnych. Wykorzystując zamieszczone w liście wartości zmiennych pierwotnych badacz dokonuje przekształcenia ich w zmienną transformowaną wg reguły, w oparciu o którą działał komputer, a następnie porównuje wynik tej operacji z wartością zmiennej transformowanej podanej na tej liście. Po stwierdzeniu, że wszystkie potrzebne transformacje są dobrze wykonane, zbiór danych można uznać za założony i "czysty".

XII. Wykonywanie obliczeń

Gdy zbiór danych jest już założony i "czysty", przystępuje się do wykonywania obliczeń na emc. Obliczenia te można wykonywać na całym zbiorze danych lub na jego podzbiorach, zależnie od potrzeb

merytorycznych. Jeśli chcemy uzyskać rozkłady na podzbiorach, programiści muszą te podzbiory założyć i sprawdzić, czy w trakcie wykonywania tej operacji nie wystąpiły jakieś błędy.

W pierwszej kolejności liczy się rozkłady zmiennych. Otrzymujemy tabulogramy zawierające zarówno liczby bezwzględne, jak i procenty. Jest wygodnie, gdy rozkłady procentowane są podwójnie: 1) od podstawy równoważnej liczbie wszystkich obserwacji, 2) od podstawy równoważnej liczbie obserwacji będących nośnikami treści merytorycznych (wyłącza się tu zatem "brak danych" i "nie dotyczy"). Trzeba zwrócić na to uwagę programistom, którzy będą mogli spełnić naszą prośbę szybko i bez wysiłku, jeśli "brak danych" i "nie dotyczy" kodowane były w sposób ujednolicony, wykorzystując znaki dokładane (X , Y). Jeśli kodowano je wg jakiejś innej zasady (np. 0,9), też nie będzie to szczególnie kłopotliwe. Problem pojawi się wówczas, gdy w kodowaniu nie było żadnego ujednolicenia. Oczywiście i wtedy możliwe jest wyłączenie tych obserwacji, ale jest to bardzo czasochłonne i wiąże się z wykonywaniem dodatkowych transformacji. Jeśli jest to potrzebne, liczy się odpowiednie współczynniki charakteryzujące rozkłady, takie jak: średnia, mediana, dominanta, kwartyle itp.

Po merytorycznej analizie rozkładów przystępujemy do określenia swoich potrzeb w zakresie dalszych obliczeń. Należy odwołać się tu do założeń koncepcyjnych (problemów badawczych) i sporządzić "ślepe" tablice, wpisując w nie zarówno nazwy, numery zmiennych, jak i wartości każdej ze zmiennych wraz z odpowiadającymi im symbolami kodowymi. Tabele te następnie dzielimy na dwuzmiennowe i wielozmiennowe. Trzeba teraz przemyśleć sposób budowy tabel wielozmiennowych i odpowiednio "rozpisać" je na tabelę dwuzmiennową. Chcemy np. otrzymać tablicę trójdzielczą. Jedną z występujących w niej zmiennych służyła jako kryterium wyróżnienia podzbiorów A_1 i A_2 . Wystarczy wówczas zaplanować uzyskanie dwu tablic dwudzielczych (jednej na zbiorze A_1 , drugiej na zbiorze A_2), które następnie złożymy w tablicę trójdzielczą.

Inny sposób budowy tablic trójdzielczych polega na sformułowaniu odpowiedniego warunku np.: tworzyć tabele dwudzielcze ze zmiennych 158, 168 dla poszczególnych wartości jakiejś innej zmiennej (np. 105). Sporządzamy teraz odrębne listy tablic dwudzielczych dla całego zbioru danych i dla poszczególnych podzbiorów. W każdej z list dokonujemy podziału tablic wg ich wielkości:

4 x 9 i mniejsze, od 5 x 9 do 9 x 9, większe niż 9 x 9. Przy sporządzaniu tych list stosuje się następujący zapis: 121, 152 283, 221 itd., co czyta się: korelacja zmiennej 121 ze zmienną 152, korelacja zmiennej 283 ze zmienną 221 itd. Tabele z warunkami zamieszcza się na końcu każdej z tych list.

Tak przygotowane listy przekazujemy programistom. Należy też poinformować programistów o konieczności wyłączenia z tablic obserwacji nie będących nośnikami treści merytorycznych i wskazać podstawę ich opocentowania. Należy też omówić z programistami rodzaj opracowania statystycznego, tzn. określić współczynniki korelacyjne oraz inne miary statystyczne, które powinny być liczone przy poszczególnych grupach tablic.

XIII. Uwagi końcowe

W związku z przedstawionym tu schematem postępowania, obowiązującym przy opracowaniu danych kwestionariuszowych na komputerze, nasuwa się kilka refleksji ogólniejszych.

1. Wbrew pozorom opracowanie komputerowe wymaga od badacza znacznie większego osobistego zaangażowania w proces badawczy, niż to ma miejsce przy "ręcznym" opracowaniu danych. Badacz zmuszony jest do intelektualnego panowania nad całością prowadzonych prac, utrzymywania stałego kontaktu z ośrodkiem obliczeniowym oraz rozwiązywania szeregu nowych problemów organizacyjnych. Najważniejszym z nich jest ustalenie trybu poprawiania błędów, jakie powstają podczas kodowania i wprowadzania danych do pamięci komputera. Czynności te muszą być wykonywane możliwie szybko, najlepiej przez specjalnie przeszkolony zespół współpracowników, którego prace badacz winien osobiście nadzorować.

2. W środowisku socjologów panuje przekonanie, iż komputerowe opracowanie danych kwestionariuszowych umożliwia znacznie szybsze otrzymanie wyników badań, niż jest to możliwe przy tradycyjnym opracowaniu. Praktyka pokazuje, że tak nie jest. Zakładanie i kontrola zbioru danych - przynajmniej obecnie - zajmuje dużo czasu. Jest to jednak rekompensowane tym, że wprowadzone do pamięci komputera dane są w niewielkim stopniu obciążone błędami, a sam proces wykonywania obliczeń przebiega szybko i są to obliczenia bezbłędne. Wydaje się, że są to korzyści współmierne w stosunku

do włożonego przez badacza wysiłku, nawet jeśli zysk czasowy nie jest tak duży, jak się tego spodziewano.

3. Opracowanie komputerowe umożliwia ponadto pełniejszą analizę danych. Przy "ręcznym" opracowaniu trudno jest bowiem wykonać dużą ilość tablic, obliczyć dla tych samych zestawień czy tablic różne miary statystyczne itp. Ponieważ obliczenia wykonywane są etapowo, istnieje też możliwość zmian w koncepcji opracowania w zależności od wyników uzyskiwanych na wcześniejszych etapach.

I. Przybyłowska, M. Szymczak

PREPARING DATA FROM QUESTIONNAIRE SURVEYS
FOR THEIR COMPUTER PROCESSING

The article is in the form of an instruction text providing basic information about the computer processing of data obtained through questionnaire surveys. It may be useful for sociology students and junior researchers. It contains basic information about:

- concept of processing materials,
- building a data collection for EMC,
- control of the built data collection,
- transforming variables,
- compiling tabulations,
- estimating statistical coefficients,
- interpreting tables.

