

DOROTA HAMAN

*Instytut Warzywnictwa*

## PRZYGOTOWANIE ANKIET DO OPRACOWANIA NA ELEKTRONICZNYCH MASZYNACH CYFROWYCH (EMC)

Znaczna część badań wymaga opracowywania wyników oraz sporej ilości obliczeń, które prowadzone metodami tradycyjnymi wymagają dość długiego czasu, co często je dezaktualizuje. Proces ten można znacznie skrócić stosując nowoczesną technikę obliczeniową. Na takie właśnie problemy napotyka się stosując badania ankietowe tak często przeprowadzane w naukach rolniczych i dające nieraz jedyną możliwość zebrania potrzebnych materiałów.

Zasadniczą trudność w opracowaniu na EMC materiałów zebranych tą metodą, w przypadku przeprowadzania bardziej złożonych analiz, jak np. testowania statystycznego itd., stanowi wielka różnorodność maszyn oraz zbyt mała ilość maszyn o dużej, a przede wszystkim szybko dostępnej pamięci, niezbędnych przy przetwarzaniu dużej ilości danych.

W przypadku gdy bada się struktury zjawisk (oblicza średnie i procenty) często wystarczają maszyny analityczne, czyli maszyny o konstrukcji elektromechanicznej lub elektronicznej przeznaczone do prostego przetwarzania danych rejestrowanych na kartach perforowanych, np. urządzenie do sortowania i zliczania tych kart — sorter.

Karty perforowane są stosowane zarówno przy obliczeniach na dużych cyfrowych maszynach elektronicznych (komputerach), jak i przy obliczeniach za pomocą maszyn analitycznych. Tak więc w zależności od potrzeb można wykorzystywać do opracowania kart obydwa rodzaje maszyn (oczywiście w przypadku gdy system dziurkowania kart jest jednakowy dla obu maszyn).

Należy dodać, że maszyny analityczne coraz rzadziej są stosowane samodzielnie, tzn. w oderwaniu od dużej maszyny elektronicznej. Z reguły są one uzupełnieniem komputera, służącym do różnorodnych czynności, takich jak: sortowanie, przeliczanie, sprawdzanie kart itp. Obecnie unowocześnia się tylko te maszyny analityczne, które są potrzebne do przygotowywania danych na kartach dla dalszych obliczeń na EMC, co nie wyklucza możliwości dokonywania przy ich pomocy prostych zestawień liczbowych bez wykorzystywania głównej maszyny cyfrowej, której koszt pracy jest o wiele wyższy.

Jeżeli z góry planuje się obliczanie np. korelacji, regresji itp. wtedy opłacalne jest przeprowadzenie wszystkich obliczeń na EMC. Duże maszyny posiadają na ogół standardowe programy liczące korelacje, regresje i różne zestawienia, podczas gdy tego rodzaju obliczenia na maszynach analitycznych trwałyby znacznie dłużej i byłyby bardziej skomplikowane. Nawet w przypadku braku programów standardowych, a więc konieczności pisania całego programu, przy dużej różnorodności i ilości obliczeń opłaca się stosować EMC.

Do maszyny wprowadzane są dane liczbowe. Aby je otrzymać należy odpowiednio zakodować uzyskane z ankiet odpowiedzi. Mówiąc ogólnie kodaż, czyli opracowanie wyników w sposób czytelny dla EMC polega na przyporządkowaniu każdej z możliwych odpowiedzi na dane pytanie symbolu liczbowego i wpisanie go w odpowiednią rubrykę przeznaczoną na kodaż, bądź w specjalnie w tym celu przygotowane tabelki. Pociąga to za sobą konieczność starannego opracowania kwestionariusza ankietowego od strony formalnej, tak aby był łatwy do okodowania. Tak więc przyspieszenie i ułatwienie obliczeń przez współpracę z maszyną cyfrową (analityczną lub EMC) zwiększa równocześnie wymagania co do jakości danych, a więc tym samym do opracowania kwestionariusza ankietowego. Jakość wyników zależy w dużej mierze od konstrukcji zarówno merytorycznej jak i formalnej kwestionariusza. Zbudowanie dobrego kwestionariusza jest, wbrew pozorom, rzeczą trudną i wymagającą dużej wiedzy oraz praktyki w badanej dziedzinie.

Pierwszym krokiem po podjęciu decyzji o przeprowadzeniu badań ankietowych powinno być wstępne spotkanie matematyka-informatyka, zajmującego się stroną formalną kwestionariusza ze specjalistą danej dziedziny opracowującym kwestionariusz od strony merytorycznej.

Pytania zawarte w kwestionariuszu mogą być różnego rodzaju. Ogólnie można je podzielić na pytania otwarte pozostawiające odpowiadającemu pełną swobodę w zakresie sposobu i objętości odpowiedzi oraz na pytania zamknięte wymagające od odpowiadającego wyboru i wskazania właściwych, gotowych odpowiedzi. Wśród pytań zamkniętych rozróżniamy z kolei pytania: alternatywne — wymagające odpowiedzi tak, nie, kafeterie — dające możliwość wyboru z szeregu gotowych odpowiedzi, skale — pytania ustalające siłę natężenia postawy i inne. Z kwestionariusza, który w dalszym etapie ma być całkowicie opracowywany na EMC należy wykluczyć pytania otwarte, które są niemożliwe do okodowania, zastępując je w miarę możliwości pytaniami zamkniętymi.

Aby lepiej zobrazować problemy budowania kwestionariusza oraz kodowania posłużymy się przykładowymi pytaniami dotyczącymi ogródków działkowych.

Przypuśćmy, że działkowicz ma wymienić gatunki kwiatów jakie uprawia. Aby okodować takie pytanie trzeba przejrzeć wszystkie wypełnione ankiety, zorientować się ile gatunków kwiatów wystąpiło we wszystkich ankietach, a następnie okodować pytanie-kafeterię z liczbą alternatywnych odpowiedzi równą ilości występujących gatunków. Kodowanie takie jest niemal tak pracochłonne jak opracowanie pytania metodami tradycyjnymi. Dlatego też jeżeli przygotowujemy kwestionariusz, który ma być opracowywany na EMC pytanie takie budujemy jako pytanie — kafeterię, uwzględniając w nim bądź gatunki szczególnie interesujące przeprowadzającego ankietę, bądź gatunki najczęściej występujące na działkach, podciągając pozostałe pod wspólną rubrykę „inne kwiaty”. Istnieją jednak sytuacje, gdy pytanie otwarte jest niezbędne do uzyskania koniecznych informacji. Przykładowo: „Jaki przepis należy Pana(i) zdaniem wprowadzić do regulaminu ogródków działkowych”. Do takiego pytania ułożenie kodażu jest praktycznie niemożliwe, gdyż w każdej ankiecie natrafimy prawdopodobnie na inną odpowiedź. Jeżeli tego rodzaju pytanie jest niezbędne trzeba się liczyć z koniecznością opracowania tej części ankiety metodami tradycyjnymi.

Okodowanie dobrze przygotowanego kwestionariusza nie jest sprawą trudną, lecz czasochłonną. Im więcej możliwych odpowiedzi na dane pytanie tym więcej symboli liczbowych trzeba dla niego zarezerwować. Należy uwzględnić również przypadek gdy brakuje odpowiedzi. W takim wypadku można umówić się np. że w kodażu odpowiada jej np. —1.

Najłatwiejsze do okodowania są pytania alternatywne z odpowiedzią tak lub nie. Kodując je, najprościej jest przyporządkować odpowiedzi twierdzącej liczbę 1, a odpowiedzi przeczącej 0. Często gdy w odpowiedzi występuje cecha ilościowa zamiast 1 wpisuje się liczbę odpowiadającą ilości jednostek, co znacznie upraszcza porównywanie i zestawianie wyników w dalszym opracowywaniu.

Jeżeli w ankiecie występuje pytanie-kafeteria takie, że trzeba wybrać dokładnie jedną z odpowiedzi, lub też pytanie-skala o tej samej własności, wtedy każdej z odpowiedzi przyporządkowuje się inną liczbę i w zależności od tego, która z możliwości została wybrana wpisuje się tę liczbę w rubrykę przeznaczoną na kodaż.

Jeżeli istnieje możliwość wyboru jednocześnie kilku z podanych odpowiedzi na pytanie typu kafeteria, lub skala opłaca się wtedy rozbić pytania na szereg alternatywnych pytań cząstkowych.

Przykładowo rozpatrzmy pytanie-kafeterię dotyczące uprawy kwiatów na działce: Jakie kwiaty uprawia się na działce? Różne, astry, piwonie, konwalie, inne kwiaty. Kodując to pytanie powinno się je rozbić na pięć pytań alternatywnych. Czy uprawia się na działce róże? Jeżeli tak to jako kodażu należy użyć np. 1, jeżeli nie to wpisać 0. Czy na działce uprawia

się astry? Jeżeli tak, to należy wpisać 1, jeżeli nie to 0 itd. W ten sposób łatwo będzie potem obliczyć na ilu działkach są np. róże i astry lub jednocześnie róże, piwonie i konwalie. Chcąc uzyskać takie informacje na podstawie odpowiedzi na pytania bez stosowania powyższego rozbicia trzeba by było użyć do kodażu 31 różnych liczb, aby uwzględnić wszystkie kombinacje jedno-, dwu-, trzy-, cztero-, i pięcio-elementowe ze zbioru pięcioelementowego. Przy rozbiciu podanym wyżej otrzymuje się pięć pytań alternatywnych co znacznie upraszcza kodaż, choć powoduje pewne rozszerzenie zbioru danych wprowadzonych do maszyny. Rozszerzenie to jest jednak bez znaczenia, wobec prędkości pracy maszyny.

Wybór metody kodowania takich pytań uwarunkowany jest pojemnością pamięci maszyny i ilością kombinacji, jakie trzeba będzie rozpatrzyć nie stosując rozbicia. Jeżeli zmniejszenie ilości danych wprowadzanych do maszyn ma nawet niewielkie znaczenie, to warto pytanie o niewielkiej ilości możliwych kombinacji (w pytaniu dotyczącym kwiatów było ich 31 chociaż uwzględniliśmy tylko 5 grup) opracować bez rozbijania. Przy pytaniach obszerniejszych kodowanie bez rozbijania jest stanowczo zbyt skomplikowane i wymaga zastosowania w kodażu bardzo dużej ilości liczb.

Jeżeli w kwestionariuszu przygotowana jest do wypełniania tabelka dotycząca drzew owocowych na działce — np.

Gatunek	Ilość drzew ogółem	Ilość drzew owocujących	Zbiór kg			
Jabłonie	1	0	0	1	0	0
Grusze	3	2	20	3	2	20

to robiąc kodaż rozbija się taką tabelkę na szereg pytań w następujący sposób. Ile jest na działce jabłoni? Ile jest na działce jabłoni owocujących? Wielkość zbioru jabłek w kg? Ile jest na działce gruszy? Ile jest na działce gruszy owocujących? Wielkość zbioru gruszek w kg?

Jeżeli odpowiadający nie posiada na działce jabłoni wtedy w kodażu wszystkich pytań dotyczących jabłoni wystąpią zera. Jeżeli posiada jabłonie to jako kodażu w pytaniu pierwszym użyje się liczby odpowiadającej ilości jabłoni na działce, w pytaniu drugim ilości drzew jabłoni owocujących, a w pytaniu trzecim liczby kilogramów odpowiadającej wielkości zbioru. Analogicznie należy postąpić z pytaniami dotyczącymi gruszy.

W przypadku tak małej tabeli można jeszcze próbować okodować ją całościowo bez rozbijania, kodując poszczególne kombinacje jabłoni i gruszy, uwzględniając owocujące i nieowocujące drzewa oraz zbiory owoców, lecz przy obszerniejszej tabeli staje się to zbyt uciążliwe.



W podanym powyżej przykładzie odpowiadający posiada jedną jabłoń natomiast nie posiada jabłoni owocujących, a więc nie ma zbioru jabłek. Na jego działce są trzy grusze z tego dwie owocujące, które w danym roku dały 20 kg owocu.

Po prawej stronie tabelki można umieścić specjalne rubryki przeznaczone na kodaż, które czasem drukuje się już w kwestionariuszu ankietowym co znacznie ułatwia późniejsze zestawianie danych przy przygotowywaniu do przenoszenia na karty perforowane lub inne mechaniczne nośniki informacji .

Ponieważ koszt użytkowania maszyny cyfrowej jest bardzo wysoki, ogromne znaczenie ma dokładne zastanowienie się nad tym jakie informacje są niezbędne i co będzie wykorzystane w dalszej pracy, by nie zbierać zbyt szczegółowych danych, których wartość poznawcza nie jest współmierna z kosztami opracowania.

Przy każdym z pytań należy zwrócić uwagę na jednostki, w jakich uzyskujemy informację, aby po wykonaniu obliczeń można było je sensownie zinterpretować.

W przypadku korzystania z urządzeń teledacyjnych do centralnej maszyny cyfrowej (urządzenie do bezpośrednich transmisji danych do centralnego ośrodka obliczeniowego) lub współpracy z dużą maszyną cyfrową, gdzie wykonywanych jest kilka lub kilkanaście programów jednocześnie, należy pamiętać o tym, że program o dużej ilości danych będzie na ogół opracowywany dość długo. Wynika to bowiem z faktu, że przy równoległym uruchamianiu kilku programów priorytet z reguły posiadają programy o małej ilości danych, które można policzyć stosunkowo szybko.

Tych kilka przykładów ma za zadanie wskazać osobom posługującym się ankietami, że pozorne uproszczenie i skróty mogą prowadzić do daleko idących komplikacji w dalszym przetwarzaniu uzyskanych wyników przy użyciu nowoczesnych urządzeń liczących. Prawidłowa konstrukcja ankiety wymaga biegłej znajomości sposobu pracy i możliwości technicznych konkretnej, użytej w danym przypadku maszyny cyfrowej. Jest to możliwe jedynie przy ścisłej współpracy badacza z matematykiem wyspecjalizowanym w obsłudze maszyn cyfrowych i w przetwarzaniu danych, znającym jednocześnie dostatecznie dokładnie problematykę, której dotyczy opracowywana ankietka.