mgr Ryszard Hall
Zakład Metod Ilościowych, Wydział Ekonomii
Uniwersytet Rzeszowski

MS Office, jako środowisko budowy aplikacji „ADA" do analizy i raportowania wyników badań kwestionariuszowych

WPROWADZENIE

Niniejszy tekst stanowi prezentację utworzonej przez autora aplikacji dedykowanej do analizy danych pochodzących z badań pierwotnych, głównie badań kwestionariuszowych, stąd nazwa „ADA" to skrót od „Analiza Danych Ankietowych". Jej filozofia jest inna niż np. zaimplementowana w pakiecie STATISTICA, głównie ze względu na jej wąskie przeznaczenie dedykowane głównie do analizy struktury i zależności pomiędzy zmiennymi opartymi na niskich skalach (jakościowych). Bardziej przypomina ona oprogramowanie służące do analizy danych zaimplementowane w serwisie „ankietka.pl"[1]. W porównaniu do obu wymienionych programów, w aplikacji ADA bardziej złożona jest definicja struktur danych obejmująca np.: definiowanie kilku zmiennych opartych na tych samych danych, definicję wielopoziomowych pytań filtrujących, analizę wg zdefiniowanych przekrojów (podzbiorów danych). Umożliwia ona również automatyzację wykonania samej analizy (tabulację i wizualizację w postaci tabel wielodzielczych, wykresów oraz współczynników statystycznych), gdzie sam badacz definiuje w prosty sposób zakres analizy, a raport uzyskuje się za pomocą przysłowiowych „kilku kliknięć".

Opisywana aplikacja ewoluowała od wersji wykorzystującej deklaratywny język programowania w postaci formuł arkuszowych o dość złożonej budowie do aplikacji hybrydowej korzystającej z możliwości klasycznego środowiska MS Office oraz procedur języka Visual Basic for Applications. Zmiany te wymuszone zostały koniecznością polepszenia jej funkcjonalności, a to wymagało bardziej złożonego opisu analizowanych danych, co z kolei komplikowało samą analizę i wymagało bardziej zaawansowanych procedur. Konieczne zatem stało się przeniesienie większości procedur analitycznych do modułów języka VBA.

Głównym środowiskiem aplikacji stał się Microsoft Excel, gdzie gromadzi się dane, przeprowadza ich edycję, kontrolę poprawności, samą analizę oraz

[1] www.ankietka.pl.

raportowanie wyników. Wykorzystano również Microsoft Word jako narzędzie do przygotowania kwestionariusza ankiety (narzędzia badawczego) oraz eksportu słownika danych (tekstu pytań i kafeterii odpowiedzi) do arkuszy Excela. Rozwijając aplikację starano się stosować zasadę, że do procedur VBA przenosi się jedynie te funkcje, których implementacja za pomocą klasycznych formuł jest niemożliwa albo nieefektywna. Powodem takiego podejścia jest to, że procedury VBA nigdy nie dorównają szybkością działania nawet najbardziej złożonym megaformułom arkuszowym.

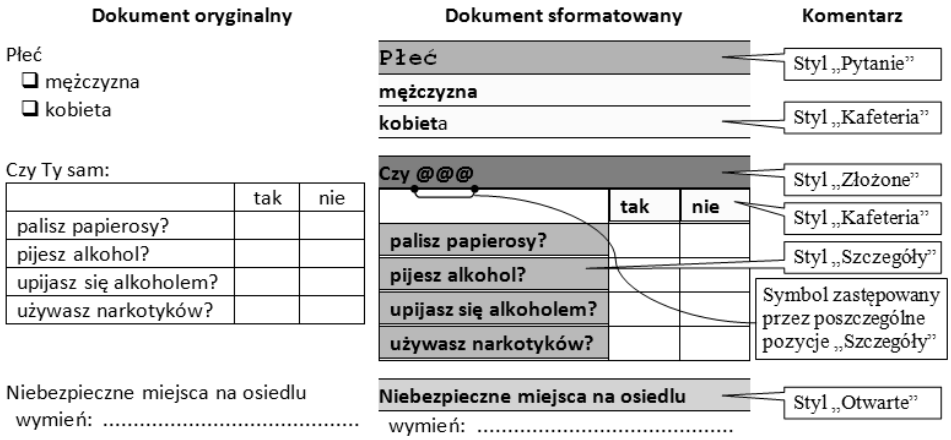
IMPORT DANYCH KWESTIONARIUSZA

Przeważnie kontakt analityka z badaniem prowadzonym przez badacza następuje już po jego fizycznym przeprowadzeniu, a więc dane najczęściej już są zgromadzone w postaci wypełnionych papierowych kwestionariuszy. Dane zgromadzone w takiej formie muszą zostać wprowadzone do arkuszy aplikacji i dotyczy to zarówno samych danych uzyskanych z badania, jak też z ich opisu słownego (czyli tekstu pytań i słownej interpretacji uzyskanych odpowiedzi zawartych w kwestionariuszu). Z praktyki autora wynika, że kwestionariusze najczęściej projektuje się za pomocą aplikacji MS Word, gdzie oprócz sformatowanego tekstu pytań i kafeterii odpowiedzi stosuje się różne symbole graficzne (np. ☐ w celu wpisaniu rangi), znaki specjalne (np. ❑ w celu zaznaczenia wariantu odpowiedzi), tabele i inne.

Niejednolitość sposobów formatowania i układu dokumentu MS Word oraz różnorodność zastosowanych w nim symboli powoduje, że trudno zautomatyzować proces przenoszenia tekstu z pliku MS Word (z ankiety) do arkuszy aplikacji ADA. Możliwe jest oczywiście „ręczne" kopiowanie poszczególnych fragmentów tekstu z pytaniami i wariantami odpowiedzi, jednak znaleziono rozwiązanie automatyzujące ten proces z zastosowaniem stylów i procedury VBA. Rys. 1 pokazuje na przykładzie trzech pytań, jak za pomocą stylów o ściśle określonej nazwie przygotować (formatować) dokument do eksportu słownika do arkusza MS Excel.

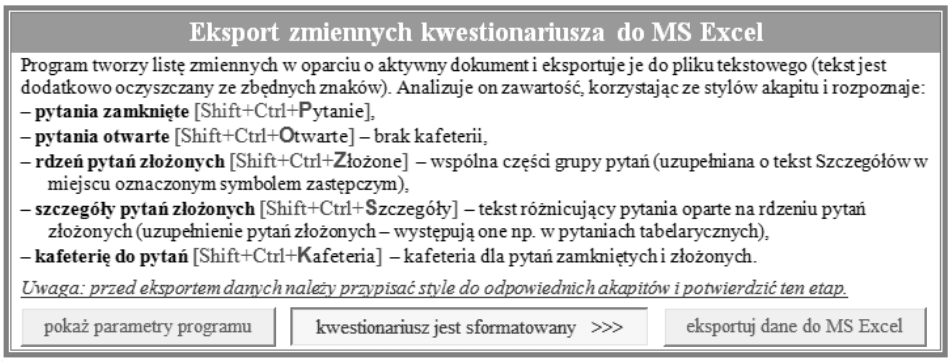
Kod procedury VBA tego modułu analizuje dokument kwestionariusza w MS Word i przenosi do arkusza MS Excel treści zmiennych i kafeterię odpowiedzi na podstawie opisujących je stylów. Mechanizm rozpoznaje zmienne kilku typów umieszczonych zarówno w akapitach, jak i tabelach. Tworzone są również zmienne złożone z połączenia wspólnego wzorca z różnicującymi je szczegółami i na rysunku 1 są to zmienne zawarte w tabeli.

Dodatkowo następuje oczyszczenia tekstu poprzez usuniecie niedozwolonych znaków jak: znaki specjalne, graficzne, wielokrotne spacje, itd. Szczegóły można zobaczyć wciskając przycisk „pokaż parametry programu (patrz rys. 2).



Rys. 1. Sposób przygotowanie kwestionariusza MS Word do eksportu danych do MS Excel

Źródło: opracowanie własne.



Rys. 2. Panel kontrolujący i uruchamiający eksport danych kwestionariusza do MS Excel

Źródło: opracowanie własne.

	A	B	C	D	E
1	**Nr**	**Pytanie**	**0**	**1**	**2**
2	1	Płeć	brak odpowiedzi	mężczyzna	kobieta
3	2	Czy palisz papierosy?	brak odpowiedzi	tak	nie
4	3	Czy pijesz alkohol?	brak odpowiedzi	tak	nie
5	4	Czy upijasz się alkoholem?	brak odpowiedzi	tak	nie
6	5	Czy używasz narkotyków?	brak odpowiedzi	tak	nie
7	6	Niebezpieczne miejsca na osiedlu	brak odpowiedzi		

Rys. 3. Fragment arkusza MS EXCEL z wyeksportowanymi danymi

Źródło: opracowanie własne.

Po przypisaniu stylów, za pomocą przycisku „eksportuj dane do MS Excel" (rys. 2) uruchamiamy z poziomu MS Word program, który umieści w nowym arkuszu MS Excel listę zmiennych o odpowiedniej dla aplikacji ADA strukturze (patrz rys. 3).

GROMADZENIE, EDYCJA I KONTROLA DANYCH

Wszystkie niezbędne dane przechowywane są w kilku arkuszach aplikacji ADA pracującej w środowisku MS Excel, z których każdy gromadzi informacje o innych właściwościach danych na różnych poziomach ich opisu. Są to:

- macierz,
- zmienne,
- słownik,
- filtry,
- przekroje,
- testy.

Opis poszczególnych struktur aplikacji ADA oparto o fikcyjny kwestionariusz ankiety złożony z 7 pytań merytorycznych i 3 merytorycznych (rys. 4). Opis rozpoczęty zostanie od arkusza „Słownik" chociaż wszystkie wyżej wymienione tworzą całość i definiuje się je we współzależności z pozostałymi.

P1. Czy brałeś udział w poprzednich wyborach?
❑ tak ❑ nie
P2. Z którą partią sympatyzujesz? (wybierz jedną)
❑ partia A ❑ partia B ❑ partia C
❑ partia D ❑ partia F
P3. Jeśli sympatyzujesz z partią B, to od kiedy?
❑ ostatni rok ❑ 1-2 lata
❑ 2-4 lata ❑ powyżej 4 lat
P4. Jeśli sympatyzujesz z partią A, C lub D to od kiedy?
❑ ostatni rok ❑ 1-2 lata
❑ 2-4 lata ❑ powyżej 4 lat
P6. W których miejsca na osiedlu, nie czujesz się bezpiecznie?
.....
.....

P5. Ponumeruj poniższe marki samochodów od najbardziej pożądanych „1" do najmniej pożądanych „6" (wybierz co najmniej 3 marki)
▭ Audi ▭ Volkswagen ▭ Peugeot
▭ Honda ▭ Toyota ▭ Skoda
P7. Do czego warto dopłacić kupując samochód?
❑ moc silnika ❑ rodzaj paliwa ❑ wielkość
❑ komfort ❑ bezpieczeństwo ❑ sylwetka

Metryczka

M1. Płeć: ❑ kobieta ❑ mężczyzna
M2. Wyksztalcenie
❑ podstawowe ❑ zasadnicze ❑ średnie ❑ wyższe
M3. Miesięczny dochód
❑ do 1 tys. ❑ 1-2 tys. ❑ 2-3 tys.
❑ 3-4 tys. ❑ powyżej 4 tys.

Rys. 4. Przykładowy kwestionariusz ankiety

Źródło: opracowanie własne.

Słownik

W uproszczeniu słownik to arkusz z tekstami poszczególnych pytań oraz skategoryzowane odpowiedzi na nie, zawarte w kwestionariuszu ankiety. Uproszczenie polega na tym, że do analizy używamy zmiennych, które niekoniecznie muszą mieć formę pytań, a dodatkowo na podstawie jednego pytania

z ankiety możemy utworzyć większą liczbę zmiennych. Podobnie kafeteria odpowiedzi na poszczególne pytania może się różnić od oryginalnej na kwestionariuszu, gdzie czasem zachodzi potrzeba jej przebudowy a posteriori.

	A	B	C	D	E	F	G	H	I
1	Nr	Pytanie	0	1	2	3	4	5	6
2	1	Udział w poprzednich wyborach	b.o.	tak	nie				
3	2	Sympatia polityczna	b.o.	partia A	partia B	partia C	partia D	partia F	
4	3	Okres sympatji z partią B	b.o.	ostatni rok	1-2 lata	2-4 lata	powyżej 4 lat		
5	4	Okres sympatji z lewicą (partie A, C, D)	b.o.	ostatni rok	1-2 lata	2-4 lata	powyżej 4 lat		
6	5	Rankink pożądanych marek samochodów	b.o.	Audi	Volkswagen	Peugeot	Honda	Toyota	Skoda
7	6	Niebezpieczne miejsca na osiedlu	b.o.						
8	7	Do czego warto doplacić kupując samochód	b.o.	moc silnika	rodzaj paliwa	wielkość	komfort	sylwetka	bezpieczeństwo
9	8	Płeć	b.o.	kobieta	mężczyzna				
10	9	Wykształcenie	b.o.	podstawowe	zasadnicze	średnie	wyższe		
11	10	Dochód miesięczny	b.o.	do 1 tys.	1-2 tys.	2-3 tys.	3-4 tys.	ponad 4 tys.	
12	11	Najbardziej pożądana marka samochodu	b.o.	Audi	Volkswagen	Peugeot	Honda	Toyota	Skoda
13	12	Trzy najpopularniejsze marki samochów	b.o.	Audi	Volkswagen	Peugeot	Honda	Toyota	Skoda

Rys. 5. Fragment arkusza „Słownik"
zawierający treści zmiennych oraz kafeterię odpowiedzi

Źródło: opracowanie własne.

Rys. 5 przedstawia fragment arkusza „Słownik", gdzie w kolejnych kolumnach mamy:

- Nr – jednoznaczny identyfikator zmiennej,
- Pytanie – treść zmiennej,
- 0 – tekst opisujący sytuację braku danych (braku odpowiedzi),
- 1…n – w kolejnych kolumnach umieszcza się kafeterię odpowiedzi.

Rys. 5 pokazuje zdefiniowanych 12 zmiennych, choć kwestionariusz przewidywał 10 pytań. Wynika to z tego, że na podstawie pytania P5 (patrz rys. 4 i 6) utworzono 3 zmienne o numerach 5, 11 i 12). Interpretacja tych dodatkowych zmiennych poczyniona zostanie przy omawianiu arkusza „Zmienne".

Arkusz „Słownik" posiada również mechanizmy kontroli poprawności danych i na rysunku widać, że nie została zamknięta zmienna nr 6 utworzona na podstawie pytania P6. Sygnalizowane jest to poprzez oznaczenie na czerwono numeru zmiennej oraz dwóch wymaganych pozycji kafeterii, gdyż każda skala pomiarowa musi przewidywać co najmniej dwie wartości cechy.

Macierz

To struktura przechowująca informacje o udzielonych przez respondentów odpowiedziach. Kod 0 (lub komórka pusta) oznacza brak odpowiedzi, a liczby 1…n numer pozycji w kafeterii zdefiniowana w arkuszu „Słownik". Nie da się tu określić, która kolumna związana jest z daną zmienną, a wynika to z tego, że (jak już wcześniej wspomniano) na tych samych kolumnach może być opisane kilka zmiennych ale i sytuacja, że określona kolumna nie jest związana z żadną zmienną, a służy jak źródło do uzyskania nowych jakościowo danych. W praktyce zdarza się, że kafeteria odpowiedzi jest zbyt rozproszona i należy pogrupować je w mniejszą liczbę kategorii. Można to uzyskać dodając do macierzy do-

datkową kolumnę, definiując w niej za pomocą klasycznej formuły MS Excel algorytm transformacji i tę nową kolumnę powiązać z definicją zmiennej.

Z tego powodu w nagłówkach kolumn (wiersz nr 1) brak identyfikatorów zmiennych a znajdują się tam przyjazne człowiekowi oznaczenia (P1, P2,…), które pozwalają raczej powiązać kolumny z pytaniami z ankiety. Jest to szczególnie przydatne podczas wprowadzania do arkusza, danych z papierowych kwestionariuszy. Wszystkie kolumny (wiersz 2) i kwestionariusze (kolumna A) są numerowane w celu uzyskania jednoznaczności przypisania danych. Arkusz posiada dodatkowe funkcjonalności, jak przyciski oprogramowane procedurami VBA umieszczone w obszarze komórek A1 i A2, służące do zarządzania strukturą i wyglądem kolumn macierzy. Procedury VBA wspomagają również poruszanie się między zmiennymi podczas edycji, a formatowanie warunkowe wykrywa błędy w definicji samej macierzy oraz błędy kodowania danych.

W przykładzie na rysunku 6 nagłówki kolumn 12 do 17 (komórki M2:R2 arkusza) oznaczone są w kolorze czerwonym, co oznacza, że żadna zmienna ich nie wykorzystuje. Powodem jest nieprzypisanie tych kolumn do żadnej zmiennej w arkuszu „Zmienne" (patrz rys. 7 i komentarz do niego). Tak, więc nawet jeśli kolumny te zawierałyby zakodowane dane, to żadna analiza aplikacji ADA ich nie widzi. Niekoniecznie musi to być błąd, bo jak wcześniej napisano, mogą to być kolumny z danymi surowymi służącymi do przeskalowania, a te przeskalowane przypisane do zmiennej.

Wykrywane są również błędy wprowadzenia wartości spoza skali i na rysunku 6 widać, że 2 wartości opisujące ankietę nr 4 (wiersz 6 arkusza) są nieprawidłowe. Aplikacja ADA sprawdza, do której zmiennej należy dana wartość, a następnie, czy wprowadzony kod jest dopuszczalny. Wartości 7 z komórki I6 należy do kolumny nr 8, a więc do zmiennej numer 5 (patrz rys. 7), a maksymalny kod tej zmiennej wynosi 6 (patrz rys. 5). Podobnie wartość 1 z komórki L6 przypisana jest do zmiennej nr 6, a ta nie ma określonej kafeterii odpowiedzi (patrz rys. 5). Błędy te są również sygnalizowane w arkuszu „Test" (patrz rys. 10 i komentarz do niego).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	✓	P1	P2	P3	P4	P5						P6	P7						M1	M2	M3
2	+	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	23
3	1	1	2	1		4	5	6											2	2	0
4	2	1	4		4	5	4	3	6	1	2								1	3	2
5	3	2	3		3	6	2	1	3										2	3	2
6	4	1	1		4	2	1	6	7			1							2	4	3
7	5	2	5			1	2												2	4	4
8	6	2	2	2		5	1	2											2	3	3
9	7	1	5			1	6	2	5										1	2	1

Rys. 6. Fragment arkusza „Macierz"
zawierający część zakodowanych kwestionariuszy

Źródło: opracowanie własne.

Zmienne

Tutaj następuje powiązanie kolumn macierzy (arkusz „Macierz") ze słownikiem (arkusz „Słownik") tworząc zmienne, jako podstawowe jednostki używane podczas analizy w aplikacji ADA. Ich definiowanie jest zautomatyzowane i wystarczy uzupełnić kolumny C i E (patrz rys. 7), a automatycznie, począwszy od kolumny F wykrywane i wpisywane są numery kolumn opisujące zmienne. W tym celu wykorzystywane są opisy zmiennych zamieszczonych zarówno w arkuszu „Zmienne" (kolumna C) oraz arkuszu „Macierz" (wiersz 1).

Można definiować wiele arkuszy zawierających zakodowane dane (kilka macierzy) i wówczas w kolumnie E należy podać nazwę arkusza, w którym znajdują się dana zmienna.

Można również zrezygnować z automatycznego wykrywania zmiennych i kolumny je opisujące wprowadzić samodzielnie. Na rysunku 7 są to definicje zmiennych 11 i 12. Nadano im nieistniejące w macierzy nazwy D1 i D2 i zdefiniowano, że opisują je kolumny 5, 6 i 7 dla zmiennej 11 oraz kolumna nr 5 dla zmiennej 12. Tak więc kolumna nr 5 macierzy używana jest w zmiennych: 5, 11 i 12.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	IV
1	Nr Zm.	Ilość kolumn	Opis Zmiennej	Max kod	Nazwa arkusza	Pierwsza kolumna	Kolejne kolumny															Wiersz w ark. Filtrów
2	1	1	P1	2	Macierz	1																
3	2	1	P2	5	Macierz	2																
4	3	1	P3	4	Macierz	3																2
5	4	1	P4	4	Macierz	4																3
6	5	6	P5	6	Macierz	5	6	7	8	9	10											
7	6	1	P6	0	Macierz	11																
8	7	0		6																		
9	8	1	M1	2	Macierz	18																
10	9	1	M2	4	Macierz	19																
11	10	1	M3	5	Macierz	23																
12	11	3	D1	6	Macierz	5	6	7														
13	12	1	D2	6	Macierz	5																

Rys. 7. Fragment arkusza „Zmienne" zawierający definicję zmiennych

Źródło: opracowanie własne.

Ponieważ pytanie, na którym je oparto, to prośba o porangowanie pożądanych marek samochodów, więc można teraz zbudować ranking na trzy sposoby:

- na podstawie zmiennej nr 5 (wszystkich 6 kolumn), licząc średnią ważoną, gdzie waga zależy od pozycji, na której daną markę wymieniono,
- na podstawie zmiennej nr 11, czyli na podstawie tych, które możemy nazwać pożądanymi markami (3 pierwsze pozycje), zakładając, że te z dalszych pozycji listy z założenia nie są pożądane, więc gdyby je uwzględnić w analizie, to jakby wybierać tę, która jest „mniej gorsza od innej",
- na podstawie zmiennej nr 12, czyli wyłącznie na podstawie marki najbardziej pożądanej, a więc tej, która jest wymieniana na pierwszym miejscu, więc wystarczy tylko zliczyć wskazania na poszczególne marki samochodów.

Teoretycznie analiza zmiennej nr 5 może dać taki wynik, że najbardziej pożądaną, jest marka, która nigdy nie była wymieniona na pierwszych trzech pozycjach. Może to mieć miejsce jeśli np. wszyscy respondenci wymienili pewną markę na 4. pozycji, ale pierwsze 3 były maksymalnie zróżnicowane. Analiza zmiennej nr 11 nie ma tej wady, choć nie można stwierdzić, że zawsze da ona lepsze wyniki niż zmienna 5. Obie wcześniej wymienione to zmienne koniunktywne (wielokrotnego wyboru), a więc i możliwości analizy statystycznej są mniejsze niż analizy zmiennej nr 12, która jest dysjunktywna (jednokrotny wybór).

Arkusz ten również ma mechanizmy sygnalizujące błędy i jeśli np. nie podano nazwy arkusza zawierającego „Macierz" to wszystkie komórki począwszy od kolumny F oznaczane są na czerwono (patrz zmienna nr 7 na rys. 7). Jeśli brakuje tylko oznaczenia zmiennej (kolumna C) to oznaczana kolorem jest tylko komórka w kolumnie F. Zawiera on również kilka kolumn informacyjnych, jak: ile kolumn przypisano do zmiennej, jaki jest maksymalny kod (rozpiętość kafeterii odpowiedzi) oraz czy, a jeśli tak to, w którym wierszu zdefiniowano filtr dla danej zmiennej. Tu filtrowane są zmienne nr 3 i 4 (istota filtrów opisana zostanie w części poświęconej arkuszowi „Filtry").

Filtry

Kwestionariusze często zawierają tzw. reguły przejścia, które kierują do konkretnego pytania, jeśli w innym udzielono określonej odpowiedzi albo nakazują pominięcie pewnego pytania. W przykładowym kwestionariuszu (patrz rys. 4) na pytanie P3 powinni udzielić odpowiedzi tylko ci respondenci, którzy zaznaczyli że sympatyzują z partią B. Zatem analizując zmienną nr 3 należy wybrać podzbiór rekordów (przypadków), których dotyczy. Liczebność przypadków dla tej zmiennej może być mniejsza niż liczebność całkowita. Jeśli zbudujemy tabelę korelacyjną w oparciu o dwie filtrowane zmienne, to liczebność przypadków poddanych analizie będzie częścią wspólną obu podzbiorów.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	**Zm** nr:	Zależy od: **Zm** filtrującej nr:	Nr kolumny arkusza **Zm** filtującej	Ilość warunków (lub warunek)	Dopuszczalne wartości zmiennej filtrującej:													
2	3	2	3	1	2													
3	4	2	3	1	1	3	4											
4																		
5																		

Rys. 8. Fragment arkusza „Filtry"
zawierający definicję 2 pytań podlegającym regułom przejścia

Źródło: opracowanie własne.

Sytuacja braku definicji filtrów przedstawiona zostanie na przykładzie próby 100-osobowej, z czego 50 osób to kobiety. Kobiety udzieliły odpowiedzi na pytanie, czy korzystały z jakiegoś produktu tylko kobietom dedykowanego. Jeśli pytanie będzie filtrowane i co druga kobieta potwierdza, że korzystała z tego

produktu, to struktura odpowiedzi to 50% „tak" i 50% „nie". Jeśli jednak nie będzie filtrowane (uwzględnieni zostaną mężczyźni) to 25% „tak", 25% „nie" i 50% „brak odpowiedzi". W sytuacji braku filtru i gdy wszyscy mężczyźni pomimo tego, że nie powinni, to udzielili odpowiedzi zaznaczając „tak", to otrzymamy jeszcze bardziej zniekształcone dane 75% „tak" i 25% „nie".

Definicje widoczne na rysunku 8 odnoszą się do zmiennej numer 3 i 4 (patrz rys. 5 i rys. 7), a więc ich liczebność może być mniejsza niż liczebność próby, gdyż przypadki, aby zostały uwzględnione muszą spełnić określone warunki:

- zmienna nr 3 „P3" (komórka A2=3) ma zdefiniowany jeden prosty warunek (komórka D2=1), gdzie dany przypadek może być uwzględniony w analizie jeśli zmienna nr 2 „P2" (komórka B2=2) przyjmuje wartość 2 „Partia B" (komórka E2=2),
- zmiennej nr 4 „P4" (komórka A3=4) ma zdefiniowany jeden warunek (komórka D3=1), gdzie dany przypadek może być uwzględniony w analizie, jeśli zmienna nr 2 „P2" (komórka B3=2) przyjmuje wartości 1, 3 lub 4 „Partia A, C lub D" (komórki E3:G3=1;3;4).

Definicja filtrów może być bardziej skomplikowana i opierać się na wartościach więcej niż jednej zmiennej (np. na dane pytanie odpowiadają mężczyźni w wieku powyżej 40 lat), ale zasada ich definiowania jest identyczna jak definiowanie przekrojów w arkuszu „Przekroje" i zostanie tam opisana.

Przekroje

Przekrój to w istocie podzbiór danych wyodrębniony na podstawie specyficznych cech (zbioru dopuszczalnych wartości) zapisanych w wybranych zmiennych. Zasada ich definiowania jest identyczna jak definiowanie filtrów z tym że filtr dotyczy tylko danej zmiennej i liczebność będzie mniejsza, jeśli analizujemy tylko tę zmienną, natomiast przekrój dotyczy wszystkich rekordów, tak jakbyśmy wyselekcjonowali pewien podzbiór danych, których dotyczą wszystkie analizy aplikacji ADA. Reszta przypadków traktowana jest jak nieistniejąca. Definiować można wiele przekrojów i jednym kliknięciem wybierać do analizy dowolny z nich.

Rys. 9 pokazuje cztery takie definicje, z czego pierwsza to prosta definicja, natomiast pozostałe trzy są bardziej złożone i używają operatorów logicznych łączących wartości kilku zmiennych. Definiuje się wszystkie kolumny z wyjątkiem kolumny D i tak:

- Przekrój nr 1 „Sympatyzujący z Partią C" – wybierany jest podzbiór przypadków, gdzie zmienna nr 2 „Sympatia polityczna" (komórka C2=2) przybiera wartość 3 „partia C" (komórka F2=3).
- Przekrój nr 2 „Kobiety o wykształceniu co najmniej średnim" – wybierany jest podzbiór przypadków spełniających jednocześnie 2 warunki (komórka E3=2), gdzie pierwszy z nich to, że zmienna nr 8 „Płeć" (komórka C3=8) ma przybrać wartość 1 „kobieta" (komórka F3=1) i jednocześnie (gdyż komórka E4=

”oraz”) zmienna nr 9 „Wykształcenie” (komórka C4=9) przybiera wartości 3 lub 4 „średnie lub wyższe” (komórka F4=3 a G4=4).

- Przekrój nr 3 „Mężczyźni nisko wykształceni sympatyzujący z partią B” – spełnione muszą być jednocześnie (E6:E7=”oraz”) 3 warunki (E5=3): zmienna nr 8 „Płeć” ma przybrać wartość 2 „mężczyzna”, zmienna nr 9 „Wykształcenie” wartości 1 lub 2 „podstawowe” lub „zasadnicze”, a zmienna 2 „Sympatia polityczna” wartość 2 „partia B”.
- Przekrój nr 4 „Osoby niewykształcone lub nisko zarabiające” – tu definicje oparte są na dwóch zmiennych: numer 9 „Wykształcenie” wybierająca wartość 1 „podstawowe” oraz numer 10 „Dochód miesięczny” przybierająca wartość 1 „do 1 tys. zł”. Aby dany rekord (ankietę) uwzględnić w analizie, wystarczy, że spełniony jest tylko jeden z nich, gdyż użyto operatora logicznego „lub” (komórka E9=”lub”).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Nr	Nazwa Przekroju	Zmienna przekroju	Nr kolumny Macierzy	Ilość warun. lub warunek	Dopuszczalne wartości zmiennej przekroju:									
2	1	Sympatyzujacy z Partią C	2	3	1	3									
3	2	Kobiety o wykształceniu co najmniej średnim	8	19	2	1									
4			9	20	oraz	3	4								
5	3	Mężczyźni nisko wykształceni sympatyzujący z parią B	8	19	3	2									
6			9	20	oraz	1	2								
7			2	3	oraz	2									
8	4	Osoby niewykształcone lub niskozarabiające	9	20	2	1									
9			10	21	lub	1									
10															
11															

Rys. 9. Fragment arkusza „Przekroje” zawierający definicję cztery podzbiorów danych

Źródło: opracowanie własne.

Testy

Jest to arkusz podsumowujący proces kodowania danych w macierzy. Można sprawdzić, czy nie ma w niej błędów wprowadzenia wartości spoza skali (kod którego nie ma w słowniku). Na rysunku 10 widać, że macierz nie jest jeszcze gotowa do przeprowadzenia analizy, gdyż zostały dwa błędy (komórka B13=2). Czerwone podświetlenie informuje, że jest to wartość wprowadzona do kolumny 8, czyli w pytaniu „P5” i jest to jedna ankieta (komórka I10=1) o numerze 4 (komórka I11=4) oraz jeden błąd w zmiennej „P6” w ankiecie o numerze 4. Obydwa te błędy widać w arkuszu „Macierz” (patrz rys. 6).

Tworzona jest również tabela „Rozkładów kodów”, która może posłużyć do ewentualnej decyzji o przeskalowaniu niektórych pytań (np. pogrupowania zbyt rozproszonych kategorii). Kolorem oznaczana jest rozpiętość skal pomiarowych i widać np., że kolumna nr 1, czyli pytanie „P1” ma 2 pozycje kafeterii i rozkład 40 osób brało udział w wyborach (kod równy 1) i 30 osób nie brało udziału w wyborach (kod równy 2). Wartość 1 w komórce L18 jest, jak widać, poza skalą, a wynika to stąd, że było to pytanie otwarte i nie zostało zamknięte (patrz rys. 5).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Ozn.	P1	P2	P3	P4	P5						P6	P7						M1	M2	M3
2		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
10	Ilość ankiet z błędem:	-	-	-	-	-	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-
11	Nr pierwszej ankiety z błędem:	-	-	-	-	-	-	-	4	-	-	4	-	-	-	-	-	-	-	-	-
13	**Razem błędów:**	2																			
15		**Rozkłady kodów**																			
16	99																				
17	0																				1
18	1	40	11	35		22	23	9		8		1							22		6
19	2	30	19	35		12	17	21			7								48	21	24
20	3		15		9			8	8											30	22
21	4		6		21	10	8													19	17
22	5		19			15	11		9												
23	6					11	11	22	13												

Rys. 10. Fragment arkusza „Test" zawierający raport błędów kodowania i rozkłady zmiennych

Źródło: opracowanie własne.

PREZENTACJA DANYCH (FORMATOWANIE I WIZUALIZACJA)

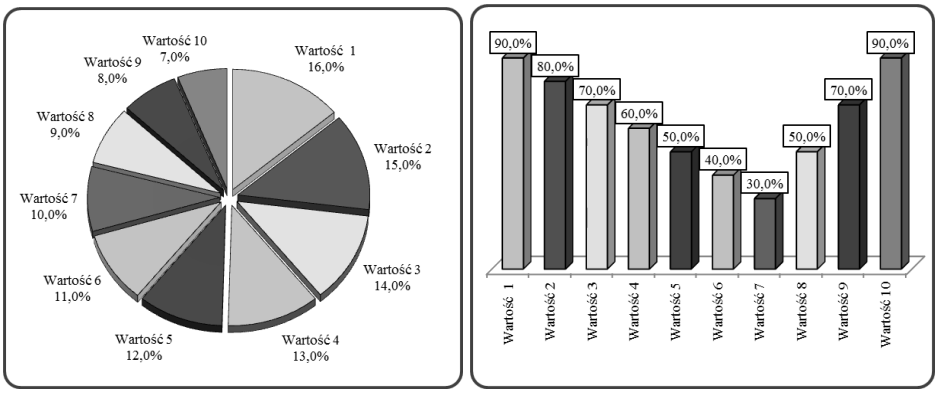
Sposób formatowania wygenerowanych wyników w postaci tabel i wykresów nie został zaimplementowany w kodzie VBA. Zdecydowano, że należy znaleźć taki sposób, aby każdy bez umiejętności programowania mógł utworzyć pożądany przez siebie wzór wykresu czy tabeli.

Wykresy

W przypadku wykresów zdecydowano się na najprostsze rozwiązanie, gdzie kopiowany będzie uprzednio przygotowany gotowy wykres, po czym zmodyfikowana zostanie jego właściwość „zakres danych", na podstawie których został utworzony. Zatem każdy może dostosować wzorzec wykresu, który użyty zostanie do wygenerowania raportu z wynikami.

Ponieważ innego typu wykresu należy użyć do wizualizacji danych dysjunktywnych, gdzie zmienną opisuje pojedyncza wartość, a innego typu wykresu dla zmiennych koniunktywnych, gdzie zmienną opisuje zbiór wartości – to należy przygotować dwa rodzaje wykresów. Dla zmiennych dysjunktywnych można wykorzystać wykres kołowy, gdyż dla wyznaczonej struktury, suma wszystkich wartości sumuje się zawsze do 100%. Nie nadaje się on dla zmiennych koniunktywnych, gdyż tu każda wartość może osiągnąć 100%, czyli wszyscy respondenci ją wskazali, ale wskazali również inne cechy dla tej zmiennej. W takim przpadku do prezentacji rozkładu takiej zmiennej należy użyć wykresu słupkowego lub kolumnowego.

Można przygotować dowolną liczbę wzorców, np. jeden dedykowany do wydruku w kolorze, a inny z wykorzystaniem szarych deseni do wydruku monochromatycznego. Przykładowe wykresy przedstawia rys. 11.



a) dla zmiennych dysjunktywnych (suma wartości równa się 100%)

b) dla zmiennych koniunktywnych (każdy słupek może osiągnąć 100%)

Rys. 11. Przykładowy wzór wykresów

Źródło: opracowanie własne.

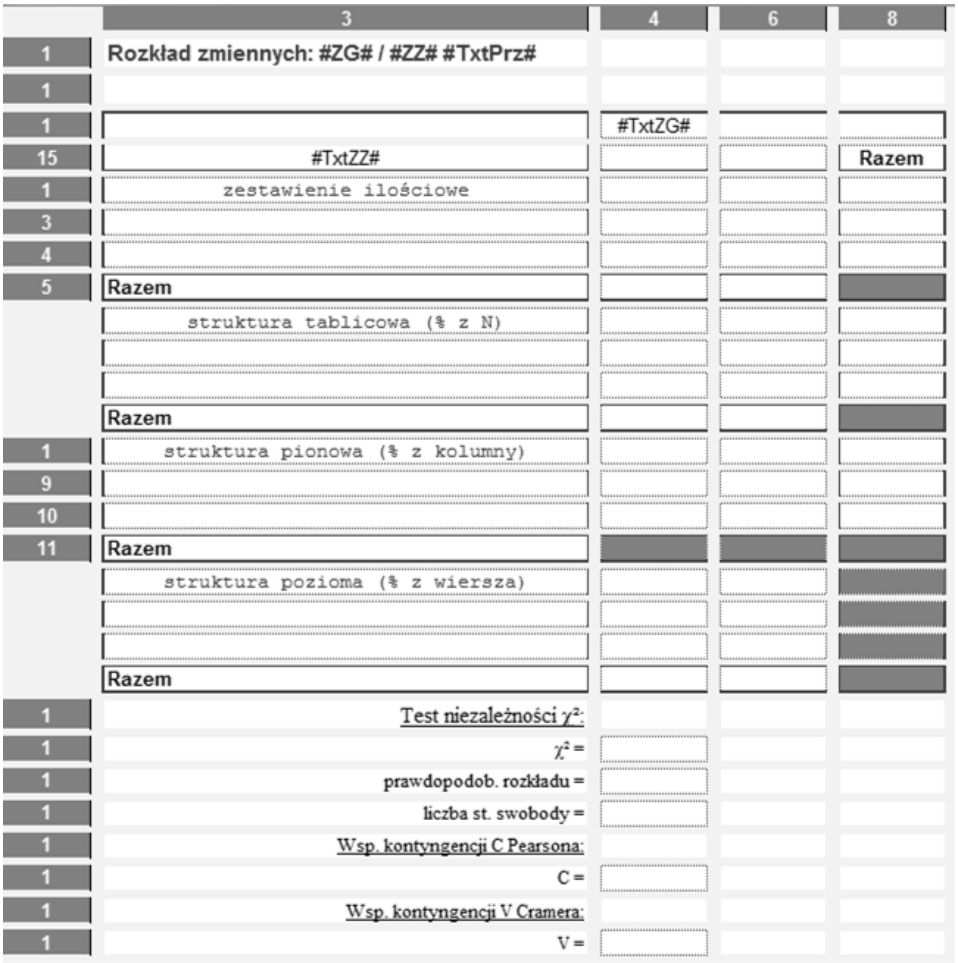
Tabele

Mechanizm definiujący strukturę tabeli zaprojektowano w ten sposób, aby można było w prosty sposób określać, jaki rodzaj danych tabela będzie zawierać. Na rysunku 12 przedstawiono przykładową definicję tabeli. Numery znajdujące się nad kolumnami oraz obok wierszy to definicje przeznaczenia poszczególnych komórek znajdujących się w środku na przecięciu danej kolumny i wiersza. Jeśli numer nie zostanie rozpoznany lub go brak, to taki wiersz czy kolumna nie znajdą się w tabeli wynikowej.

Nie będzie tu opisywany pełny leksykon z przeznaczeniem poszczególnych numerów, a tylko wybrane, na przykładzie których przybliżona zostanie zasada budowy tabel. Przykładowo, wszystkie komórki znajdujące się w wierszach oznaczone cyfrą 1 zostaną wiernie skopiowane do tabeli wynikowej (tak z założenia interpretowana jest wartość 1). Inne wartości mają specjalne znaczenie i tak wartość 3 definiuje wiersz dla zliczonych braków odpowiedzi zmiennej znajdującej się w boczku tabeli. Wartość 4 definiuje wiersz ze zliczonymi wartościami dla poszczególnych wariantów odpowiedzi zmiennej znajdującej się w boczku tabeli, a więc będzie powtarzany tyle razy, ile jest pozycji w kafeterii w arkuszu „Słownik". Takie samo znaczenie ma kolumna oznaczona wartością 6 i kolumna powtarzana będzie tyle razy, ile jest pozycji kafeterii, ale tym razem zmiennej umieszczonej w główce tabeli. Jeśli wygenerowalibyśmy dla takiej definicji tabelę, gdzie w główce tabeli umieszczono by płeć, a w boczku wykształcenie, to komórka oznaczona szustym numerem kolumny i czwartym numerem wiersza, powielona będzie osiem razy (2 odmiany płci i 4 warianty wykształcenia).

Definicja struktury tabeli polega zatem na wpisaniu odpowiednich numerów kolumn i wierszy z predefiniowanej listy. I tak wiersz oznaczony numerem 10

ma podobne znaczenie jak numer oznaczony wartością 4, z tym że służy do wprowadzenia wartości obliczonych jako ułamek danej wartości do sumy kolumny (procent z kolumny). Pomimo tego, że w definicji przygotowano część na strukturę poziomą, to nie znajdzie się ona w tabeli wynikowej, gdyż nie wprowadzono odpowiednich numerów wierszy, które je uaktywniają. Jeśli jednak obok tych wierszy wprowadzono by kolejno 1, 12, 13 i 14, to w raporcie w tabeli wynikowej pojawiłaby się sekcja z policzonymi odsetkami wg wierszy.



Rys. 12. Fragment arkusza z definicją struktury przykładowej tabeli rozkładu dwu zmiennych

Źródło: opracowanie własne.

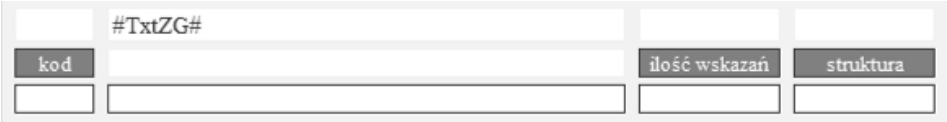
W definicji można umieszczać coś w rodzaju zmiennych, które zastąpione zostaną odpowiednimi wartościami podczas tworzenia gotowej tabeli. Zmienne

rozpoznawane są poprzez znaki „#" umieszczone na początku i końcu nazwy zmiennej. W opisywanym przykładzie użyto zmiennych o następującym przeznaczeniu:

- #ZG# – nazwa zmiennej umieszczonej w główce tabeli, ta którą wprowadzona do kolumny C arkusza „Zmienne" (patrz rys. 7),
- #ZZ# – nazwa zmiennej umieszczonej w boczku tabeli,
- #TxtZG# – treść zmiennej umieszczonej w główce tabeli, ta którą wprowadzona do kolumny B arkusza „Słownik" (patrz rys. 5),
- #TxtZZ# – treść zmiennej umieszczonej w boczku tabeli,
- #TxtPrz# – nazwa przekroju, wg którego analizowane są dane, zdefiniowana w kolumnie B arkusza „Przekroje" (patrz rys. 9), a wybranego w komórce A16 arkusza „Panel" (patrz rys. 9).

Wygląd (formatowanie) tabeli zdefiniowano poprzez użycie stylów. Każda elementarna komórka składająca się na tabelę ma przypisany styl o unikatowej nazwie i formatowanie tabeli wynikowej polega na modyfikacji dedykowanego poszczególnym komórkom stylu. Takie podejście umożliwia przygotowanie wielu wzorców formatowania wyników, ale również zmianę wyglądu już gotowych tabel poprzez scalenie (skopiowanie) do gotowego raportu, stylów z innego szablonu.

Dokładnie ta sama zasada dotyczy budowania i formatowania tabel prezentującej rozkłady pojedynczych zmiennych (przykład zamieszczono na rys. 13).



Rys. 13. Fragment arkusza z definicją struktury przykładowej tabeli rozkładu jednej zmiennej

Źródło: opracowanie własne.

ANALIZA DANYCH

Wszystkie mechanizmy analizy danych zgromadzonych w wyżej opisanych strukturach przeniesiono do modułów VBA. Żadne procedury napisane za pomocą tego języka nie będą tutaj opisywane, choćby dlatego, że aplikacja ADA składa się z kilkudziesięciu funkcji i procedur napisanych w tym języku, liczących w sumie około 1500 linii kodu programu, a jego wydruk zająłby 50 stron formatu A4.

Uwaga skoncentrowana zostanie zatem na interfejsie uruchamiającym poszczególne analizy i jego podstawowych funkcjonalnościach. Projektując aplikację włożono wiele wysiłku, aby generowanie wyników było maksymalnie proste i nie trzeba było każdorazowo ustawiać parametrów jego działania.

Sam skoroszyt, z programem i interfejsem nim sterującym, składa się z kilku arkuszy z których najważniejsze to:

- Definicje,
- Panel.

Definicje

Tu definiuje się, gdzie znajdują się dane, które będą analizowane i jaki ma być zakres analiz. W tabeli „Lokalizacja Danych" widocznej na rysunku 14 określa się, gdzie znajdują się poszczególne (wcześniej opisane) struktury danych, a więc nazwa skoroszytu oraz arkusza. Definiując lokalizację macierzy nie podaje się nazwy arkusza, gdyż definiuje się to w arkuszu „Zmienne". W momencie uruchomienia aplikacji wszystkie struktury danych muszą być dostępne a więc otwarte. Trzecia kolumna tej tabeli to test dostępności i jak widać, nie został odnaleziony skoroszyt „Wzór_KolorPełny".

Lokalizacja Danych	Plik (skoroszyt)	Arkusz	Dostępny?
Macierz	Przykład		Tak
Zmienne	Przykład	Zmienne	Tak
Słownik	Przykład	Słownik	Tak
Filtry	Przykład	Filtry	Tak
Przekroje	Przykład	Przekroje	Tak
Projekt Rozkładów 2D	Zakres analizy	Analiza_3	Tak
Wzorzec Formatowania Wyników	Wzór_KolorPełny		NIE
Lokalizacja Wyników	Wyniki-przykład		Tak

Generowanie Wyników Analiz	1 Wymiarowe		2 Wymiarowe	
Tworzenie Tablic Rozkładów	1	Tak	1	Tak
Sortowanie Wyników	1	Tak	0	Nie
Tworzenie Wykresów	1	Tak	0	Nie
Wyznaczanie Statystyk	0	Nie	1	Tak

Rys. 14. Fragment arkusza „Definicje" zawierający lokalizację struktur danych i zakres analiz

Źródło: opracowanie własne.

Przeznaczenie trzech dodatkowych, widocznych na rysunku 14, a nieopisanych wcześniej struktur jest następujące:

- Arkusz „Projekt rozkładów 2D" jest tablicą o strukturze kwadratowej tabeli, gdzie w nagłówkach kolumn i wierszy umieszczone są numer zmiennych a na przecięciu wartość 1 oznacza, że chcemy wygenerować tablicę korelacyjną i wyznaczyć statystyki dla tej pary zmiennych. Rys. 15 pokazuje, że utworzone zostaną 32 tabele korelacyjne, gdzie w główce gotowych tabel umieszczone będą poszczególne zmienne metryczkowe (zmienne 8, 9 i 10, 11 i 12) a w boczku poszczególne zmienne merytoryczne.
- „Wzorzec formatowania Wyników" to skoroszyt, w którym zdefiniowany jest wygląd gotowego raportu, a więc tabel i wykresów (patrz rysunki: 11, 12 i 13 i komentarze do nich).

- „Lokalizacja wyników” to skoroszyt, do którego raport z gotowymi tabelami i wykresami zostanie wprowadzony.

	1	2	3	4	5	6	7	8	9	10	11	12
1								1	1	1	1	1
2								1	1	1		
3								1	1	1		
4								1	1	1		
5								1	1	1		
6								1	1	1		
7								1	1	1		
8									1	1		
9										1		
10												
11								1	1	1		
12								1	1	1		

Rys. 15. Fragment arkusza ze zdefiniowanym zakresem analizy współzależności dwu zmiennych „Projekt rozkładów 2D”

Źródło: opracowanie własne.

W tabeli „Generowanie Wyników Analiz” (patrz rys. 14) metodą zero-jedynkową definiuje się jakie analizy (działania) zostaną wykonane, gdy analizować będziemy pojedyncze zmienne „1 Wymiarowe” (a więc powstanie tabela rozkładu jednej zmiennej) oraz gdy zmienne analizowane będą parami „2 Wymiarowe” (a więc powstanie tabela rozkładu 2 zmiennych). Tak więc, gdy utworzymy rozkład np. zmiennej nr 2 „sympatie polityczne”, to zgodnie z konfiguracją jak na rysunku 14, powstanie tabela ze zliczonymi liczebnościami, jej zawartość zostanie posortowana malejąco zaczynając od partii, z którą sympatyzuje największa liczba badanych, a następnie zostanie na jej podstawie zbudowany wykres. Sam układ, rodzaj informacji oraz wygląd tabeli oraz rodzaj i wygląd wykresu definiuje się w skoroszycie „Wzorzec formatowania Wyników”.

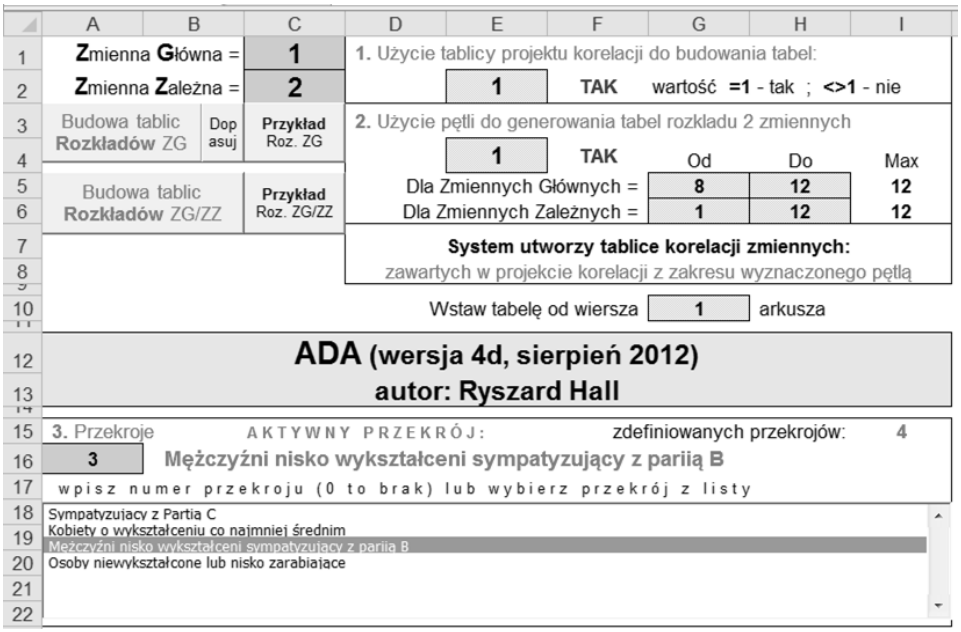
Panel

W arkuszu tym definiuje się rodzaj i zakres analizy i uruchamia sam proces tworzenia raportu. Na rysunku 16 widać panel z powiązanymi strukturami opisanymi w części „Gromadzenie, edycja i kontrola danych”. Aplikacje uruchomić można w trybie „Przykładu”, a więc po wprowadzeniu do komórki C1 numeru zmiennej i wciśnięciu przycisku „Przykład Roz. ZG” generowany jest w arkuszu „Przykład” wynik analizy (zgodnie z ustawieniami w tabeli „Generowanie Wyników Analiz” – patrz rys. 14). Podobnie wprowadzając dwie zmienne w komórkach C1 i C2 i wciskając przycisk „Przykład Roz. ZG/ZZ” otrzymujemy tabelę korelacyjna dla tych zmiennych. Arkusz „Przykład” zawsze przechowuje tylko jednę analizę, a więc można tego trybu używać korzystając z aplikacji w trybie interaktywnym.

Innym trybem jest wygenerowanie raportu do innych arkuszy dowolnego skoroszytu. Uruchamia się ją poprzez wciśnięcie przycisku „Budowa tablic Rozkładów ZG” dla analiz jednej zmiennej lub „Budowa tablic Rozkładów

ZG/ZZ" dla analiz współzależności dwu zmiennych. Wcześniej należy jednak wybrać kombinację ustawień komórek E2 i E4 (patrz rys. 16). Jeśli tylko komórka E2 zawiera wartość równą 1, to o tym, które zmienne zostaną wzięte do analizy decyduje arkusz „Projekt rozkładów 2D" (patrz rys. 15). Jeśli tylko komórka E4 zawiera wartość równą 1, to o zakresie decyduje pętla określona przez wartości komórek G5 i H5 dla analiz jednej zmiennej oraz dodatkowo komórek G6 i H6 dla współzależności dwu zmiennych. Jeśli w obu komórkach E2 i E4 jest wartość 1, to zakres determinuje „Projekt rozkładów 2D", ale tylko w zakresie ograniczonym przez zakres zmiennych, określony w komórkach G5:H6.

Dodatkowo w komórce A16 można wprowadzić numer przekroju, który zostanie poddany analizie (lub kliknąć dany przekrój na poniższej liście). Jeśli wartością komórki jest 0 (lub pusta, to analizowany jest pełny zakres danych. W konfiguracji jak na rysunku 16 analizować będziemy jedynie mężczyzn nisko wykształconych, którzy sympatyzują z partią B.



**Rys. 16. Fragment arkusza „Panel"
z powiązanymi strukturami danych**

Źródło: opracowanie własne.

Jeśli zakres analizy będzie taki jak na rysunku 16, to pomimo tego, że ustawienia w komórkach E4 oraz G5:H5 określają, iż wyniki generowane będą dla zmiennych M1, M2 i M3, D1 i D2 (zmienne od 8 do 12), to te 2 ostatnie wyklucza z analizy uaktywniona w komórce E2 definicja zakresu analizy w arkuszu „Projekt rozkładów 2D" (patrz rys. 15), gdzie ich nie uwzględniono. Gotowe

tabele wynikowe umieszczane są w arkuszach skoroszytu o nazwach takich jak zmienna tworząca główkę tabeli, więc jeśli będzie to taki zakres jak zdefiniowany na rysunku 15, to 9 tabel, gdzie zmienną główną jest zmienna nr 10 umieszczonych zostanie w arkuszu o nazwie „M1", 10 tabel dla zmiennej 11 w arkuszu „M2" i 11 tabel dla zmiennej 12 w arkuszu „M3".

Panel zawiera również komentarze pomagające ustawić parametry działania programu jak np.: jaki jest maksymalny numer zmiennej (komórki I5:I6), jaki jest zakres analizy (komentarz w obszarze D8:I8), wybrany (lub brak wyboru) przekrój danych (komórki B16:I16). Obszar w wierszach 12 i 13 to obszar informacyjny w którym podczas wykonywaniu analiz wyświetlany jest aktualny postęp obliczeń.

Przykładowy raport z analizy

Poniżej zaprezentowano kilka tabel analitycznych wygenerowanych za pomocą aplikacji ANK, wykonanych na postawie rzeczywistych badań kwestionariuszowych. Zaprezentowane w tabelach zmienne oraz liczebności nie mają nic wspólnego z przykładem użytym w poprzedniej części niniejszego tekstu.

Rozkład zmiennych: 25 (m1) / 8 (p5)

Opinie innych	Płeć			
	b.o.	kobieta	mężczyzna	**Razem**
zestawienie ilościowe				
b.o.	–	1	1	2
jednego lub obojga rodziców	5	74	25	**104**
rodzeństwa	2	49	19	**70**
znajomych, przyjaciół	2	135	71	**208**
nieznajomych osób (np..na forach)	–	17	6	**23**
kogoś inne	–	–	4	**4**
nie kierowałem się opiniami innych osób	5	51	29	**85**
Razem	**11**	**223**	**118**	N=352
struktura pionowa (% z kolumny)				
b.o.	–	0,4%	0,8%	0,6%
jednego lub obojga rodziców	45,5%	33,2%	21,2%	29,5%
rodzeństwa	18,2%	22,0%	16,1%	19,9%
znajomych, przyjaciół	18,2%	60,5%	60,2%	59,1%
nieznajomych osób (np..na forach)	–	7,6%	5,1%	6,5%
kogoś inne	–	–	3,4%	1,1%
nie kierowałem się opiniami innych osób	45,5%	22,9%	24,6%	24,1%
Razem	100,0%	100,0%	100,0%	100,0%

Test niezależności χ^2: *** ***NIE MOŻNA LICZYĆ (zmienna koniunktywna)***

Rys. 17. Przykładowa tabela korelacyjna zmiennej koniunktywnej i dysjunktywnej

Źródło: opracowanie własne.

Rys. 17 pokazuje tabelę automatycznie utworzoną przez aplikację ADA dla dwu zmiennych, z których jedna (płeć) jest zmienną dysjunktywną, a druga

(opinie innych) koniunktywną, mierzącą, czyimi opiniami kierowała się osoba badana. Mogła się ona kierować opiniami wielu osób, więc suma z kolumny „Razem" nie sumuje się do liczebności całkowitej liczby respondentów, czyli 352 (podobnie jak odsetki z kolumny „Razem" nie sumują się do 100%). Nie jest również wyznaczana statystyka χ^2 i mierniki na niej oparte, gdyż można to robić wtedy, gdy obydwie zmienne są jednokrotnego wyboru (mierzy je jedna wartość skali).

Rozkład zmiennych: 25 (m1) / 29 (d2)

Tok studiów	Płeć			
	b.o.	kobieta	mężczyzna	**Razem**
zestawienie ilościowe				
b.o.	–	–	–	–
dzienne	1	180	70	**251**
zaoczne	10	43	48	**101**
Razem	**11**	**223**	**118**	N=352
struktura pionowa (% z kolumny)				
b.o.	–	–	–	–
dzienne	9,1%	80,7%	59,3%	71,3%
zaoczne	90,9%	19,3%	40,7%	28,7%
Razem	100,0%	100,0%	100,0%	100,0%

Test niezależności χ^2:
wartość statystyki χ^2 = 18,055225
prawdopodob. rozkładu = 0,000021
liczba stopni swobody = 1
Wsp. kontyngencji C Pearsona:
C = 0,3171289

Rys. 18. Przykładowa tabela korelacyjna dla dwóch zmiennych dysjunktywnych, gdzie nie ma podstaw do odrzucenia hipotezy o niezależności tych zmiennych

Źródło: opracowanie własne.

Rys. 18 pokazuje tabelę korelacyjną, gdzie obie zmienne są dysjunktywne. Wyznaczona została statystyka χ^2, ale prawdopodobieństwo, że zmienne te są niezależne jest niewielkie, a więc można odrzucić hipotezę H0 o niezależności i przyjąć alternatywną H1, że istnieje zależność między zmiennymi. W takim przypadku można wyznaczyć statystyki określające siłę tej zależności, więc wyznaczony został współczynnika kontyngencji C Pearsona.

Kolejne dwie tabele pokazują, w jaki sposób działa filtr. Tabela na rysunku 19 pokazuje zależność pomiędzy płcią a informacją, czy respondent spotkał się z billboardem uczelni i jak widać w sumie 26 osób z 352 (7,4%) widziało billboard i pamięta, gdzie się on znajdował.

Jeśli kolejna zmienna o numerze 11 „Miejsce billboardu uczelni" mierząca, gdzie respondent widział billboard jest filtrowana wartością zmiennej 10 „Czy spotkał się z billboardem uczelni", to tabele oparte na tej zmiennej będą miały

mniejszą liczebność (patrz rys. 19). Tak więc rys. 20 pokazuje tabelę o liczebności N=26 osób, gdyż tylu respondentów pamięta, gdzie billboard widziało. Są to więc są te przypadki, gdzie wartością zmiennej 10 jest kod 1 „tak". Definicja taka musi być poczyniona w arkuszu „Filtry", więc wykorzystując pusty 4 wiersz na rys. 8 należałoby wprowadzić następujące wartości: A4=11, B4=10, D4=1, E4=1.

Rozkład zmiennych: 25 (m1) / 10 (p7a)

Czy spotkali się z bilbordem uczelni?	Płeć			
	b.o.	kobieta	mężczyzna	**Razem**
zestawienie ilościowe				
b.o.	1	1	2	4
tak	–	19	7	26
tak, ale nie pamiętam gdzie	6	94	55	155
nie	4	109	54	167
Razem	11	223	118	N=352
struktura pionowa (% z kolumny)				
b.o.	9,1%	0,4%	1,7%	1,1%
tak	–	8,5%	5,9%	7,4%
tak, ale nie pamiętam gdzie	54,5%	42,2%	46,6%	44,0%
nie	36,4%	48,9%	45,8%	47,4%
Razem	100,0%	100,0%	100,0%	100,0%

Test niezależności χ^2:
wartość statystyki χ^2 = 1,1780568
prawdopodob. rozkładu = 0,5548661
liczba stopni swobody = 2

Rys. 19. Przykładowa tabela korelacyjna zawierająca zmienną filtrującą

Źródło: opracowanie własne.

Rozkład zmiennych: 25 (m1) / 11 (p7a.t)

Miejsce bilbordu Uczelni	Płeć			
	b.o.	kobieta	mężczyzna	**Razem**
zestawienie ilościowe				
b.o.	–	–	–	–
obok budynku głównego	–	14	5	19
inne	–	5	2	7
Razem	–	19	7	N=26
struktura pionowa (% z kolumny)				
b.o.	–	–	–	–
obok budynku głównego	–	73,7%	71,4%	73,1%
inne	–	26,3%	28,6%	26,9%
Razem	–	100,0%	100,0%	100,0%

Rys. 20. Przykładowa tabela korelacyjna zawierająca zmienną filtrowaną

Źródło: opracowanie własne.

Zmienna może być również filtrowana wewnętrznie, a więc definicja filtru danej zmiennej odwołuje się do wartości tejże zmiennej. Rys. 21 pokazuje efekt definicji filtru dla zmiennej 47 „Kto w Twojej rodzinie zarabia więcej", który można by zdefiniować w arkuszu „Filtry" (patrz rys. 8) następująco: A5=47, B5=47, D5=1, E5=2, F5=3. Tak więc jeśli w jakiejkolwiek tabeli użyta będzie zmienna 47, to zostaną uwzględnione tylko te przypadki, gdzie wartością tej zmiennej jest 2 lub 3 (198 badanych osób znajduje się w związku formalnym lub nieformalnym). Eliminuje to błędne dane, udzielone przez respondentów w wyniku niezrozumienia pytania, czy niezastosowania się do reguł przejścia. Inna zatem będzie liczebność tabeli, struktura oraz wyznaczone współczynniki statystyczne dla zmiennej niefiltrowanej, gdyż oparte o dane, które w tabeli nie powinny się znaleźć.

Rozkład zmiennych: 61 (M1) / 47 (P36)

Kto w Twojej rodzinie zarabia więcej? (dotyczy osób w związku)	Stan cywilny						
	b.o.	panna	mężatka	konkubina	rozwie-dziona	wdowa	**Razem**
zestawienie ilościowe							
b.o.	–	–	15	1	–	–	**16**
ja	–	–	46	2	–	–	**48**
mąż (partner)	–	–	80	1	–	–	**81**
porównywalnie	–	–	51	2	–	–	**53**
Razem	–	–	**192**	**6**	–	–	N=198
struktura pionowa (% z kolumny)							
b.o.	–	–	7,8%	16,7%	–	–	8,1%
ja	–	–	24,0%	33,3%	–	–	24,2%
mąż (partner)	–	–	41,7%	16,7%	–	–	40,9%
porównywalnie	–	–	26,6%	33,3%	–	–	26,8%
Razem	–	–	100,0%	100,0%	–	–	100,0%

Test niezależności χ^2:
wartość statystyki χ^2 = 1,26463
prawdopodob. rozkładu = 0,53136
liczba stopni swobody = 2

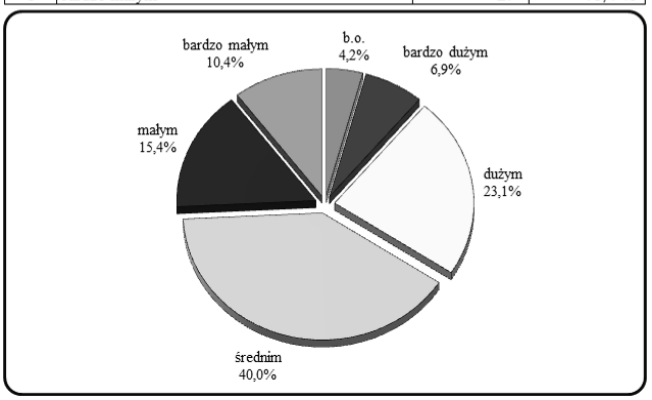
Rys. 21. Przykładowa tabela korelacyjna, zawierająca zmienną wewnętrznie filtrowaną

Źródło: opracowanie własne.

Kolejne dwie tabele przedstawiają rozkład pojedynczej zmiennej, gdzie dla analiz „1 Wymiarowych" (patrz rys. 14) uaktywniono tylko dwie opcje: tworzenie tabel rozkładów oraz tworzenie wykresów. Odsetki z kolumny „Struktura" sumują się do 100% tylko dla zmiennych dysjunktywnych (rys. 22), natomiast każda pozycja dla zmiennej koniunktywnej może przybrać 100%. Tak więc wszyscy respondenci mogli wziąć udział w każdym z wymienionych zjazdów (rys. 23). Inne są również typy wykresów dla obu rodzaju zmiennych (ich definicja opisana została w komentarzach do rysunku 11).

Jakim wydatkiem dla Twojego budżetu jest koszt uczestnictwa w Zjeździe? **N= 260**

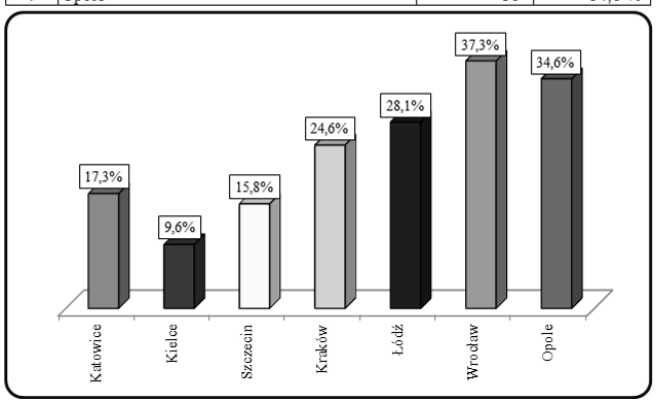
kod		ilość wskazań	struktura
0	b.o.	11	4,2 %
1	bardzo dużym	18	6,9 %
2	dużym	60	23,1 %
3	średnim	104	40,0 %
4	małym	40	15,4 %
5	bardzo małym	27	10,4 %



Rys. 22. Przykład tabeli analitycznej i wykresu dla zmiennej dysjunktywnej
Źródło: opracowanie własne.

W których zjazdach brałeś udział? **N= 260**

kod		ilość wskazań	struktura
0	b.o.	0	0,0 %
1	Katowice	45	17,3 %
2	Kielce	25	9,6 %
3	Szczecin	41	15,8 %
4	Kraków	64	24,6 %
5	Łódź	73	28,1 %
6	Wrocław	97	37,3 %
7	Opole	90	34,6 %



Rys. 23. Przykład tabeli analitycznej i wykresu dla zmiennej koniunktywnej
Źródło: opracowanie własne.

KIERUNKI ROZWOJU APLIKACJI

W zamierzeniu aplikacja ADA powinna umożliwiać wygenerowanie w prosty sposób raportu z analizą techniczną. Jakie analizy będą wykonane zależy od ustawień poczynionych w arkuszach „Definicje" oraz „Panel", jak również zdefiniowanych wzorców tabel i wykresów, opisanych w części „Prezentacja danych (formatowanie i wizualizacja)". Generalnie wybieramy gotowy wzorzec, wybieramy zakres analizowanych zmiennych i uruchamiamy aplikację. Taki automatyzm wymaga jednak wcześniejszego przygotowania danych (zdefiniowanie zmiennych, filtrów itp.) co zostało opisane w części „Gromadzenie, edycja i kontrola danych". Im bardziej precyzyjnie opisze się dane, tym bardziej precyzyjnie można dobrać narzędzia, które będą automatycznie zastosowane do ich analizy i w tym kierunku aplikacja ADA będzie rozwijana.

Wydaje się, że kolejnym logicznym kierunkiem rozwoju aplikacji powinno być zwiększenie zastosowanych testów i współczynników statystycznych. Aby jednak nie zmieniać filozofii aplikacji i za każdym razem nie wskazywać, jakie testy dla jakich zmiennych będą policzone, należy umożliwić automatyczny ich dobór przez aplikację. Wymaga to lepszego opisu zmiennych, gdyż w istocie w obecnej wersji rozpoznawane są zmienne dysjunktywne i koniunktywne, a do mierzenia współzależności zastosowane są testy nieparametryczne nadające się do skal każdego typu.

Dodanie w definicji zmiennych informacji o skali, na jakiej dokonano pomiaru jej wartości umożliwiłoby automatyczne zastosowanie różnych testów zależnie od typu zmiennych. Tak więc dodanie jednej kolumny w opisie zmiennej umożliwi rozróżnienie skal nominalnych od porządkowych, interwałowych oraz stosunkowych i zależnie od ich rodzaju pozwoli zdefiniować inny zakres testów statystycznych. Testy nieparametryczne będą wykonywane dla zmiennych opartych skalach nominalnych, dla porządkowych można będzie wyznaczać np. współczynnik rang Spearmana itd. Oczywiście przypisanie wykonywanych testów do poszczególnych skal będzie konfigurowalne.

LITERATURA

Bullen S., Bovey R., Green J., *Excel programowanie dla profesjonalistów*, Wydawnictwo HELION, Gliwice 2006.

Gibbs G., Gribbs G., *Analizowanie danych jakościowych*, PWN, Warszawa, listopad 2011.

Krzymowski B., *Microsoft Office. Podstawy programowania w języku Visual Basic dla Aplikacji*, Komputerowa Oficyna Wydawnicza „Help", Michałowice 2004.

Malarska A., *Statystyczna analiza danych wspomagana programem SPSS*, SPSS Polska, Kraków 2005.

Walkenbach J., *2003 PL Excel. Programowanie w VBA. Vademecum profesjonalisty*, Wydawnictwo HELION, Gliwice 2004.

Wątroba J., *Od surowych danych do profesjonalnego raportu – przykład kompletnej statystycznej analizy danych medycznych*, StatSoft Polska 2006.

www.ankietka.pl.

Streszczenie

Artykuł opisuje podstawowe funkcjonalności aplikacji ADA służącej do gromadzenia i analizy danych ankietowych. Zasadnicza praca twórcza włożona została w samo oprogramowanie napisane przez autora w języku VBA z wykorzystaniem środowiska MS Excel. Opisano, jakie struktury danych są gromadzone przez aplikację, interfejs uruchamiający poszczególne analizy, ich zakres oraz przykłady samych analiz.

MS Office, as an application development environment "ADA" for analyzing and reporting the results of questionnaires

Summary

The text describes the basic functionality of the application of ADA for the collection and analysis of survey data. The main creative work has been inserted in the software written by the author in VBA environment using MS Excel. Describes the structure of the data is collected by the application, the interface starts individual analysis, their scope and examples of the same analysis.