PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
Nr 86 — 2009
Wizualizacja wyników badań marketingowych – podejścia, metody i zastosowania

Jolanta Perek-Białas
Szkoła Główna Handlowa w Warszawie

RAPORTOWANIE WYNIKÓW BADAŃ – NAUCZANIE A PRAKTYKA

1. Wstęp

W celu przygotowania do pracy przyszłych badaczy prowadzone są na wyższych uczelniach różnego rodzaju zajęcia (z reguły o charakterze warsztatowym), które mają – także z założenia – za zadanie pokazać sztukę prezentacji wyników z różnego rodzaju badań, np. marketingowych czy rynkowych, i nauczyć jej. Usystematyzowaną wiedzę na temat, jak powinien wyglądać raport z badań sondażowych (w tym marketingowych), można znaleźć w wielu pozycjach literatury przedmiotu, m.in.: [Churchill 2002; Mazurek-Łopacińska 2005]. Dodatkowo prezentacje rezultatów zastosowania wielowymiarowej analizy danych są obecne w rozdziałach publikacji, które dotyczą zagadnień metod analizy danych, np.: [Walesiak 1996; Sagan 1998; Bazarnik i in. 1992; Gatnar, Walesiak 2004; Rószkiewicz 2002]. Jednak z reguły na zajęciach nie ma zwykle zarówno czasu, jak i możliwości, aby odpowiednio skonfrontować finalny raport z zastosowanymi różnymi metodami analizy danych – napisany według wytycznych podręczników – z oczekiwaniami i wymaganiami prawdziwego klienta. Studenci (szczególnie kierunków tzw. ilościowych) są nastawieni głównie na przekazywanie szczegółowych informacji dotyczących samych wyników, jak również oceny jakości modeli, które zbudowali, gdyż to potwierdza ich wiarygodność i rzetelność jako badaczy. Przekazują więc informacje dotyczące założeń zastosowania określonych rozwiązań modelowych, sposobu doboru zmiennych, określenia, które z nich są istotne statystycznie i czy ich model daje moc predykcyjną czy tylko wyjaśniającą. Z kolei w praktyce osoby odbierające raporty z badań (niekoniecznie zaznajomione z tajnikami analizy) nie są zainteresowane wszystkimi szczegółami „technicznymi", ale takim sposobem prezentacji wyników, który pozwoli im nie tylko na ich szybkie zrozumienie, ale też na ich przekazanie dalej – czyli do osób decyzyjnych w firmie – w sposób zrozumiały i jasny.

Celem niniejszego artykułu jest więc głównie pokazanie różnych możliwych prezentacji wyników, które są dostępne za pomocą programów statystycznych, ale dodatkowo uatrakcyjniając wizualizację wyników, przy zastosowaniu specjalnego

języka programowania. Od razu będzie to wskazanie, w jaki ciekawy sposób można pokazywać, wydawałoby się, proste, typowe i banalne wyniki. Pytania, które można zadać w kontekście tematyki opracowania, to:

- Na ile w ramach zajęć trzeba, a na ile można, wygospodarować czas na pokazanie roli i znaczenia dobrej wizualizacji wyników?
- Czy „ograniczona" wizualizacja oznacza brak kompetencji analitycznych badacza/analityka?
- Jak i gdzie znaleźć złoty środek między nauczaniem a praktyką, dla której wizualizacja decyduje też o przewadze konkurencyjnej firm badawczych?

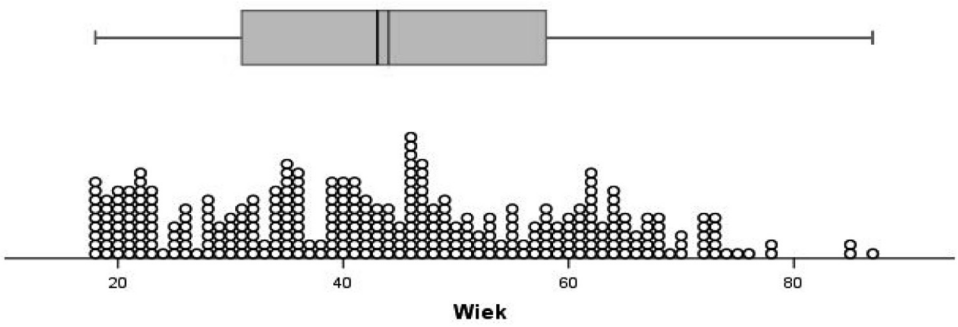
2. Przykłady prostych, a ciekawych rozwiązań

Dostępne oprogramowanie[1] umożliwia tworzenie różnego rodzaju wykresów, które dają możliwości łatwej i szybkiej oceny wyników. Z reguły jednak nawet takie proste wykresy, jak słupkowy, kołowy czy nawet rozrzutu są prezentowane oddzielnie, gdyż nie ma jednoczesnych funkcji prezentacji łącznie informacji na kilku wykresach w ramach jednej procedury. To nie oznacza, że nie można takich „innych" wykresów, jak zaprezentowano poniżej (tj. na rys. 1-5), utworzyć w ramach dostępnego programu statystycznego. Na przykład rys. 1 pokazuje nałożenie oprócz wykresu punktowego na wykres skrzynkowy także średniej obok mediany. Rysunek 2 pokazuje z kolei możliwość jednoczesnego zestawienia zagregowanych i niezagregowanych rozkładów tej samej zmiennej. Wszystkie odpowiedzi: *bardzo ufam*, *trochę ufam* zostały raz zliczone i zsumowane w jedną, podobnie: *niezbyt ufam*, *w ogóle nie ufam* w drugą, a w wariancie poniżej kategorie te zostały rozdzielone, dając możliwość szczegółowego wglądu w rozkład odpowiedzi respondentów. Z badań sondażowych wynika, że często standardem jest budowanie różnych tabel krzyżowych (z 2, 3 zmiennymi i więcej). Rysunek 3 jest przykładem prezentacji tabeli krzyżowej, gdzie można zobaczyć nie tylko rozkład w ramach jednej zmiennej „zaufania do banków", ale od razu widać różnice (lub ich brak) między kobietami i mężczyznami. Wykresy rozrzutu z histogramami i wykresy skrzynkowe (rys. 4 i 5) nie tylko pokazują, jaki charakter zależności mają dwie zmienne w analizie, ale też jednocześnie otrzymujemy informacje, czy są to zmienne o rozkładach symetrycznych, czy skośnych, ile wynosi mediana i jakie wartości (minimum, maksimum) przyjmują użyte zmienne.

Zaprezentowane poniżej wykresy są alternatywą do standardowych wizualizacji wyników, spełniając tym samym cel, aby w ramach jednego wykresu pokazywać dwie, trzy informacje możliwe do uzyskania w ramach kilku procedur. Jest to na pewno kierunek, który w przyszłości będzie chętnie wykorzystywany przez osoby przygotowujące raporty.

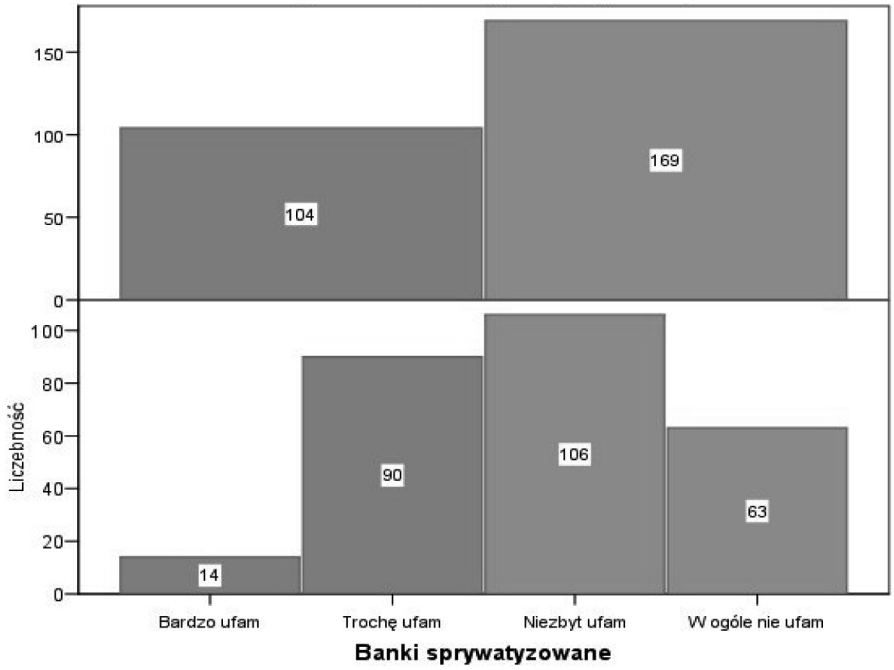
---

[1] W artykule wykorzystano PASW 17 – produkt firmy SPSS Inc. Materiały SPSS Polska zostały przekazane do prezentacji i publikacji przez Janusza Wachnickiego, za co składam podziękowania.



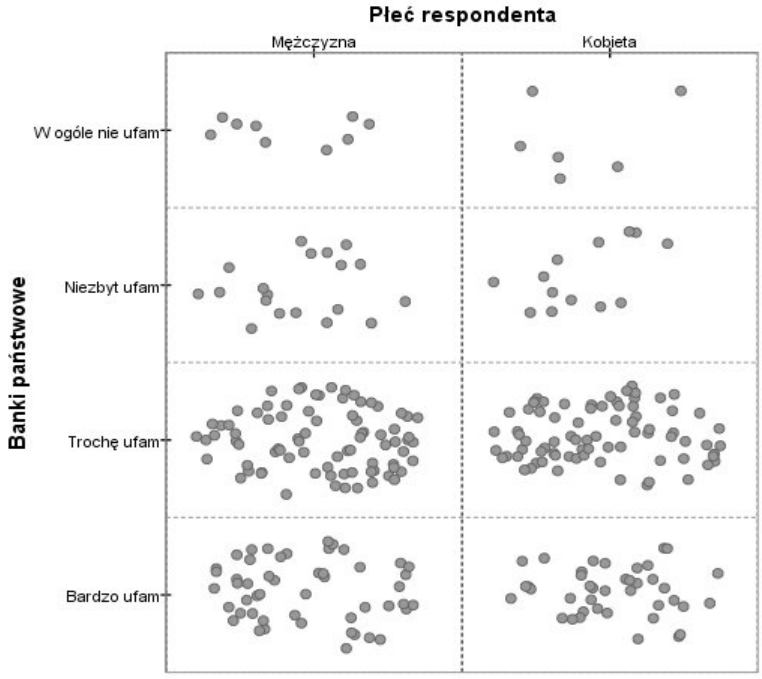
Rys. 1. Przykład wykresu skrzynkowego z zaznaczoną średnią i wykresem punktowym

Źródło: materiały SPSS Polska.



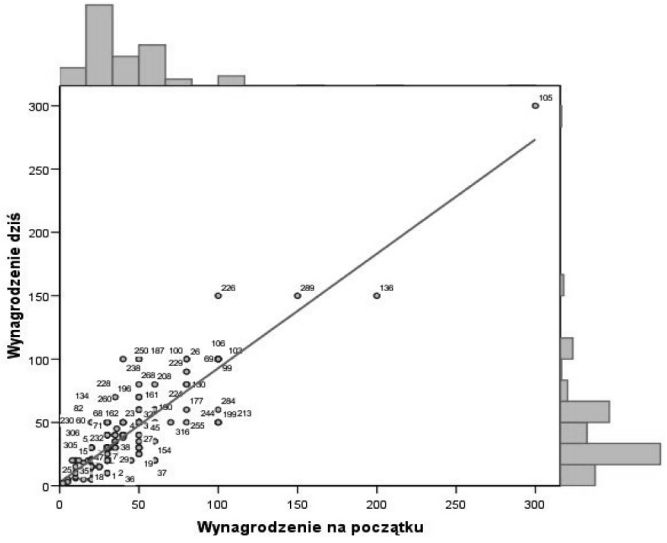
Rys. 2. Przykład wykresu słupkowego – agregowanego

Źródło: materiały SPSS Polska.



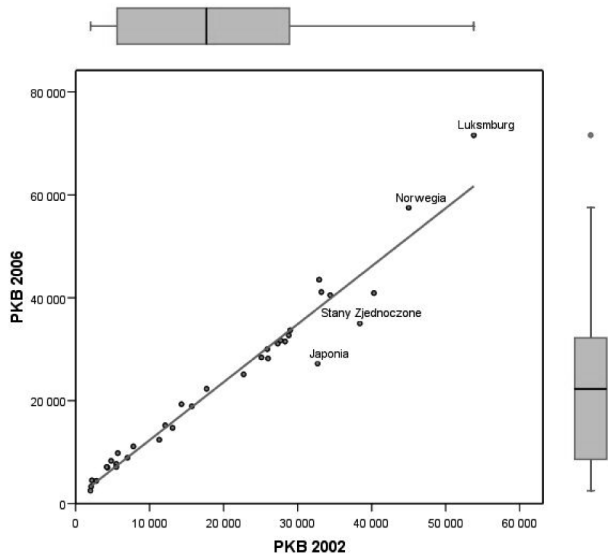
Rys. 3. Przykład wykresu punktowego z rozpraszaniem

Źródło: materiały SPSS Polska.



Rys. 4. Przykład wykresu rozrzutu z histogramami

Źródło: materiały SPSS Polska.



Rys. 5. Przykład wykresu rozrzutu z wykresami box-plots

Źródło: materiały SPSS Polska.

3. GPL – skąd się biorą takie ciekawe prezentacje?

Można się zastanowić, co zrobić, aby takie wykresy samodzielnie tworzyć. Odpowiedzią jest GPL (*Graphics Production Language*), który jest językiem projektowania wykresów i wizualizacji zorientowanym obiektowo. Jest to język, który jest oparty na tzw. gramatyce wykresów zaprezentowanej szeroko m.in. w książce *The Grammar of Graphics* Lelanda Wilkinsona [Wilkinson 2005]. Można też używać ViZml (*Visualization Markup Language*), który jest językiem podobnym do GPL pod względem zastosowań, bazującym na strukturze XML.

W języku GPL wykorzystuje się kilka składni, które warto przedstawić, np.:

- SOURCE – oznacza deklarację źródła danych, np. deklarację źródła danych, dane z SPSSa albo zewnętrzny plik csv lub baz SQL;
- DATA – deklaracje zmiennych do wykresu, czyli deklaracje zmiennych ze zbioru wyspecyfikowanego przez SOURCE lub utworzonych np. w ramach funkcji iteracyjnej. Tutaj podana jest nazwa, format, poziom pomiaru, zakres wartości zmiennych;
- SCALE – deklaracja skali wykresu, czy skala jest liniowa, jakościowa czy wykładnicza. Służy do wskazania, na jakiej skali zmienne mają być reprezentowane na wykresie;
- GUIDE – deklaracja opisu osi wykresu lub legendy, która wskazuje etykiety do osi, linie referencyjne, legendę;

– ELEMENT – specyfikuje, czy dane wykresu będą reprezentowane przez punkty, linie, słupki, obszary itp. oraz specyfikuje postępowanie w przypadku nakładania się danych, np. rozpraszanie punktów.

Dodatkowo istnieje wiele opcjonalnych instrukcji, takich jak: COMMENT – komentarz, PAGE – instrukcja deklarująca rozmiar strony, na której ma być tworzony wykres, GRAPH – podział wykresu na bloki, czyli właśnie opcja dająca możliwość umieszczania wielu wykresów na jednym obszarze, TRANS – instrukcja służąca do przekształcania zmiennych, np.: wyliczenia sumy czy innych przekształceń z użyciem funkcji EVAL, COORD – instrukcja dająca przejścia z klasycznego układu 2D z osiami X i Y na układ polarny (np. wykresy radarowe), równoległe koordynaty (np. wiele osi Y).

4. Przykład zastosowania GPL w praktyce wizualizacji wyników

Aby zaprezentować różnice między „klasycznym" a GPL-owskim budowaniem wykresów, posłużmy się pewnym przykładem dotyczącym modelu regresji logistycznej. Najpierw wyniki modelu regresji logistycznej zostaną zaprezentowane tak, jak na zajęciach z reguły jest to omawiane i pokazywane (nazwijmy to podejście umownie „studenckim"), a następnie w taki sposób, w jaki można byłoby postępować w końcowych raportach (nazwijmy to drugie podejście umownie „praktycznym").

Oszacowano pewien model regresji logistycznej dotyczący informacji o posiadanym kredycie vs. braku kredytu. Standardowo otrzymalibyśmy wyniki jak w tab. 1.

Nie wchodząc w merytoryczne uzasadnienie i interpretację tych wyników, możemy zauważyć, że jest to standardowy wydruk z zastosowania procedury regresji logistycznej, który dodatkowo może być opisany, aby było wiadomo, jak interpretować wyniki i czy są one istotne statystycznie. Głównie jednak skupiamy się na Exp(B), na tym, kiedy jest > 1, a kiedy < 1, tym samym oznaczając stymulujący/destymulujący efekt zmiennej. Jest to jedno z poprawnych i często spotykanych podejść w prezentacji wyników modeli regresji logistycznej w publikacjach czy wystąpieniach. Jednak poniżej pokazane zostanie, jak można to zrobić bardziej atrakcyjnie właśnie przy zastosowaniu języka programowania GPL.

Może warto jeszcze przypomnieć, że zaznaczając odpowiednio ostatnią kolumnę w tab. 1 przez podświetlenie, można bezpośrednio „wyprodukować" wykres, który w języku poleceń (SYNTAX) programu statystycznego PASW 17 byłby zapisany tak jak poniżej; ilustrację tego polecenia zamieszczono na rys. 6.

```
DATASET ACTIVATE ZbiórDanych1.
GRAPH
 /BAR(SIMPLE)=COUNT BY Exp(B).
```

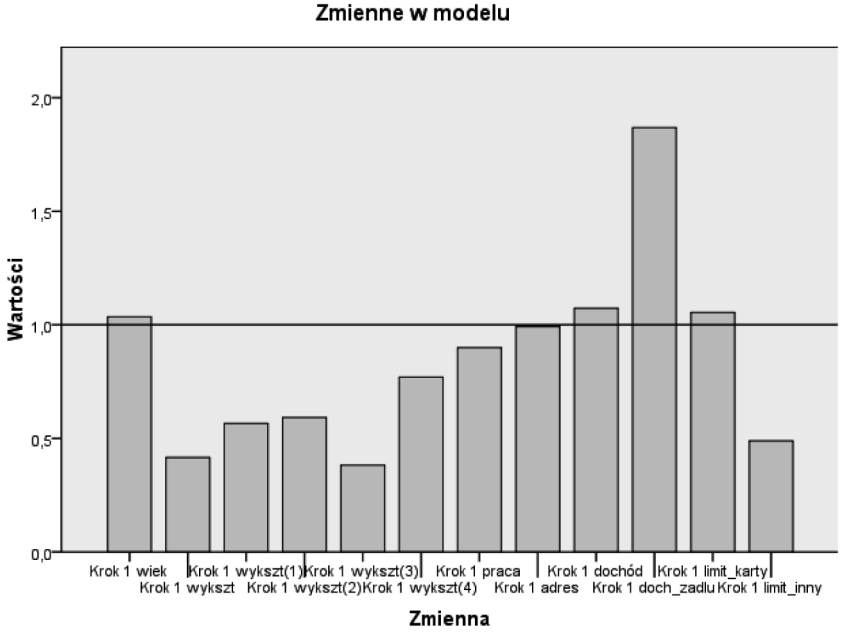
Tabela 1. Wyniki regresji logistycznej dotyczącej posiadania kredytu

Zmienne w modelu

		B	Błąd standardowy	Wald	df	Istotność	Exp(B)
a	wiek	,035	,018	4,074	1	,044	1,036
	wykszt			2,662	4	,616	
	wykszt(1)	-,876	1,294	,459	1	,498	,416
	wykszt(2)	-,569	1,294	,193	1	,660	,566
	wykszt(3)	-,524	1,304	,161	1	,688	,592
	wykszt(4)	-,961	1,334	,519	1	,471	,382
	praca	-,261	,033	60,888	1	,000	,771
	adres	-,105	,023	20,539	1	,000	,900
	dochód	-,008	,008	1,010	1	,315	,992
	doch_zadlu	,071	,031	5,340	1	,021	1,073
	limit_karty	,625	,113	30,635	1	,000	1,868
	limit_inny	,053	,078	,456	1	,499	1,054
	Stała	-,714	1,463	,238	1	,625	,490

a. Zmienne wprowadzone w kroku 1: wiek, wykszt, praca, adres, dochód, doch_zadlu, limit_karty, limit_inny.

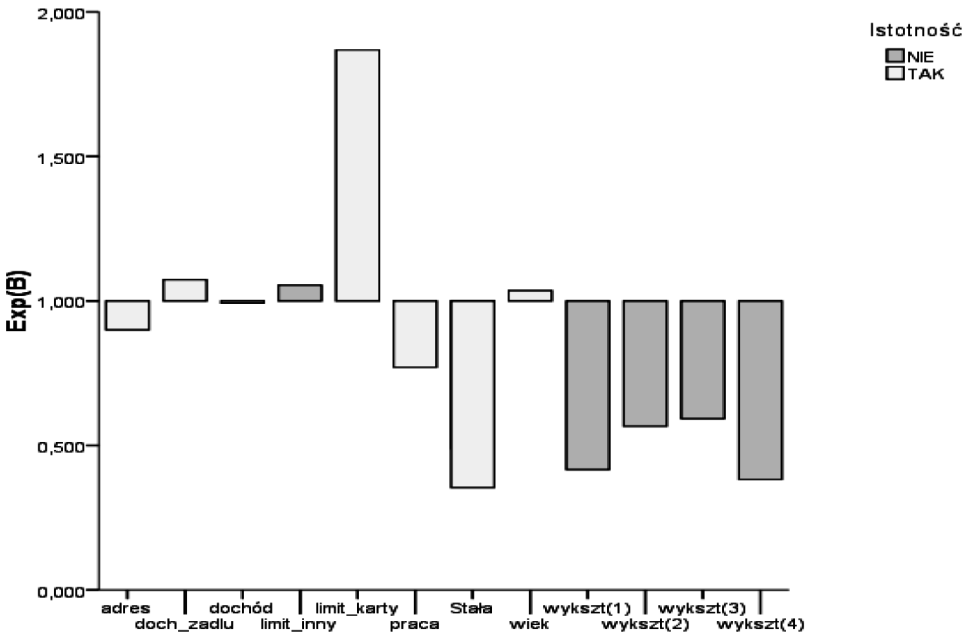
Źródło: opracowanie przy użyciu PASW 17.



Rys. 6. Przykład „zwykłego" wykresu słupkowego z zaznaczonymi wartościami Exp(B)

Źródło: opracowanie własne.

```
* Kreator wykresów.
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=Var2
MEAN(ExpB)[name="MEAN_ExpB" LEVEL=SCALE]
    Istotność MISSING=LISTWISE REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: Var2=col(source(s), name("Var2"), unit.category())
  DATA: MEAN_ExpB=col(source(s), name("MEAN_ExpB"))
  DATA: Istotność=col(source(s), name("Istotność"),
unit.category())
  GUIDE: axis(dim(3), label("Var2"))
  GUIDE: axis(dim(2), label("Exp(B)"))
  GUIDE: legend(aesthetic(aesthetic.color.interior),
label("Istotność"))
  SCALE: linear(dim(2), origin(1))
  ELEMENT: interval(position(Var2*MEAN_ExpB),
color.interior(Istotność), shape.interior(shape.square))
END GPL.
```



Rys. 7. Przykład wykresu słupkowego Exp(B) ze wskazaniem zmiennych istotnych w modelu

Źródło: opracowanie własne.

Aby inaczej zaprezentować to, co jest najważniejsze w tabeli wyników regresji, czyli które zmienne są istotne i które osiągają wartość Exp(B) większą niż 1, a które mniejszą niż 1, dobrze jest posłużyć się właśnie odpowiednio zapisanym kodem GPL, który w tym przypadku brzmi tak jak poniżej; wykonanie tego polecenia zaprezentowane jest na rys. 7:

Jak widać, dzięki zastosowaniu GPL otrzymaliśmy wykres, który nie tylko pokazuje, które zmienne są istotne statystycznie (różne kolory), ale też od razu widać, które zmienne działają stymulująco, a które destymulująco na fakt otrzymania kredytu. Cały ten wykres w takiej postaci i kolorystyce uwzględnia wszystkie najważniejsze informacje, które z reguły przedstawiono by w postaci tabeli, tak jak w opcji „studenckiej".

5. Podsumowanie

Prezentacja wyników wielowymiarowej analizy danych jest często marginalizowana i niedoceniana, a tak naprawdę temat wizualizacji powinien być szerzej omawiany nie tylko w ramach zajęć dla studentów, ale także w ramach ich dodatkowej studenckiej aktywności (np. projekty studenckich kół naukowych mogą być okazją do ćwiczenia takich umiejętności). Jest to odpowiedź na pierwsze pytanie postawione na początku artykułu.

Okazuje się, że nie zawsze bezpośrednie wykorzystanie raportów z określonych programów do analizy danych jest właściwe, ale właśnie kreatywne podejście (z uwzględnieniem rozszerzenia graficznych możliwości programów statystycznych) do prezentacji najważniejszych wyników jest jak najbardziej preferowane i powinno być nauczane, a następnie stosowane w praktyce. Chodzi o to, aby wiedzieć, co i jak chce się zaprezentować (np. tak jak w artykule wielkości Exp(B) z zaznaczonymi według różnych kolorów istotnymi statystycznie współczynnikami), aby było to zrozumiałe dla odbiorców. Odpowiadając na drugie pytanie, należy stwierdzić, że obecne możliwości programowania wraz z dostępnymi pakietami statystycznymi powinny w sposób łatwy i nieskomplikowany umożliwiać badaczom takie prezentacje, aby mogli łatwo rozszerzyć swoje umiejętności analityczne o te „wizualizacyjne". Zdaję sobie sprawę, że wymaga to od nas badaczy/analityków posiadania kolejnych umiejętności (innych niż typowe posługiwanie się metodami ilościowymi). Jednak chyba warto podejmować trud zdobywania kolejnych umiejętności w tym zakresie, aby mieć pełną satysfakcję z tego, że nasze analizy są zrozumiałe i czytelne dzięki „wizualizacjom".

Trudno jednoznacznie odpowiedzieć na ostatnie pytanie, bo to właśnie odpowiednia prezentacja wyników powinna uwzględniać to, co jest „statystycznie ważne", z tym, co „marketingowo" będzie akceptowalne i zrozumiałe, tym samym uatrakcyjniając nie tylko raporty firm badawczych, ale także wykłady i ćwiczenia ze studentami.

Literatura

Bazarnik J., Grabiński T., Kącіak E., Mynarski S., Sagan A., *Badania marketingowe. Metody i oprogramowanie komputerowe*, AE, Kraków 1992
Churchill G.A., *Badania marketingowe. Podstawy metodologiczne*, Wyd. PWN, Warszawa 2002.
Gatnar E., Walesiak M., *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, AE, Wrocław 2004.
Mazurek-Łopacińska K., *Badania marketingowe. Teoria i praktyka*, Wydawnictwo Naukowe PWN, Warszawa 2005.
Rószkiewicz M., *Metody ilościowe w badaniach marketingowych*, Wydawnictwo Naukowe PWN, Warszawa 2002.
Sagan A., *Badania marketingowe. Podstawowe kierunki*, AE, Kraków 1998.
Walesiak M., *Metody analizy danych marketingowych*, Wyd. PWN, Warszawa 1996.
Wilkinson L., *The Grammar of Graphics*, Springer, 2nd Edition, 2005.

PRESENTATION OF RESEARCH RESULTS – TEACHING AND PRACTICE

Summary

In the following paper, various graphical possibilities of presenting research results are shown. First, there are presented simple charts, however not independently but together on one graph (as i.e. scatterplots with histogram, or box-plots with histogram). Further, there are shown how – on the graph – "typical" results of logistic regression could be illustrated by only statistically significant Exp(B). The Graphical Programming Language (GPL) could be used in order to obtain different graphs. GPL could be used via available statistical packages (i.e. PASW 17) and so it gives possibilities of more interesting presentations and visualization of the results.