

SPSS 16.0 – nowe funkcje, nowe możliwości

SPSS INC. kontynuując tradycję udoskonalania narzędzi statystycznych, prezentuje nową wersję programu SPSS. SPSS 16.0 jest dostępny na platformach: Windows, Mac i Linux. Oferuje nową formę interfejsu graficznego, jeszcze bardziej funkcjonalną i intuicyjną w obsłudze. Wraz z wersją 16.0 pojawia się nowy moduł analityczny SPSS NEURAL NETWORKS, a kilka modułów dostępnych w poprzednich wersjach zostało rozbudowanych o zaawansowane opcje. SPSS 16.0 to także pełniejsza integracja ze środowiskiem SPSS PREDICTIVE ENTERPRISE SERVICES i językiem Python oraz integracja z programem R. Kolejne funkcjonalności pozwalają na jeszcze wygodniejsze zarządzanie danymi i dostosowywanie postaci raportów do potrzeb użytkowników.

- nowy interfejs użytkownika
- nowy moduł SPSS NEURAL NETWORKS
- uzupełnienie istniejących technik analitycznych
- zarządzanie danymi i raportowanie
- udoskonalona programowalność
- integracja z PREDICTIVE ENTERPRISE SERVICES

nowy interfejs użytkownika

Nowy interfejs użytkownika został całkowicie zaprojektowany w języku Java. Czynie on SPSS jeszcze łatwiejszym w obsłudze i w pracy z danymi. Użytkownik może zmieniać rozmiar okien dialogowych, co umożliwia na przykład pełny podgląd długich nazw i etykiet zmiennych. W nowej wersji możliwa jest także szybsza selekcja oraz przenoszenie zmiennych między panelami metodą przeciągnięcia i upuszczenia.

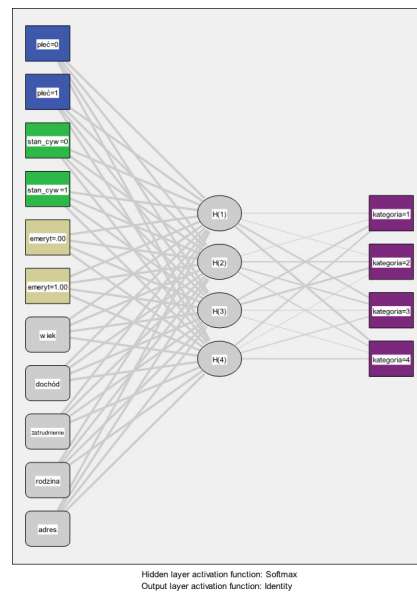
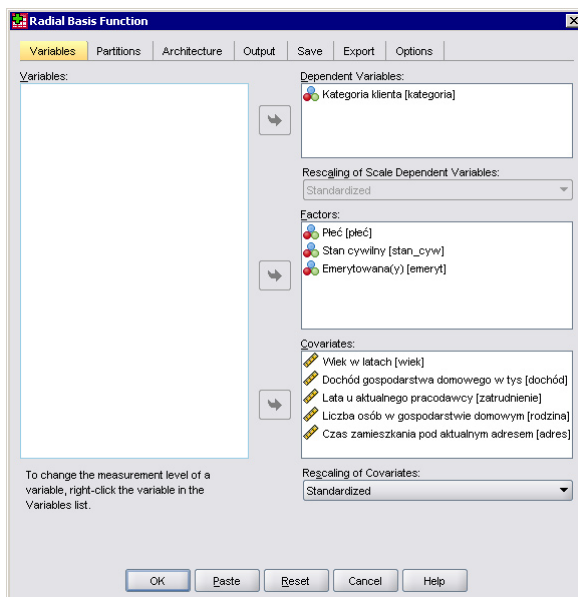
nowy moduł SPSS NEURAL NETWORKS

SPSS NEURAL NETWORKS pozwala na odkrywanie bardziej złożonych związków pomiędzy danymi. Sieć neuronowa to nieliniowe narzędzie do klasyfikacji i predykcji. Typowa sieć neuronowa składa się z pewnej liczby neuronów zorganizowanych w warstwy, które możemy podzielić na wewnętrzne i zewnętrzne oraz dodatkowo na jedną lub więcej warstw ukrytych. W trakcie procesu uczenia sieci neuronowych, czyli procesu odkrywania przez sieć związków pomiędzy zmiennymi wyjaśniającymi i zmienną wyjaśnianą, połączeniom pomiędzy neuronami przypisane zostają odpowiednie wagi wyznaczone przez algorytm tak, aby minimalizować błąd i zapewnić jak najlepszą predykcję. Działanie sieci neuronowych (w pewnym dużym uproszczeniu) przypomina działanie ludzkiego mózgu.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
kredyty	rata	akt_adres	wiek	kredyty	saldo	cel	szczegolnosc	zatrudnienie	stan_cyw	innu_dluz	dom	telefon			
169,00	4,00	4,00	67,00	2,00	debet	AGD	nieznane	>7	ZonatyMe...	brak	wlasny	nie			
951,00	2,00	2,00	22,00	1,00	debet	AGD	<1000	1-4	ZonatyMe...	brak	wlasny	tak			
996,00	2,00	3,00	49,00	1,00	debet	edukacja	<1000	4-7	ZonatyMe...	brak	wlasny	tak			
982,00	2,00	4,00	45,00	1,00	debet	meble	<1000	4-7	ZonatyMe...	porozczenie	wlasny	tak			
870,00	3,00	4,00	53,00	2,00	brak debetu	samochod	<1000	1-4	ZonatyMe...	brak	wlasny	tak			
955,00	2,00	4,00	35,00	1,00	debet	edukacja	nieznane	1-4	ZonatyMe...	brak	wlasny	nie			
935,00	3,00	4,00	53,00	1,00	debet	meble	5000-10000	>7	ZonatyMe...	brak	wlasny	tak			
948,00	2,00	2,00	35,00	1,00	debet	samochod	<1000	1-4	ZonatyMe...	brak	wynajem	nie			
959,00	2,00	4,00	61,00	1,00	debet	AGD	>10000	4-7	ZonatyMe...	brak	wlasny	tak			
234,00	4,00	2,00	28,00	2,00	debet	samochod	<1000	bezzrobotny	Wolny	brak	wlasny	tak			
295,00	3,00	1,00	25,00	1,00	debet	samochod	<1000	<1	ZonatyMe...	brak	wynajem	tak			
308,00	3,00	4,00	24,00	1,00	debet	inwestycja	<1000	<1	ZonatyMe...	brak	wynajem	tak			
567,00	1,00	1,00	22,00	1,00	debet	AGD	<1000	1-4	ZonatyMe...	brak	wlasny	nie			
1199,00	4,00	4,00	60,00	2,00	debet	samochod	<1000	>7	ZonatyMe...	brak	wlasny	tak			
1403,00	2,00	4,00	28,00	1,00	debet	samochod	<1000	1-4	ZonatyMe...	brak	wynajem	tak			
1282,00	4,00	2,00	32,00	1,00	debet	AGD	1000-5000	1-4	ZonatyMe...	brak	wlasny	tak			
2424,00	4,00	4,00	53,00	2,00	debet	AGD	nieznane	>7	ZonatyMe...	brak	wlasny	tak			
8072,00	2,00	3,00	25,00	3,00	debet	inwestycja	nieznane	<1	ZonatyMe...	brak	wlasny	tak			
12579,00	4,00	2,00	44,00	1,00	debet	samochod	<1000	>7	ZonatyMe...	brak	wlasny	nie			
3430,00	3,00	2,00	31,00	1,00	debet	AGD	5000-10000	>7	ZonatyMe...	brak	wlasny	nie			
2134,00	4,00	4,00	48,00	3,00	debet	samochod	<1000	1-4	ZonatyMe...	brak	wlasny	nie			
2647,00	2,00	3,00	44,00	1,00	debet	AGD	5000-10000	1-4	ZonatyMe...	brak	wynajem	tak			
2241,00	1,00	3,00	48,00	2,00	debet	samochod	<1000	<1	ZonatyMe...	brak	wynajem	tak			
1804,00	3,00	4,00	44,00	1,00	debet	samochod	1000-5000	<1	ZonatyMe...	brak	wlasny	tak			
2069,00	2,00	1,00	26,00	2,00	debet	meble	nieznane	1-4	Wolny	brak	wlasny	tak			
6,00	1374,00	1,00	2,00	36,00	1,00	debet	meble	<1000	1-4	ZonatyMe...	brak	wlasny	nie		

Rysunek 1.
Edytor danych SPSS 16.0

W ramach SPSS NEURAL NETWORKS dostępne są: sieć typu wielowarstwowy perceptron oraz sieć typu RBF.



Classification

Sample	Obszar	Predicted				
		Podstawowa obsługa	E-obsługa	Dodatkowa obsługa	Pełna obsługa	Percent Correct
Training	Podstawowa obsługa	100	4	52	26	54,9%
	E-obsługa	67	3	56	24	2,0%
	Dodatkowa obsługa	56	6	97	23	53,3%
	Pełna obsługa	72	4	60	33	19,5%
	Overall Percent	43,2%	2,5%	38,8%	15,5%	34,1%
Testing	Podstawowa obsługa	28	0	15	13	50,0%
	E-obsługa	13	1	19	6	2,6%
	Dodatkowa obsługa	32	1	35	5	47,9%
	Pełna obsługa	20	3	18	9	18,0%
	Overall Percent	42,7%	2,3%	39,9%	15,1%	33,5%
Holdout	Podstawowa obsługa	14	0	10	4	50,0%
	E-obsługa	10	0	8	10	0%
	Dodatkowa obsługa	6	0	16	4	61,5%
	Pełna obsługa	8	0	8	1	5,9%
	Overall Percent	38,4%	0%	42,4%	19,2%	31,3%

Dependent Variable: Kategoria Klienta

Rysunek 2.

Okno dialogowe tworzenia sieci neuronowej typu RBF. Schemat sieci. Tabela oceny modelu (przewidywane wartości vs. obserwowane wartości)

uzupełnienie istniejących technik analitycznych

SPSS 16.0 oferuje dodatkowe lub ulepszone techniki w ramach SPSS COMPLEX SAMPLES, SPSS ADVANCED MODELS, AMOS oraz dzięki SPSS PROGRAMMABILITY EXTENSION dostęp do metod analitycznych oprogramowanych w pakiecie statystycznym R.

SPSS COMPLEX SAMPLES udostępnia metodę regresji Cox'a w celu przeprowadzenia analiz przeżycia dla danych pochodzących z badań o złożonym schemacie doboru próby. Jeśli użytkownik dysponuje takimi danymi można użyć powyższej techniki do dokładnego przewidywania czasu

pozostałego do danego wydarzenia, np. jak długo klient o wysokiej wartości będzie aktywny. SPSS COMPLEX SAMPLES Cox Regression (CSCOXREG) umożliwia analizę różnic w podgrupach, jak również wyników dla zbioru predyktorów. Procedura bierze pod uwagę schemat doboru próby w trakcie wyliczania statystyk i może być przeprowadzona na danych zawierających powtarzające się przypadki, np. wielokrotne wizyty pacjenta.

SPSS ADVANCED MODELS oferuje dodatkowe opcje dla uogólnionych modeli liniowych (GENLIN) oraz uogólnionych równań estymujących (GEE), które zostały wprowadzone w SPSS 15.0. Procedury te umożliwiają pełniejszą predykcję zmiennych porządkowych, takich jak np. satysfakcja klientów mierzona na skali porządkowej. Udoskonalenia dostępne w SPSS 16.0 pozwalają analitykom przewidywać zmienne zależne, które mają rozkład złożony Tweedie'go i są wykorzystywane na przykład w analizach portfela ryzyk.

AMOS jest wielozadaniowym narzędziem z rodziny SPSS, służącym do modelowania równań strukturalnych (SEM), teraz także umożliwiającym analizę klas latentnych (latent class analysis) oraz modeli mieszanych (mixture modeling).

W SPSS 16.0 możliwe jest także stosowanie, po uprzedniej instalacji PLS EXTENSION MODULE, regresji metodą cząstkowych najmniejszych kwadratów (Partial Least Squares – PLS). Jest to alternatywna metoda dla klasycznej regresji metodą najmniejszych kwadratów (Ordinary Least Squares – OLS). PLS jest techniką predykcyjną, dzięki której można prowadzić analizy dla wielu zmiennych niezależnych, nawet jeśli wykazują one silną liniową korelację. Wybór techniki PLS zamiast OLS jest uzasadniony na przykład, gdy użytkownik dysponuje dużą liczbą zmiennych w porównaniu z ilością obserwacji – w sytuacji, która często występuje w badaniach ankietowych.

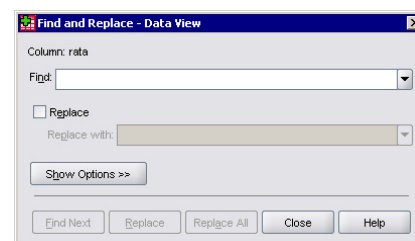
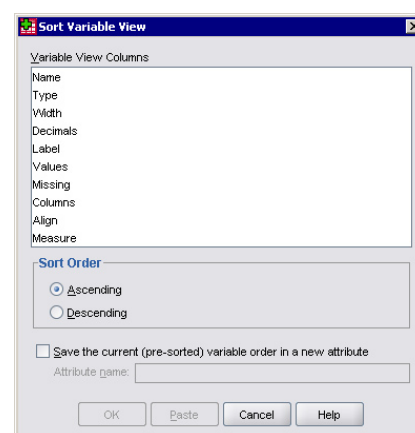
zarządzanie danymi i raportowanie

SPSS 16.0 zawiera wiele użytecznych funkcjonalności do obsługi danych:

- Zmiana długości lub typu istniejących zmiennych tekstowych przy użyciu języka poleceń
- Definiowanie etykiet wartości i braków danych dla zmiennych tekstowych bez względu na ich szerokość
- Korzystanie z nowych funkcjonalności edytora danych, m.in. wyszukiwania i zastępowania wartości danych, sprawdzania poprawności pisowni etykiet wartości i etykiet zmiennych, sortowania zmiennych wg nazw, typu, formatu itp.
- Wyszukiwanie i zastępowanie tekstu w oknie raportów – np. wyszukiwanie ostrzeżeń w celu zidentyfikowania pojawiających się problemów w raporcie
- Zapis i odczyt danych w plikach Excel 2007
- Wybór pomiędzy pracą z wieloma lub z jednym zbiorem danych w tym samym czasie
- Definiowanie domyślnego folderu wczytywania i zapisywania danych
- Zastąpienie edytora wykresów interaktywnych IGRAPH nowym, bardziej efektywnym narzędziem wizualizacji wykresów, które tworzy i edytuje wykresy szybciej i łatwiej
- Możliwość użycia standardu Unicode przy wczytywaniu i zapisywaniu plików danych oraz plików języka poleceń przy pracy ze zbiorem danych w wielu językach SPSS.

udoskonalona programowalność

SPSS PROGRAMMABILITY EXTENSION wydajnie rozszerza możliwości SPSS poprzez możliwość korzystania z zewnętrznych języków programowania, takich jak Python i Microsoft.NET. Integrujące wtyczki (plug-ins) dostępne są na stronie internetowej SPSS DEVELOPER CENTRAL.

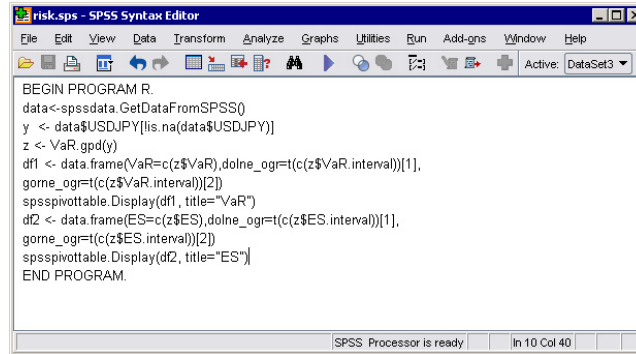


Rysunek 3.

Sortowanie zmiennych oraz wyszukiwanie i zastępowanie danych w edytorze danych.

SPSS 16.0 wprowadza język Python jako podstawowy język skryptowy. Python zastąpił wbudowany w poprzednich wersjach język SAX BASIC dla takich zadań jak: automatyzacja oraz dostosowywanie raportów do potrzeb użytkownika. Podobnie jak przy użyciu SAX BASIC, użytkownik może stosować autoskrypty do modyfikowania raportów otrzymywanych po wykonaniu jednej lub wielu procedur SPSS. Utworzone wcześniej skrypty w języku SAX BASIC będą mogły być nadal obsługiwane w nowej wersji SPSS.

W SPSS PROGRAMMABILITY EXTENSION integracja wtyczki dla pakietu R umożliwia skorzystanie w ramach języka poleceń spss z bogactwa procedur analitycznych dostępnych w środowisku R.



```

BEGIN PROGRAM R.
data<-spssdata.GetDataFromSPSS()
y <- data$USDJPY[!is.na(data$USDJPY)]
z <- VaR.gpd(y)
df1 <- data.frame(VaR=c(z$VaR),dolne_ogr=t(c(z$VaR.interval))[1],
gorne_ogr=t(c(z$VaR.interval))[2])
spsspivottable.Display(df1, title="VaR")
df2 <- data.frame(ES=c(z$ES),dolne_ogr=t(c(z$ES.interval))[1],
gorne_ogr=t(c(z$ES.interval))[2])
spsspivottable.Display(df2, title="ES")
END PROGRAM.
  
```

VaR		
VaR	dolne_ogr	gorne_ogr
1,637	1,485	1,882
ES		
ES	dolne_ogr	gorne_ogr
2,113	1,813	2,599

Rysunek 4.

Wyliczenie metodą Peaks over Threshold (POT) wartości narażonej na ryzyko (VaR) i oczekiwanej warunkowej straty (ES) dla wskaźnika USD/JPY wraz z 95% przedziałami ufności, przy pomocy procedury pakietu R.

Więcej informacji o możliwościach programowalności na stronie: www.spss.com/devcentral

integracja z PREDICTIVE ENTERPRISE SERVICES

spss 16.0 dodatkowo zapewnia integrację z SPSS PREDICTIVE ENTERPRISE SERVICES – funkcjonalnym środowiskiem uruchomieniowym składanym w nim procedur analitycznych, stworzonym przez SPSS INC. SPSS PREDICTIVE ENTERPRISE SERVICES udostępnia i umożliwia między innymi:

- ustalanie harmonogramów wykonywania modeli i stworzenie systemu powiadomień o przebiegu analiz automatycznych,

- zarządzanie wynikami analiz,
- organizację przechowywania danych dostosowaną do potrzeb analitycznych,
- zaawansowane i przejrzyste zarządzanie użytkownikami, grupami, autoryzacjami, zabezpieczeniami oraz uprawnieniami.

Integracja aplikacji SPSS i PES możliwa jest dzięki SPSS ADAPTER FOR PREDICTIVE ENTERPRISE SERVICES.

Więcej informacji o SPSS PREDICTIVE ENTERPRISE SERVICE na stronie:

www.spss.com/predictive_enterprise_services

SPSS POLSKA

ul. Raclawicka 58
30-017 Kraków
tel./faks 012.636.07.91
tel./faks 012.636.45.35
tel./faks 012.636.96.80
e-mail: info@spss.pl
www.spss.pl
www.analizadanych.pl
www.webmining.pl