

**Opis przedmiot zamówienia: Oprogramowanie do zaawansowanej i szeroko
pojętej kompleksowej statystycznej analizy danych
– licencja bezterminowa 16 stanowisk**

Cechy użytkowe oprogramowania:

- wersja jedno stanowiskowa dla jednego nazwanego użytkownika z licencją producenta bez ograniczenia w czasie – 16 sztuk
- główne środowisko pracy w programie w języku polskim (anglojęzyczne zwroty jedynie w niewielkim zakresie całego interfejsu)
- prawo do korzystania z serwisu (obejmującego pomoc techniczną i aktualizacje do nowych wersji oprogramowania) w ciągu **12** miesięcy od zakupu
- pomoc techniczna świadczona w języku polskim; rozbudowana pomoc elektroniczna zawierająca opisy poszczególnych opcji programu oraz dla wybranych modułów opisane krok po kroku przykłady analiz (podręcznik internetowy użytkownika w zakładce pomoc i wsparcie - samouczek; poradnik statystyczny w zakładce pomoc i wsparcie)

Środowisko pracy z programem i korzystanie z zewnętrznych danych:

- Dane są składowane w arkuszu danych umożliwiającym interakcyjne wprowadzanie i przekształcanie danych (sortowanie, transformacje zmiennych, ułoż w stertę/rozrzuć po zmiennych) oraz import i eksport danych (m.in. z plików Excel i plików tekstowych).
- Integracja oprogramowania ze standardowymi bazami danych (SQL Server, MS Access i inne) przez OLE DB.
- Wczytywanie i zapis danych w formacie Excel (.xls, .xlsx, .xlsb, .xlsm), tekstowym, csv, html i innych.
- Wczytywanie i zapis plików danych w formatach minimum: Statistica, SPSS, SAS, JMP, Minitab.
- Oprogramowanie zawiera wbudowany, zgodny ze standardami język programowania Visual Basic, który umożliwia dostęp programowy do funkcji programu, programowanie własnych procedur analitycznych oraz automatyzację prac.
- Środowisko użytkownika umożliwiające graficzne definiowanie projektu analitycznego w postaci schematu (grafu), w którym źródła danych, procedury przetwarzania danych i wyniki reprezentowane są przez ikony, a przepływ danych obrazują strzałki.
- Możliwość uruchamiania procedur w językach R i Python w projektach analitycznych zdefiniowanych jako schemat graficzny (graf).
- Oprogramowanie działa na stanowisku komputerowym pod kontrolą systemu operacyjnego Windows 7/8/10.

Zarządzanie wynikami

- Oprogramowanie zapewnia możliwość tworzenia raportów z analizy, z możliwością zapisania w formacie PDF (bezpośrednio, bez użycia emulatora drukarki).
- Przesyłanie wyników (tabel, wykresów) do dokumentów edytora tekstowego (np. MsWord).
- Raport otrzymywany przy pomocy oprogramowania przypomina dokument edytora tekstu, a poszczególne obiekty (np. wykresy, arkusze, arkusz czy wykres MS Excel) umieszczane są w nim kolejno, jeden za drugim. Raporty mogą być zapisywane nie tylko we własnym formacie oprogramowania, ale także w postaci plików RTF, HTML.
- Oprogramowanie pozwala na zapis dokumentów (arkuszy danych i wyników, raporty) w postaci plików HTML, gotowych do opublikowania w Internecie lub Intranecie.
- Możliwość aktualizacji utworzonych wykresów po zmianie danych źródłowych (automatycznie lub przez użytkownika).
- Możliwość edycji wykresów po ich wstawieniu do dokumentu edytora tekstowego (tzn. wykresy mogą być wstawiane jako obiekty OLE) na komputerach z zainstalowanym oprogramowaniem.
- Wynikowe tabele mają format pliku danych, dzięki czemu można na nich (tzn. na wynikowych tabelach) łatwo wykonywać kolejne analizy.

Funkcjonalność oprogramowania:

Oprogramowanie udostępnia w jednym środowisku użytkownika następujące funkcje analityczne:

- Statystyki podstawowe i tabele
- Możliwość wykonywania analiz w grupach
- Wykresy: histogramy, wykresy rozrzutu, wykres workowy, wykresy średnia i błędy, wykresy ramkowane, wykres składowych zmienności, wykresy zakresu, wykres rozrzutu z błędem, obrazkowe wykresy rozrzutu, wykresy rozrzutu z rysunkami, wykresy rozrzutu z histogramami, wykresy normalności, wykresy kwantyl-kwantyl, wykresy prawdopodobieństwo-prawdopodobieństwo, wykresy słupkowe/kolumnowe, wykresy liniowe, wykresy sekwencyjne/nakładane, wykresy kołowe, wykresy brakujących danych i spoza zakresu, histogramy dwóch zmiennych, wykresy powierzchniowe, wykresy warstwowe, wykresy wafłowe, wykresy trójkątne, skategoryzowane wykresy XYZ, skategoryzowane wykresy trójkątne, wykresy macierzowe, wykresy obrazkowe, wykresy XYZ 3W, wykresy trójkątne 3W
- Dopasowanie rozkładów
- Regresja wieloraka
- Analiza wariancji (ANOVA)
- Statystyki nieparametryczne
- Ogólne modele liniowe
- Uogólnione modele liniowe i nieliniowe

- Ogólne modele regresji
- Modele cząstkowych najmniejszych kwadratów
- Komponenty wariacyjne
- Analiza przeżycia
- Estymacja nieliniowa
- Linearyzowana regresja nieliniowa
- Analiza log-liniowa tabel liczości
- Szeregi czasowe i prognozowanie
- Modelowanie równań strukturalnych
- Analiza skupień
- Analiza czynnikowa
- Składowe główne i klasyfikacja
- Algorytm NIPALS dla analizy składowych głównych i metody cząstkowych najmniejszych kwadratów
- Analiza kanoniczna
- Analiza rzetelności i pozycji
- Drzewa klasyfikacyjne
- Analiza korespondencji
- Skalowanie wielowymiarowe
- Analiza dyskryminacyjna
- Ogólne modele analizy dyskryminacyjnej
- Analiza mocy testów
- Standardowe karty kontrolne (m.in.: karty X-średnie i R, P, C, U, CUSUM, EWMA, Pareto)
- Wielowymiarowe karty kontrolne (Hotellinga, MEWMA, MCUSUM)
- Analiza zdolności procesu: wskaźniki zdolności procesów (np. Cp, Cr, Cpk, Cpl, Cpu, K, Cpm, Pp, Pr, Ppk, Ppl, Ppu i inne),
- Plany badania i analiza powtarzalności i odtwarzalności pomiarów (R&R)
- Analiza Weibulla
- Planowanie doświadczeń (DoE): standardowe plany frakcyjne dwuwartościowe 2(k-p), plany frakcyjne 2(k-p) o najmniejszej aberracji i maksymalnym niewykłaniu, plany eliminacyjne (Placketta-Burmana), plany frakcyjne trójwartościowe typu 3(k-p) z podziałem na bloki oraz plany Boxa-Behnkena, plany centralne kompozycyjne (powierzchnia odpowiedzi), plany kwadratów łańcuchowych, doświadczenia wg metody Taguchi, plany dla mieszanin i powierzchni o podstawie trójkątnej, plany dla ograniczonych powierzchni i mieszanin, plany D i A-optymalne
- MSPC (Multivariate Statistical Process Control):
 - Analiza składowych głównych (PCA)
 - Cząstkowe najmniejsze kwadraty (PLS)

- Wielokierunkowe cząstkowe najmniejsze kwadraty wg partii (BMPLS)
- Wielokierunkowa analiza składowych głównych według czasu (TMPCA)
- Wielokierunkowe cząstkowe najmniejsze kwadraty wg czasu (TMPLS)
- Tabele raportujące
- Automatyczne sieci neuronowe
- Ogólne modele drzew klasyfikacyjnych i regresyjnych (GTrees)
- Ogólne modele CHAID (Chi-square Automatic Interaction Detection)
- Interakcyjne drzewa klasyfikacyjna i regresyjne (C&RT, CHAID)
- Drzewa klasyfikacyjne i regresyjne ze wzmacnianiem (Boosted Trees)
- Losowy las (Random Forests)
- Analiza składowych niezależnych (ICA – Independent Component Analysis)
- Analiza skupień uogólnioną metodą EM i k-średnich
- Analiza koszykowa (Association Rules)
- Uogólnione modele addytywne (GAM - General Additive Models)
- Multivariate Adaptive Regression Splines (MAR Splines)
- Inne metody uczenia maszyn (Machine Learning): wektory nośne (Support Vector Machines), naiwny klasyfikator Bayesa, K-najbliższych sąsiadów (K-Nearest Neighbors).
- Szybkie wdrażanie modeli predykcyjnych
- Dobroć dopasowania, klasyfikacji, predykcji (Goodness of fit)
- Dobór i eliminacja zmiennych (dla dużych zbiorów danych)
- Łączenie grup (klas) z wykorzystaniem algorytmu CHAID
- Interakcyjne drążenie danych
- Wykrywanie reguł asocjacji
- Analiza sekwencji
- Analiza skojarzeń
- Wykresy zmienności
- Wykresy wielokrotne, pozwalające bezpośrednio porównywać wiele zmiennych zależnych
- Komponenty wariacyjne z przedziałami ufności
- Elastyczne operowanie wieloma zmiennymi zależnymi: jednoczesne analizowanie wielu miennych wg tego samego lub różnych planów
- Wykresy komponentów wariacyjnych
- Analiza podstawowych przyczyn
- Optymalizacja wielkości wyjściowych
- Ogólna optymalizacja
- Funkcjonalność text mining
- Analiza dokumentów zapisanych w formacie MS Word

- Zliczanie wystąpień słów
- Różne miary częstości występowania słów: prosta częstość, częstość binarna (ang. binary frequency), odwrotna częstość dokumentowa (ang. inverse document frequency), częstość logarytmiczna
- Możliwość określania własnej stop-listy
- Możliwość określania synonimów
- Wykonywanie rozkładu według wartości osobliwych (ang. singular value decomposition) dla miar częstości występowania słów w zbiorze dokumentów
- Definiowanie reguł poprawności danych
- Reguły poprawności danych
- Analiza brakujących danych
- Przekodowanie na zmienne sztuczne
- Szybkie rekodowanie
- Przekształcenia zmiennych
- Zliczanie wystąpień
- Porządkuj zmienne wielokrotnych odpowiedzi
- Kalkulator liczebności próby
- Ważenie wieńcowe przypadków
- Propensity score matching
- Podsumowanie skali pozycyjnej
- Podsumowanie skali rangowej
- Wykres dyferencjału semantycznego
- Wykres dla skali Stapela
- Rzetelność skali
- Metoda ocen porównawczych Thurstone'a
- Współczynniki zgodności sędziów
- Krzywe ROC
- Metaanaliza i metaregresja
- Kreator regresji logistycznej
- Kreator regresji liniowej
- Analiza conjoint
- Aglomeracja z punktem odcięcia
- Analiza PROFIT
- Uogólniona metoda składowych głównych (PCA)
- Porządkowanie liniowe
- Bootstrap

- Miary powiązania/efektów dla tabel 2x2
- Analiza koncentracji
- Standaryzowane miary efektu
- Test post hoc ANOVA Friedmana
- CATANOVA
- Karta CUSUM ważona ryzykiem
- Indeks KMO oraz Test sferyczności Bartletta
- Wykres Blanda-Altmana
 - Badanie ciągów pomiarów
 - Przedziały odniesienia
 - Przedział ufności dla ilorazów
 - Profile ryzyka
- Wykres słupkowy (kolorowe słupki)
- Wykres sekwencyjny
- Wykres radarowy
- Wykres mozaikowy
- Wykres kołowy (SPie plot)
- Piramida populacyjna
- Możliwość wizualizacji danych na mapach:
 - a) Gotowe szablony map dostępne w programie obejmują podział Polski na: województwa, powiaty, gminy, okręgi wyborcze, województwa w podziale na powiaty, województwa w podziale na gminy, województwa w starym podziale
 - b) Możliwość wczytywania innych niż zawarte w programie szablonów map w formacie *.shp
 - c) Kolorowanie na mapach tła obszarów wartościami zadanej zmiennej (predefiniowane palety do wyboru, możliwość ustalenia palety użytkownika, możliwość ustalenia własnych granic dla przedziałów legendy, możliwość zapisu/wczytania palety kolorów z/do pliku)
 - d) Generowanie wykresów kołowych i słupkowych (możliwość ręcznej zmiany wielkości wykresu, możliwość ręcznego ustalenia jego położenia, możliwość zmiany skalowania wysokości słupka względem wiersza/kolumny/całości, zmienny promień wykresu kołowego zależny od wartości ze zmiennej)
 - e) Wyświetlanie etykiet tekstowych pobranych z zadanej zmiennej lub zmiennej zawierającej mapowanie elementów wraz z formatowaniem zadanych przez użytkownika (kolor, krój itp.), oraz ręczną korekcją położenia etykiety względem innych elementów wykresu
 - f) Różne stany wyświetlania elementów obszaru – aktywny, nieaktywny, ukryty
 - g) Rodzaj i grubość linii rysowanych jako granice może być zmieniana przez użytkownika
 - h) Możliwość zapisu/odczytu z i do pliku wszystkich opcji wyglądu mapy

- i) Możliwość ręcznej edycji przez użytkownika szablonów map wczytanych w programie (usuwanie obszarów, scalanie obszarów) i zapisu jako nowy szablon
- j) Możliwość zarejestrowania wygenerowanej mapy (z wizualizacją danych) w postaci makra

Ponadto dostawca przeszkoli pracowników Zamawiającego w formie minimum 4 - godzinnego szkolenia w siedzibie Zamawiającego z obsługi dostarczonego oprogramowania w terminie nie później niż miesiąc od daty dostarczenia oprogramowania. Czas i termin szkolenia zostanie ustalony wspólnie przez obie strony.

Dostarczane oprogramowanie ma być zakupione u autoryzowanego partnera producenta programu, razem z odpowiednim wsparciem („maintenance”) wymaganym przez Zamawiającego, w najnowszej stabilnej wersji znajdującej się na rynku.