

## Opis przedmiotu zamówienia

Dostawa aktualizacji do najnowszej wersji posiadanego przez Zamawiającego oprogramowania SPSS.

Zamawiający dopuszcza dostawę nowej licencji na oprogramowanie równoważne obejmujące funkcjonalność statystyczną, prezentacyjną i obsługową na nie niższym stopniu zaawansowania statystycznego i prezentacyjnego oraz programowania niż najnowsza dostępna wersja oprogramowania SPSS. Oprogramowanie równoważne musi spełniać wymagania określone w pkt 4.

## 1. Opis dotychczasowego środowiska SPSS

Statystyka publiczna obecnie wykorzystuje oprogramowanie SPSS w wersji 21. Oprogramowanie udostępniane jest głównie poprzez farmę terminalową Citrix. Środowisko obejmuje licencje pływające pozwalające na jednoczesną pracę określonej liczby użytkowników (licencja Concurrent Users). Przyznawaniem dostępu do oprogramowania zarządza serwer licencji SPSS.

W statystyce publicznej oprogramowanie SPSS w wersji 21, jest oparte zarówno o serwery licencji jak i instalacje jednostanowiskowe, obejmuje:

- a. Licencje Sieciowe obejmujące komponenty:
  - IBM SPSS Statistics Base 80
  - IBM SPSS Custom Tables 10
  - IBM SPSS Regression 4
  - IBM SPSS Advanced Statistics 4
  - IBM SPSS Categories 4
  - IBM SPSS Forecasting 3
  - IBM SPSS Missing Values 3
  - IBM SPSS Complex Samples 2
  - IBM SPSS Conjoint 1
- b. Licencje jednostanowiskowe obejmujące komponenty:
  - IBM SPSS Statistics Base 10
  - IBM SPSS Custom Tables 10
  - IBM SPSS Regression 1
  - IBM SPSS Advanced Statistics 5
  - IBM SPSS Categories 2
  - IBM SPSS Forecasting 8
  - IBM SPSS Missing Values 7
  - IBM SPSS Complex Samples 3
  - IBM SPSS Conjoint 4
  - IBM SPSS Exact Tests 4

### 1.1. Opis infrastruktury sprzętowo-systemowej oraz oprogramowania posiadanego przez Zamawiającego i dedykowanych dla wdrożenia.

#### Infrastruktura

W Centrum Przetwarzania Danych GUS, Zamawiający posiada środowisko serwerowe oparte na systemach operacyjnych MS Windows Server, stanowiącym platformę systemową dla następujących komponentów infrastruktury informatycznej:

- 1) Systemu usług katalogowych bazującego na MS Active Directory w wersji Windows Server 2019 o funkcjonalności lasu i domeny co najmniej na poziomie wersji Windows Server 2012 R2.
- 2) Środowisko do wirtualizacji serwerów bazujących na oprogramowaniu VMware Cloud Foundation Advanced z funkcjami NSX i vSAN oraz vCenter Server 6.7.

- 3) Oprogramowania do wykonywania kopii zapasowych Veeam Backup@Replication Enterprise Plus.
- 4) Środowisk terminalowych CITRIX wykorzystujących oprogramowanie Citrix Virtual Apps 1912 LTSR.

Na stacjach roboczych zainstalowane są następujące systemy operacyjne posiadane przez Zamawiającego: MS Windows 7 Pro, MS Windows 8.1. Pro i MS Windows 10 Pro.

Analityczne aplikacje dla użytkowników końcowych instalowane i uruchamiane są w środowisku usług terminalowym CITRIX.

Zamawiający posiada na całe środowisko wirtualne licencje serwerowe Microsoft Windows 2019 oraz na wybrane serwery oprogramowanie bazodanowe Microsoft SQL Server 2019.

Środowisko wirtualizacyjne jest umiejscowione na infrastrukturach serwerowo-sieciowych HPE Synergy12000 Frame z dwu procesorowymi (20 i 8 rdzeniowymi) serwerami HPE SY 480 Gen10 pracującymi jako hosty do tworzenia serwerów wirtualnych pod kontrolą oprogramowania wirtualizacyjnego VMware Cloud Foundation Advanced oraz modułów pamięci masowej HPE Synergy D3940 z dyskami SSD.

## 2. Przedmiot zamówienia

### 2.1. Opis zapotrzebowania na oprogramowanie SPSS

Dotychczasowo wykorzystywane oprogramowanie SPSS w wersji 21 nie może być dłużej wykorzystywane z powodu zaprzestania wsparcia technicznego przez firmę IBM dla starych wersji IBM Statistics.

Analiza potrzeb Zamawiającego wskazuje na konieczność zmniejszenia posiadanej liczby licencji:

nazwa	Maksymalna liczba jednoczesnych użytkowników
IBM SPSS Statistics Base	30
IBM SPSS Advanced Statistics	4
IBM SPSS Complex Samples	4
IBM SPSS Custom Tables	16
IBM SPSS Exact Tests	4
IBM SPSS Forecasting	2
IBM SPSS Regression	2

## 3. Szczegółowa specyfikacja oprogramowania aktualizowanego

Zadanie polega na dostarczeniu Zamawiającemu bezterminowych licencji sieciowych, pozwalających na uruchomienie w środowisku terminalowym najnowszej dostępnej wersji oprogramowania SPSS zgodnie z określonym zapotrzebowaniem:

kod	nazwa	Ilość licencji
D0ELQLL	IBM SPSS Statistics Base concurrent user perpetual license	30
D0ELLLL	IBM SPSS Advanced Statistics concurrent user perpetual license	4
D0EL9LL	IBM SPSS Complex Samples concurrent user perpetual license	4
D0ELULL	IBM SPSS Custom Tables concurrent user perpetual license	16
D0EMELL-1	IBM SPSS Exact Tests concurrent user perpetual license	4
D0EM8LL	IBM SPSS Forecasting concurrent user perpetual license	2
D0ELL1L	IBM SPSS Regression concurrent user perpetual license	2

## 4. Wymagania dla oprogramowania równoważnego

Oprogramowanie równoważne musi spełniać poniższe wymagania:

### 4.1. Wymagania ogólne:

- 1) Licencja ważna na czas nieoznaczony.
- 2) Licencja na jednoczesnych użytkowników.
- 3) Wersja oprogramowania instalowana na serwerze terminalowym.
- 4) Oprogramowanie musi umożliwiać logowanie domenowe z wykorzystaniem mechanizmu Active Directory, bez konieczności kolejnego logowania.
- 5) Wersja oprogramowania możliwa do zarządzania poprzez serwer licencji.
- 6) Pełne wsparcie dla systemu Windows 10.
- 7) Roczny serwis techniczny dla licencji.  
Serwis techniczny musi obejmować dostarczenie nowej wersji oprogramowania, jeśli ta pojawi się w okresie 1 roku od dnia podpisania Umowy.
- 8) Oprogramowanie musi być wyposażone w samouczek – czyli zestaw lekcji multimedialnych umożliwiający samodzielną naukę programu.
- 9) Oprogramowanie musi:
  - a. pracować na danych zapisanych w formacie SPSS (pliki z rozszerzeniem „sav”),
  - b. wykorzystywać komendy języka SPSS i pliki poleceń SPSS (pliki z rozszerzeniem „sps”),
  - c. otwierać (także do edycji) oraz zapisywać raporty w formacie SPSS (pliki z rozszerzeniem „spv”),
  - d. umożliwiać korzystanie i edycję skryptów utworzonych w SPSS (pliki z rozszerzeniem „sbs”).

### 4.2. Wymagania w zakresie operacji na danych.

- 1) Oprogramowanie musi umożliwiać import danych w różnych formatach, w tym co najmniej: Excel (XLS, XLSX), SAS (SAS7BDAT), SPSS (SAV), oraz plików tekstowych za pomocą dedykowanego kreatora (TXT, DBF, ASCII, CSV).
- 2) Oprogramowanie musi zapewniać możliwość tworzenia wykresów z wykorzystaniem interaktywnego kreatora – obsługiwanego za pomocą mechanizmu „przeciągnij i upuść” zapewniającego podgląd układu wykresu w trakcie tworzenia.
- 3) W ramach interfejsu oprogramowanie powinno zapewniać, co najmniej przygotowanie następujących wykresów:
  - a. słupkowy,
  - b. histogram (w tym piramida populacyjna)
  - c. liniowy,
  - d. kołowy, pierścieniowy,
  - e. skrzynkowy,
  - f. rozrzutu, wielowymiarowy wykres rozrzutu,
  - g. radarowy, szeregu,
  - h. punktowy,
  - i. na mapach Polski (mapa gradientów, słupkowe, kołowe, liniowe). Podział na województwa, gminy i powiaty,
  - j. tabelowe, które umożliwiają połączenie wykresu (do wyboru co najmniej słupkowy, szeregu, histogram) i tabeli prezentującej statystyki (do wyboru co najmniej minimum, maksimum, średnia, mediana, odchylenie standardowe) w ramach jednego obiektu raportowego,
- 4) Oprogramowanie musi zapewniać dostęp do danych gromadzonych w ramach popularnych systemów bazodanowych za pomocą sterowników ODBC lub mechanizmu OLE DB. Dostęp do baz danych powinien być realizowany za pomocą kreatora pozwalającego tworzyć i zapisywać kwerendy w języku SQL.
- 5) Oprogramowanie musi dawać możliwość przeszukiwania i zastępowania w zbiorze danych określonych wartości (i ciągów tekstowych).
- 6) Musi istnieć możliwość zarządzania brakami danych co najmniej w zakresie:
  - a. zastępowanie braków danych wartościami średniej, mediany, dominanty,
  - b. wykorzystania mechanizmu wielokrotnej imputacji braków danych,

- c. wyłączenia wszystkich obserwacji z brakami.
- 7) Oprogramowanie musi zapewniać możliwość definiowania różnych rodzajów braków danych, w tym co najmniej braków odpowiadających odpowiedziom kwestionariuszy: "nie wiem", "nie dotyczy" i "odmowa odpowiedzi".
- 8) Oprogramowanie musi zapewniać możliwość kopiowania własności zmiennych w obrębie jednego zbioru danych oraz kopiowania własności pomiędzy zbiorami danych.
- 9) Musi być możliwe tworzenie zestawów wielokrotnych odpowiedzi, tj. obiektów łączących wartości pochodzących z różnych zmiennych, w których zakodowane są odpowiedzi na pytania wielokrotnego wyboru.
- 10) Oprogramowanie musi zapewniać możliwość podglądu z poziomu kreatora podstawowych informacji o zmiennej (typ zmiennej, etykieta zmiennej, poziom pomiaru, kategorie zmiennej - w przypadku zmiennych nominalnych i porządkowych). Utworzony zestaw wielokrotnych odpowiedzi musi być możliwy do użycia przy konstruowaniu obiektów tabelarycznych i budowaniu wykresów.
- 11) Oprogramowanie musi zapewniać następujące funkcjonalności w zakresie przekształcania i przygotowania danych:
  - a. Sortowanie rekordów na podstawie jednej lub większej liczby zmiennych.
  - b. Podział zbioru danych na części, tak by analizy były automatycznie wykonywane dla różnych grup jednostek analizy (np. jednostek terytorialnych).
  - c. Identyfikacja powtarzających się rekordów na podstawie pojedynczej zmiennej lub zestawu wybranych zmiennych.
  - d. Wyszukiwanie i usuwanie duplikatów rekordów oraz duplikatów zmiennych w zbiorze danych.
  - e. Selekcja losowa i warunkowa rekordów, permanentna lub tymczasowa - bez utraty zawartości zbioru danych (niewybrane rekordy pozostają nieaktywne - nie są brane pod uwagę podczas przekształcania danych i analizy).
  - f. Operacje wykonywane wg. zmiennych takie jak: zliczanie wystąpień określonych wartości lub przedziałów wartości w zestawie zmiennych, sumowanie wartości po zmiennych, wyciąganie wartości maksymalnej i minimalnej z zestawu zmiennych, wyliczanie średniej, mediany, odchylenia standardowego dla grup zmiennych.
  - g. Sprawdzanie poprawności danych poprzez:
    - i. weryfikację poprawności obserwacji obejmuje sprawdzenie identyfikatorów pod względem występowania braków danych i duplikatów oraz oznaczenie pustych obserwacji,
    - ii. weryfikację poprawności zmiennych obejmuje między innymi oznaczenie zmiennych z dużą ilością braków danych i zmiennych z dużą liczbą kategorii mało licznych,
    - iii. poszukiwanie błędów logicznych i wartości spoza dopuszczalnego zakresu,
    - iv. wyliczanie dla każdej obserwacji indeksu anomalii, który mierzy nietypowość obserwacji z uwzględnieniem jej przynależności do grupy o zbliżonych wartościach.
  - h. Restrukturyzacja zbioru danych, w tym co najmniej możliwe jest przekształcenie wybranych zmiennych w grupy powiązanych ze sobą obserwacji, przekształcenie grup obserwacji w zmienne oraz pełna transpozycja zbioru danych.
  - i. Agregacja zbioru danych z użyciem co najmniej takich miar jak: suma, średnie, mediana, minimum, maksimum, odchylenie standardowe, liczba agregowanych rekordów, pierwsza i ostatnia w grupie. W przypadku agregacji oprogramowanie powinno dawać możliwość przypisania jako nowych zmiennych do istniejącego zbioru danych miar wyliczonych z wykorzystaniem funkcjonalności agregacji.
  - j. Rekodowanie wartości zmiennych na podstawie liczebności kategorii w zbiorze danych, w taki sposób, że w wyniku rekodowania cecha najliczniej reprezentowana w danych otrzymuje najniższą/najwyższą wartość na nowej liście wartości, a kolejne wartości cechy w zależności od liczebności uzyskują kolejne wartości numeryczne.
  - k. Automatyczne rekodowanie cech małowartościowych w oparciu o zdefiniowane przez użytkownika kryteria, w celu ułatwienia analizy i bardziej przejrzystej prezentacji wyników. Na przykład kategorie cech, których reprezentacja w zbiorze jest mniejsza

- niż założona przez użytkownika zostają zgrupowane do kategorii „inne” (np. wszystkie wartości stanowiące poniżej X%, zostają zgrupowane w ramach oddzielnej kategorii).
- l. Kategoryzacja wizualna zmiennych ilościowych z możliwością automatycznego wyboru punktów podziału (w oparciu o wyznaczenie równolicznych kategorii, percentyli, punktów przedstawiających miary odchylenia standardowego) i możliwością tworzenia etykiet opisujących granice przedziałów dla utworzonych kategorii.
  - m. Automatyczne przekształcanie zmiennych tekstowych w zmienną numeryczną, gdzie kolejnym wartościom tekstowym np. "{Zmienna-A X", " Zmienna-A Y", " Zmienna-A Z" itd. są przypisywane kolejne wartości 1,2,3 itd. Jednocześnie z wartości tekstowych są automatycznie tworzone etykiety wartości.
  - n. Kodowanie dychotomiczne, które pozwala w oparciu o wartości wybranej cechy przyjmującej określone wartości utworzyć zestaw zmiennych dychotomicznych, zawierających wartość informacyjną cechy wyjściowej.
  - o. Operacje na datach wyciągnię z daty: dni, miesiący, kwartałów, roku, dnia tygodnia, dnia w roku, tygodnia. Tworzenie daty z oddzielnych zmiennych reprezentujących dzień, miesiąc, rok.
  - p. Kreator wyliczania nowych zmiennych przy użyciu funkcji arytmetycznych, daty i czasu, logicznych, statystycznych.
  - q. Ważenie rekordów wartościami wybranej zmiennej i zakresie rekordów na podstawie wybranych zmiennych.
  - r. Standaryzacji lub też normalizacji zmiennych poprzez wskazanie odpowiednio wartości średniej i odchylenia standardowego dla zmiennych standaryzowanych lub wartości minimalnej i maksymalnej dla zmiennych normalizowanych.
- 12) Oprogramowanie musi zapewniać co najmniej następujące techniki analityczne:
- a. regresja liniowa, estymacja krzywej regresji,
  - b. tabele krzyżowe z możliwością prezentacji miar współzależności zmiennych (liczebności obserwowane i liczebności oczekiwane, test proporcji kolumnowych, skorelowane zmienne V-Cramera),
  - c. jednoczynnikowa analiza wariancji,
  - d. korelacja zmiennych,
  - e. porównywanie średnich,
  - f. statystyki ilorazowe dla zmiennych ilościowych,
  - g. analiza dyskryminacyjna,
  - h. testy nieparametryczne: McNemara, Kołmogorowa-Smirnowa, Kendalla,
  - i. hierarchiczna analiza skupień i analiza skupień metodą k-średnich,
  - j. analiza czynnikowa i analiza głównych składowych,
  - k. analizy symulacyjne,
  - l. analiza miar nierówności.
- 13) Oprogramowanie musi umożliwiać konstruowanie tabel wykorzystujących trzy wymiary prezentacji danych: wiersze, kolumny i warstwy, w oparciu o interaktywny kreator – obsługiwany przy pomocy mechanizmu „przeciągnij i upuść” zapewniający podgląd układu tabeli w trakcie jej tworzenia. W ramach interfejsu oprogramowanie powinno zapewniać, co najmniej:
- a. możliwość sortowania wierszy etykiet lub statystyk,
  - b. możliwość prezentowania zestawów wielokrotnych odpowiedzi, obejmująca tworzenie tabel dla pytań kwestionariusza z możliwością wyboru wielu odpowiedzi,
  - c. możliwość umieszczania w tabeli więcej niż jednej zmiennej poprzez ustawianie zmiennych obok siebie i ich zagnieżdżanie,
  - d. umieszczania w wierszach, kolumnach i warstwach podsumowań w zależności od prezentowanych statystyk (liczebności, procenty, statystyki opisowe).
- 14) Oprogramowanie musi zapewniać następujące możliwości edycji utworzonych tabel i wykresów:
- a. edycji graficznej poszczególnych elementów tabeli oraz wykresu: wyświetlanie nagłówków, komentarzy, linii referencyjnych i innych elementów,

- b. automatyczne kolorowanie komórek tabeli (gradientowe - w zależności od natężenia cechy oraz warunkowe - dla ustalonych wartości, a także pomijania kolorowania przekątnej),
  - c. edytowania tabel oraz wykresów w zakresie: zmiany porządku kategorii (sortowanie na podstawie wartości, etykiet, statystyki, itp.),
  - d. możliwość tworzenia szablonów tabel i wykresów użytkownika oraz korzystania z predefiniowanych szablonów wizualizacji, umożliwiającą automatyzację raportowania z zachowaniem jednolitego standardu graficznego prezentacji wyników,
  - e. edytowanie elementów tabeli oraz wykresu i prezentowania statystyk w zależności od typu i charakteru wybranych zmiennych.
- 15) Oprogramowanie musi zapewniać możliwość importowania map.
  - 16) Wymagane jest, aby oprogramowanie dysponowało możliwością wizualizacji na mapach Polski, co najmniej w podziale na województwa i powiaty.
  - 17) Oprogramowanie musi zapewniać wygodny eksport obiektów wynikowych (tabel i wykresów), co najmniej do dokumentu Ms-Word, MS-Excel, MS-PowerPoint oraz do obiektów HTML. W przypadku wykresów konieczne jest zachowanie formatowania (kolory, linie i obramowania), a w przypadku tabel posiadających warstwy, zachowanie możliwości wyboru czy eksportowana ma być tylko pierwsza widoczna warstwa tabeli, czy wszystkie warstwy jako kolejne tabele.
  - 18) Oprogramowanie musi zapewniać następujący wachlarz zaawansowanych technik analitycznych takich jak:
    - a. regresja liniowa,
    - b. uogólnione modele liniowe,
    - c. regresja logistyczna,
    - d. uogólnione równania estymujące,
    - e. procedury ANOVA,
    - f. liniowe modele mieszane,
    - g. ogóle modele wielowymiarowych tabel kontyngencji dostępne z poziomu języka polecań (syntax),
  - 19) Oprogramowanie daje możliwość wykonywania testów dla jednej, dwu i k prób (zarówno zależnych, jak i niezależnych), testy zgodności, także testy miary siły związków.

#### 4.3. Wymagania w zakresie raportowania.

- 1) Oprogramowanie musi umożliwiać tworzenie raportów.
- 2) Oprogramowanie musi umożliwiać użycie komentarzy i notatek dodawanych do obiektów wynikowych jako opisów w treści raportu.
- 3) Musi istnieć możliwość eksportu gotowych raportów do formatu, przynajmniej: HTML i PDF.

## 5. Wsparcie

- 1) Dostarczone oprogramowanie objęte będzie dwunastomiesięcznym wsparciem technicznym producenta oprogramowania.
- 2) Wsparcie musi być dostępne w języku polskim.
- 3) Wsparcie obejmować będzie dostarczenie nowej wersji oprogramowania w razie jej opublikowania przez producenta oprogramowania.