

Marcin SKOWRONEK

Ewa STARZEWSKA-KARWAN

ANALSTAT – SYSTEM ANALIZY STATYSTYCZNEJ

Streszczenie. W artykule przedstawiono podstawowe cechy funkcjonalne, możliwości systemu i repertuar dostępnych badań systemu analizy statystycznej ANALSTAT. Zaprezentowano również zasadę przekazywania danych między programami systemu za pośrednictwem wspólnego obszaru pamięci operacyjnej.

ANALSTAT – THE SYSTEM OF STATISTICAL ANALYSIS

Summary. In the paper the basic functional features of the statistical analysis system ANALSTAT are presented, together with the description of its structure, possibilities and services. The rule of data passing between the system programs, using the common area in the operational memory, is also explained.

ANALSTAT – DAS SYSTEM DER STATISTISCHEN ANALYSE

Zusammenfassung. Im Artikel wurden die fundamentalen Merkmale, die Möglichkeiten und die Zusammenstellung der zugänglichen Forschungen des Systems der statistischen Analyse ANALSTAT vorgestellt. Das Verfahren der Datenübertragung zwischen den Systemprogrammen durch Vermittlung des gemeinsamen Speicherbereich wurde ebenso gezeigt.

1. Wprowadzenie

Statystyczna ocena danych liczbowych jest zadaniem często spotykanym w praktyce. Dane liczbowe mogą być wynikiem obserwacji zjawisk zachodzących w przyrodzie, wynikiem zaplanowanych eksperymentów albo też wynikiem ankiet lub testów socjologicznych. Do zapamiętywania tych wyników wykorzystywane są zazwyczaj pliki baz danych.

Wiele programów (systemów), takich jak systemy zarządzania bazami danych, programy obsługi arkuszy kalkulacyjnych czy też programy prezentacji graficznej wyników, ma wbudowane mechanizmy umożliwiające wykonywanie prostych badań statystycznych. Istnieją również specjalizowane systemy programowe przeznaczone wyłącznie do wykonywania analiz statystycznych i prezentacji graficznej danych liczbowych i wyników analizy statystycznej. Przykładem mogą tu być pakiet CSS oraz STATGRAPHICS [1].

Prezentowany system ANALSTAT jest przykładem prostego systemu, pracującego pod systemem operacyjnym MS-DOS, przeznaczonego do wykonywania analizy statystycznej i prezentacji graficznej wyników tej analizy, danych liczbowych zgromadzonych w plikach baz danych dBase, Informix lub Paradox.

System ANALSTAT był projektowany równolegle z systemem UNIVERS, umożliwiającym wyszukiwanie danych przy użyciu języka zapytań opartego na modelu relacji uniwersalnej [2]. Przyjęto więc, że ewentualna selekcja danych liczbowych poddawanych analizie statystycznej dokonana zostanie za pomocą systemu UNIVERS, udostępniającego użytkownikowi wysokiego poziomu język zapytań zbliżony w pewien sposób do języka naturalnego.

Interaktywna forma wyboru rodzaju analizy (badania) i deklarowania potrzebnych jej parametrów, możliwość przechowywania opisu badań i następnie powtórnej ich realizacji, a także możliwość graficznej prezentacji wyników badań stwarzają z systemu ANALSTAT wygodne narzędzie programowe, użyteczne szczególnie w różnorodnych badaniach naukowych.

Systemy UNIVERS i ANALSTAT są odrębnymi programami, które mogą być używane niezależnie od siebie bądź też wspólnie w celu rozwiązywania złożonych zadań. Zadania takie mogą dotyczyć analizy statystycznej danych rozmieszczonych w kilku plikach bądź też danych wymagających wstępnej selekcji, często przy uwzględnieniu powiązań między kilkoma plikami. Wtedy etap analizy statystycznej musi zostać poprzedzony operacją wyszukiwania, która może być zrealizowana przy użyciu systemu UNIVERS. Wyniki analizy statystycznej mogą być następnie prezentowane graficznie przy użyciu programu CHART2D będącego elementem systemu ANALSTAT. Łączne użycie obydwu tych systemów stwarza nowe jakościowo narzędzie badawcze. Pozwala ono bowiem na

bezpośrednie włączenie do analizy statystycznej różnorodnych kryteriów wyboru przetworzonych danych.

Systemy zostały wykonane w Instytucie Informatyki Politechniki Śląskiej. Prace nad obydwoma systemami rozpoczęto w ramach CPBR 8.6, ukończono zaś w ramach grantu Komitetu Badań Naukowych nr 8 0269 91 01 pt. *Rozwój narzędzi programowych dla mikrokomputerowych baz danych*.

2. Charakterystyka systemu

Podstawowym trybem pracy systemu ANALSTAT jest tryb interakcyjny, w którym użytkownik za pośrednictwem systemu menu, wspomaganego przez system podpowiedzi, wprowadza dane wejściowe, inicjuje badania, obserwuje i analizuje wyniki badań oraz dokonuje ich rejestracji.

Oprócz tego system posiada również mechanizmy umożliwiające:

- zapis danych wejściowych do pliku opisu badań,
- odczyt danych wejściowych z pliku opisu badań,
- automatyczny odczyt danych wejściowych z pliku opisu badań łącznie z wykonywaniem, obserwacją i rejestracją wyników badań.

Dzięki tym mechanizmom możliwe jest powtarzanie badań, bez konieczności ponownego wprowadzania danych wejściowych, jak również praca bez ingerencji użytkownika.

Dane wejściowe wprowadzane przez użytkownika systemu ANALSTAT, to:

- a) nazwy plików, w których będą lub są zapamiętane opisy badań oraz nazwy plików, w których będą zapamiętane wyniki badań,
- b) nazwy plików baz danych (relacji) i nazwy atrybutów, w których zgromadzone są wyniki eksperymentów (obserwacje) poddawane analizie statystycznej,
- c) dane szczegółowe dla poszczególnych badań,
- d) opcje sterujące wykonywaniem akcji odczytu, prezentacji i rejestracji.

Wyniki każdego wykonywanego badania zapamiętywane są w plikach roboczych. Użytkownik może je wielokrotnie przeglądać i zarejestrować ich zawartość na urządzeniu piszącym (drukarkę). Ponadto wyniki badań mogą zostać dopisane do odpowiednich plików wyjściowych. Oprócz wyników badań również parametry badań są zapamiętywane w pliku roboczym. Wyniki i parametry zapamiętane w plikach roboczych mogą być dopisane do odpowiednich plików wyjściowych. W systemie przewidziano 3 typy plików wyjściowych:

- pliki opisu badań (rozszerzenie .BDN),
- pliki wyników badań w postaci tekstowej (znakowej) (rozszerzenie .ZNK),

- pliki wyników pośrednich dla programu prezentacji graficznej (rozszerzenie .GRF).

Wspólna nazwa (bez rozszerzenia) tych plików jest jedną z danych wejściowych. System ANALSTAT umożliwia wcześniejsze przygotowanie danych typu b), c) i d) w pliku opisu badań oraz umożliwia odczytanie kolejnych zestawów takich danych z tzw. wejściowego pliku opisu badań, w tym samym lub innym przebiegu programu. Wprowadzanie nazw plików wejściowych i wyjściowych jest opcjonalne. Nazwy plików wyjściowych i wejściowego mogą być wielokrotnie zmieniane w czasie przebiegu programu.

Obligatoryjnymi danymi wejściowymi są nazwy relacji (pliki o rozszerzeniu .DBF, .DAT i .DB) i nazwy atrybutów, w których umieszczone są wyniki eksperymentów poddawane badaniom statystycznym. Dane te muszą być wprowadzone przed wykonaniem pierwszego badania. Wprowadzanie tych danych wejściowych sprowadza się do:

- określenia typu bazy danych,
- określenia ścieżki dostępu do katalogu zawierającego relacje wybranego typu bazy danych,
- wyboru określonego pliku (relacji) z listy wyświetlanych plików bazy danych,
- wyboru określonych atrybutów z wybranej relacji.

Atrybuty mogą pochodzić z różnych relacji. W czasie przebiegu programu można wielokrotnie modyfikować listę wybranych atrybutów. Wartości atrybutów (obserwacje określonych zmiennych eksperymentu) są wczytywane jednokrotnie po każdej modyfikacji listy wybranych atrybutów.

Większość oferowanych w systemie badań wymaga, oprócz określenia listy atrybutów i wyboru badania, wprowadzenia pewnych dodatkowych (szczegółowych) parametrów. Parametry te precyzują sposób realizacji badania, cechy atrybutów, np. wskazują, które obserwacje są wejściowe, a które wyjściowe lub określają sposób prezentacji wyników badania.

W trybie pracy interaktywnej użytkownik systemu ANALSTAT może inicjować pojedyncze akcje, takie jak:

- wykonanie badania,
- prezentacja wyników ostatniego badania,
- rejestracja wyników ostatniego badania,
- prezentacja wyników wcześniejszych badań,
- zapis wyników ostatniego badania,
- zapis opisu badań,
- odczyt zestawu opisu badań.

W większości przypadków po wykonaniu badania użytkownik będzie inicjował akcje prezentacji i/lub rejestracji wyników badań. System umożliwi automatyczne wykonanie ciągu akcji, zainicjowanego akcją wykonania badania. Taki tryb pracy umożliwiają opcje

sterujące pracą systemu. Użytkownik za pomocą opcji sterujących ma możliwość tworzenia określonych ciągów akcji prezentacji, rejestracji i odczytu zestawu opisu badań.

3. Struktura systemu i zasady współpracy programów

System ANALSTAT jest zbiorem programów, które można podzielić na:

- program komunikacji z użytkownikiem,
- programy realizujące badania statystyczne,
- programy prezentacji wyników badań.

Przyjęcie takiej struktury systemu wynikało przede wszystkim z chęci uniezależnienia wielkości kodu programu od liczby oferowanych badań statystycznych. W trakcie pracy systemu w pamięci operacyjnej umieszczony jest tylko program komunikacji i co najwyżej jeden dodatkowy program realizujący badanie lub prezentację wyników. Przy takim założeniu dodanie nowego badania tylko nieznacznie zwiększa kod programu komunikacji, gdyż kod związany z realizacją badania umieszczony jest w oddzielnym programie.

Rozwiązanie takie umożliwiło również wykorzystanie istniejących programów prezentacji (Pokaz.exe – prezentacja tekstu, Chart2d.exe – prezentacja wykresów) oraz ułatwiło proces uruchamiania systemu.

Program komunikacji z użytkownikiem jest podstawowym (głównym) programem systemu ANALSTAT. Do jego zadań, oprócz obsługi dialogu między systemem a użytkownikiem, należą:

- odczyt obserwacji z relacji,
- kontrola poprawności danych wejściowych,
- kompletowanie danych wejściowych i wywoływanie programów realizujących badania statystyczne,
- wywoływanie programów prezentacji wyników badań statystycznych,
- dopisywanie wyników do plików wyjściowych,
- odczyt danych wejściowych z plików opisu badań.

Na czas reakcji systemu (czas od momentu zainicjowania badania do momentu rozpoczęcia obserwacji wyników), oprócz czasu realizacji badania, istotny wpływ ma czas odczytu obserwacji oraz czas przekazywania danych wejściowych i wyników badania. W większości przypadków, dla jednego zestawu atrybutów wykonywanych jest kilka badań statystycznych. Przyjęto więc, że wartości atrybutów odczytywane będą jednorazowo przez program komunikacji i przekazywane do programu realizującego badanie łącznie z jego parametrami szczegółowymi. Celem skrócenia czasu przekazywania danych do programu

realizujących badania, zastosowano metodę przekazywania za pośrednictwem *wspólnego obszaru pamięci*.

```

Program Główny; {Szkielec programu głównego}
{$M 4096, 1024, 2048} {Dyrektywa określająca wielkość stosu i stogu}
Uses
  Dos;
Type
  TypDane = record
    par1: String;
    par2: Integer;
  end;
Var
  ior: Integer;
  dane: TypDane;

Procedure ExecProgram( prog: String; param: String );
{ Procedura bada, czy istnieje program o nazwie właściwej prog oraz wywołuje
  program z parametrem wejściowym param. Brak programu lub jego błędne dzia-
  łanie sygnalizowane jest przez zmienną ior }
Var
  command :String[127];
begin
  if ( FSearch( prog + '.exe', GetEnv( 'PATH' ) ) = '' ) and
    ( FSearch( prog + '.com', GetEnv( 'PATH' ) ) = '' ) then
    begin { brak programu }
      ior := 2; { kod braku programu }
    end
  else
    begin { wywołanie programu }
      command := prog + ' ' + param;
      SwapVectors;
      Exec( GetEnv( 'COMSPEC' ), '/C ' + command );
      SwapVectors;
      ior :=DosError; { kod wykonania programu }
    end;
end;

Procedure WolajProgram( prog: String; wskaz: Pointer );
{ Utworzenie łańcucha adr_rek (adres rekordu) na podstawie wskaźnika początku
  rekordu wskaz i wywołanie programu prog z parametrem adr_rek }
Type
  wordarr = array[1..2] of Word;
Var
  segm, offs: String[5];
  code:      Integer;
  tabl:      wordarr absolute wskaz;
  adr_rek:   String;
begin
  Str( tabl[1], segm ); { utworzenie napisu adresu segmentu }
  Str( tabl[2], offs ); { utworzenie napisu adresu przesunięcia }
  adr_rek := segm + ' ' + offs;
  ExecProgram( prog, adr_rek ); { wywołanie programu }
end;

begin
  dane.par1 := 'Przykładowy napis'; { Ustawienie przekazywanego parametru }
  dane.par2 := 13;
  WolajProgram ( 'Badanie', Addr( dane ) ); { Wywołanie programu Badanie }
end.

```

Rys. 1. Przekazywanie danych za pośrednictwem wspólnego obszaru pamięci

Fig. 1. The data passing using the common area in the operational memory

Istota tej metody polega na tym, że program główny umieszcza przekazywane dane w spójnym obszarze pamięci operacyjnej, na przykład w rekordzie, o strukturze znanej

programowi wywoływanemu, oraz przekazuje wywoływanemu programowi, jako parametry wywołania, adres tego rekordu. Program wywoływany na podstawie otrzymanych parametrów tworzy wskaźnik do rekordu przekazywanych danych wejściowych. Metoda ta może być również zastosowana do przekazywania wyników z programu realizującego badanie do programu głównego.

```

Program Badanie; { Szkielet programu wywoływanego }
Type
  TypDane = record { Typ rekordu wspólnego obszaru pamięci }
    par1: String;
    par2: Integer;
  end;
  TypWskRek = ^TypDane;
Var
  WskRek: TypWskRek; { Wskaźnik rekordu wspólnego obszaru pamięci }
  Wskaznik: pointer absolute WskRek;

Procedure UstawWskRek( Var wskaz: pointer);
{ Ustawienia wskaźnika rekordu danych wskaz, przekazywanego w parametrach
  wywołania programu }
Type
  tabadr= array[1..2] of word;
Var
  segm, offs: String[5];
  code: Integer;
  tab: tabadr absolute wskaz;
begin
  segm := ParamStr(1); { napis adresu segmentu }
  offs := ParamStr(2); { napis adresu przesunięcia }
  Val( segm, tab[1], code); { odtworzenie wskaźnika rekordu }
  Val( offs, tab[2], code);
end;

begin
  UstawWskRek( Wskaznik ); { Odtworzenie wskaźnika przekazywanego rekordu }
  { Przykładowe wykorzystanie przekazanych danych }
  Writeln ( 'par1: ', WskRek^.par1, ', par2: ', WskRek^.par2 );
end.

```

Rys. 2. Odtwarzanie wskaźnika wspólnego obszaru pamięci

Fig. 2. The pointer reproduction to the common area in the operational memory

Na rys. 1 przedstawiono szkielet programu głównego w języku Turbo Pascal ilustrującego metodę przekazywania danych za pośrednictwem wspólnego obszaru pamięci. Adres (wskaźnik) wspólnego obszaru pamięci zamieniany jest na dwa napisy: napis adresu segmentu i napis przesunięcia w segmencie, przekazywane następnie jako parametry wywołania programu. W programie wywoływanym (rys. 2), na podstawie parametrów wywołania, odtwarzany jest wskaźnik wspólnego obszaru pamięci.

Wyniki działania programów realizujących badania statystyczne przekazywane są do programu głównego za pośrednictwem plików roboczych. Rozwiązanie takie przyjęto ze względu na właściwości użytych programów prezentacji.

4. Charakterystyka dostępnych badań

System ANALSTAT służy do przeprowadzenia podstawowych badań statystycznych, takich jak: wyznaczenie podstawowych statystyk, histogramów, współczynników korelacji, równań regresji, zrealizowanie wybranych testów parametrycznych i nieparametrycznych, dokonanie analizy wariancyjnej, wyznaczenie funkcji autokowariancji, kowariancji wzajemnej, wygładzenie serii pomiarowej i prognozowanie jej wartości oraz przedstawienie funkcji gęstości prawdopodobieństwa i dystrybuanty w postaci tablicy lub wykresów, dla wybranych rozkładów.

Badania te zostały podzielone na 10 kategorii:

- elementarne statystyki,
- histogramy,
- korelacje,
- regresje,
- dystrybuanty,
- średnie,
- nieparametryczne,
- wariancje,
- serie czasowe,
- rozkłady.

Niektóre kategorie badań są złożone, np. w ramach kategorii **Histogramy** można zrealizować histogram jednej lub dwóch zmiennych. Jak już wspomniano, każde badanie jest oddzielnym programem. Opracowano 23 programy. Wyniki badania mogą być wyświetlone na monitorze lub zapisane do pliku. Niektóre wyniki można przedstawić w postaci graficznych wykresów. Dalej scharakteryzowano poszczególne badania. Ograniczono się do wymienienia wielkości wyznaczanych w badaniu. Szczegółowy opis algorytmów badań umieszczono w pracy [2].

1. Elementarne statystyki. W ramach tego badania wyznaczane są: średnia z próby, odchylenie standardowe, wartość minimalna i maksymalna wszystkich obserwacji z próby, suma wszystkich obserwacji w próbie.

2. Histogramy. W ramach tej kategorii badań można zrealizować histogram jednej lub dwóch zmiennych. W badaniu tym są również wyznaczane wartości niektórych statystyk.

W badaniu **Histogram jednej zmiennej** wyznaczane są następujące statystyki: suma wszystkich obserwacji, średnia z próby, odchylenie standardowe, wartości minimalna

i maksymalna dla wszystkich obserwacji w próbie, histogram, czyli empiryczna dystrybucja, funkcja gęstości prawdopodobieństwa (w procentach). Histogram może być przedstawiony w postaci wykresu.

W badaniu **Histogram dwóch zmiennych** wyznacza się częstość względną występowania par wartości dwóch badanych zmiennych w dwuwymiarowych przedziałach. Wyznaczane są również podstawowe statystyki oraz realizowany jest test niezależności chi-kwadrat. Statystyki te, to wyznaczone dla każdej zmiennej i każdego przedziału: suma wartości zmiennej w przedziale, średnia wartość zmiennej w przedziale, odchylenie standardowe zmiennej w przedziale. Dla testu niezależności dwóch zmiennych (test zgodności chi-kwadrat) obliczane są: wartość statystyki chi-kwadrat i liczba stopni swobody.

3. Korelacje. W ramach tej kategorii badań wyznaczane są współczynniki korelacji liniowej i potęgowej oraz podstawowe statystyki dla wskazanych przez użytkownika atrybutów.

W badaniu **Korelacja liniowa** obliczane są statystyki elementarne: wartość średnia, odchylenie standardowe oraz macierz współczynników korelacji.

W badaniu **Korelacja potęgowa** obliczane są takie statystyki, jak w korelacji liniowej, z tą tylko różnicą, że badanymi zmiennymi są: zmienna zależna, zmienna niezależna i jej kolejne potęgi.

4. Regresje [5]. W ramach tej kategorii badań wyznaczane są równania regresji dwuwymiarowej, wielokrotnej, potęgowej oraz podstawowe statystyki, wyznaczone dla wskazanych przez użytkownika atrybutów.

W badaniu **Regresja dwuwymiarowa** wyznacza się, dla wskazanych zmiennych, statystyki elementarne: wartość średnią, odchylenie standardowe, moment centralny trzeciego i czwartego rzędu, macierz współczynników korelacji. Dla każdej pary atrybutów wyznaczone są współczynniki równania regresji oraz standardowy błąd współczynników regresji. Proste regresje mogą być przedstawione w postaci wykresu.

W badaniu **Regresja wielokrotna** wyznaczane są wartości statystyk elementarnych dla wskazanych atrybutów: wartość średnia, odchylenie standardowe, macierz współczynników korelacji. Dla liniowego równania regresji wielokrotnej wyznaczane są: współczynniki regresji wielokrotnej, standardowy błąd współczynników regresji, wartość statystyki o rozkładzie Studenta, służącej do przeprowadzenia testu istotności dla każdego współczynnika regresji wielokrotnej, współczynnik korelacji wielokrotnej, standardowy błąd współczynnika korelacji wielokrotnej, suma kwadratów zmiennej w regresji i jej liczba stopni swobody, suma kwadratów zmiennej poza regresją i jej liczba stopni swobody, statystyka

o rozkładzie Snedecora, która służy do realizacji testu o zerowych wartościach współczynników regresji.

W badaniu **Regresja potęgowa** obliczane są takie statystyki, jak dla regresji wielokrotnej, przy czym zmiennymi niezależnymi w równaniu regresji mogą być zmienne bądź ich potęgi.

5. Dystrybuanty [3]. W badaniu realizowany jest nieparametryczny test zgodności Kołmogorowa. Dystrybuanta empiryczna porównywana jest z jednym z następujących rozkładów teoretycznych: normalnym, wykładniczym, Cauchy'ego i równomiernym. Wynikiem badania jest: wartość statystyki i prawdopodobieństwo określające poziom ufności hipotezy zgodności rozkładu.

6. Średnie. W badaniu obliczana jest wartość statystyki o rozkładzie Studenta i jej liczba stopni swobody. Parametry te służą do realizacji następujących czterech testów: testu istotności dla wartości średniej przy nie znanej wariancji, testu istotności dla wartości średniej w dwóch populacjach o identycznych, ale nie znanych wariancjach, testu istotności dla wartości średniej w dwóch populacjach o różnych i nie znanych wariancjach i testu istotności dla wartości średniej w dwóch populacjach, w których o wariancjach nic się nie zakłada.

7. Nieparametryczne [3, 8, 9, 10]. W tej kategorii badań realizowane są cztery testy: **Test Kendalla**, **Test Spearmana**, **Test Wilcoxona** i **Test Manna-Whitneya**, w których dla każdego elementu z próby określa się rangę będącą numerem elementu w uporządkowanym niemalejąco ciągu.

Test Kendalla jest testem istotności dla współczynnika korelacji rangowej Kendalla. W ramach tego testu obliczana jest wartość współczynnika korelacji rangowej Kendalla, jego standardowy błąd i wartość statystyki, która służy do testowania hipotezy o niezależności dwóch zmiennych losowych. Statystyka ta ma standaryzowany rozkład normalny.

Test Spearmana jest testem istotności dla współczynnika korelacji rangowej Spearmana. W ramach tego testu obliczana jest wartość współczynnika korelacji rangowej Spearmana, wartość statystyki, która służy do testowania hipotezy o niezależności dwóch zmiennych losowych i jej liczba stopni swobody. Statystyka ta ma rozkład Studenta.

Test Wilcoxona weryfikuje hipotezę, czy dwa ciągi obserwacji o jednakowej długości mają identyczny rozkład. W ramach tego testu obliczana jest wartość statystyki Wilcoxona oraz statystyki o standaryzowanym rozkładzie normalnym która służy do testowania weryfikowanej hipotezy.

Test Manna-Whitneya weryfikuje hipotezę, czy dwa ciągi obserwacji o różnej długości mają identyczny rozkład. W ramach tego testu obliczana jest wartość statystyki Manna-Whitneya oraz statystyki o standaryzowanym rozkładzie normalnym, która służy do testowania weryfikowanej hipotezy.

8. Wariancje [3, 6]. W badaniu przeprowadzona jest analiza wariancyjna dla 1, 2, 3, 4 lub 5 czynników. Wyznaczane są następujące parametry: wektor sum kwadratów dla poszczególnych kombinacji czynników, wektor liczby stopni swobody wyznaczonych wcześniej sum kwadratów dla poszczególnych kombinacji czynników, wektor wariancji empirycznej dla poszczególnych kombinacji czynników.

9. Serie czasowe. W tej kategorii badań realizowane są cztery badania: **Autokowariancja, Kowariancja, Wygładzanie, Prognozowanie.**

W badaniu **Autokowariancja** wyznaczana jest funkcja autokowariancji. Funkcja ta może być przedstawiona w formie wykresu.

W badaniu **Kowariancja** wyznaczane są funkcje kowariancji wzajemnej. Funkcje te mogą być przedstawione w formie wykresu.

W badaniu **Wygładzanie** wyznaczany jest ciąg wygładzonych wartości funkcji. Zastosowano w tym celu filtr nierekurencyjny, uśredniający. Wartości uśredniane mogą być selekcjonowane. Wyznaczona funkcja wygładzona może być przedstawiona w formie wykresu.

W badaniu **Prognozowanie** wyznacza się ciąg przyszłych wartości z wyprzedzeniem o jeden krok w stosunku do obserwowanego ciągu opierając się na algorytmie Browna [12, 13]. Jako funkcję prognozy przyjęto parabolę, której parametry są „dopasowywane” w każdym kroku do danych obserwowanych. Wyznaczoną funkcję prognozy można przedstawić w postaci wykresu.

10. Rozkłady. W tej kategorii badań wyznaczane są w formie tablicy lub wykresu teoretyczne dystrybuanty i funkcje gęstości prawdopodobieństwa dla rozkładu normalnego, wykładniczego, chi-kwadrat i beta.

5. Podsumowanie

Przedstawiony w artykule system ANALSTAT umożliwia analizę statystyczną danych liczbowych zgromadzonych w plikach baz danych różnych typów. Interakcyjny sposób

obsługi, wspomagany przez system podpowiedzi, czyni z systemu ANALSTAT proste i wygodne narzędzie badawcze, nie wymagające dużego nakładu na opanowanie jego obsługi. Te cechy, jak i oferowany zestaw badań stanowią, że system może być wykorzystywany nie tylko przez informatyków, ale np. przez lekarzy, socjologów, ekonomistów lub inżynierów.

Przyjęta koncepcja realizacji systemu jako zbioru niezależnych programów umożliwia łatwą rozbudowę systemu o nowe badania i dodatkowe formaty danych wejściowych, a równocześnie gwarantuje, że system nie ma wygórowanych wymagań pamięciowych. Podstawowa część systemu, tj. program komunikacji z użytkownikiem, wymaga aktualnie mniej niż 250 kB pamięci operacyjnej.

Program prezentacji graficznej, będący częścią składową systemu, umożliwia prezentację niektórych wyników badań w postaci histogramów, przebiegów schodkowych (konturowych) lub ciągłych.

Przedstawiona w artykule metoda przekazywania danych między programami może znaleźć zastosowanie przy realizacji innych, dużych systemów informatycznych, gwarantując w przyszłości łatwą ich rozbudowę.

LITERATURA

- [1] Dąbkowski J.: Statgraphics. Komputerowa Oficyna Wydawnicza "HELP", Warszawa 1992.
- [2] Kozielski S., Skowronek M. (red): System wyszukiwania i system analizy statystycznej w relacyjnych bazach danych. Skrypt Pol. Śl. nr 1814, Gliwice 1993.
- [3] Fisz M.: Rachunek prawdopodobieństwa i statystyka matematyczna. PWN, Warszawa 1969.
- [4] Zubrzycki S.: Wykłady z rachunku prawdopodobieństwa i statystyki matematycznej. PWN, Warszawa 1966.
- [5] Platt Cz.: Problemy rachunku prawdopodobieństwa i statystyki matematycznej. PWN, Warszawa 1977.
- [6] Brzeziński J., Stachowski R.: Zastosowanie analizy wariancji w eksperymentalnych badaniach psychologicznych. PWN, Warszawa 1984.
- [7] Hellwig Z.: Elementy rachunku prawdopodobieństwa i statystyki matematycznej. PWN, Warszawa 1967.
- [8] Zieliński R., Zieliński W.: Podręczne tablice statystyczne. PWN, Warszawa 1987.
- [9] Siegel S.: Nonparametric statistics for behavioral sciences. McGraw-Hill, New York 1956.

- [10] Zieliński R.: Siedem wykładów wprowadzających do statystyki matematycznej. PWN, Warszawa 1990.
- [11] Papoulis A.: Prawdopodobieństwo, zmienne losowe i procesy stochastyczne. WNT, Warszawa 1972.
- [12] Box G., Jenkins G.: Analiza szeregów czasowych. PWN, Warszawa 1983.
- [13] Brown R.: Smoothing, forecasting and prediction of discrete time series. Prentice-Hall, New York 1963.
- [14] Otnes R., Enochson L.: Analiza numeryczna szeregów czasowych. WNT, Warszawa 1978.

Recenzent: Dr hab. inż. Stanisław Wołek

Wpłynęło do Redakcji 1 października 1993 r.

Abstract

In the paper the basic functional features of the statistical analysis system ANALSTAT are presented, together with the description of its structure, possibilities and services. The system allows to read numerical data directly from the database files, performs statistical analysis and presents the results in the graphical form. The data for analysis is read from the files of the data bases like dBase, Informix or Paradox.

The system is built as a set of service programs, called by the main program (user communication program). The rules of data passing between main program and service programs used in the ANALSTAT system are explained on the example programs (fig. 1 and fig. 2).

The system contains 23 service programs, each of them performs the specific set of statistical analysis. The short characteristic and evaluated statistics of each service are also enclosed.

The rule of data passing between the system programs, using the common area in the operational memory, is also explained.