Marek PRZYBYŁ

WYKORZYSTANIE PAKIETU STATGRAPHICS DO ANALIZY PRZEMIAN JAKOŚCIOWYCH W CZASIE

Celem niniejszego artykułu jest wskazanie, że pakiet Stratgraphics jest narzędziem, dzięki któremu w sposób szybki i przystępny można przedstawić opis matematyczny i statystyczny procesów gospodarczych i ich prognozowanie. Oczywiście przy spełnieniu kilku uwarunkowań związanych z bazą danych.
Przedsiębiorstwo prowadząc swoją działalność, zakłada bazę danych, a następnie stara się zgromadzone dane analizować w różnorodny sposób m.in. pod względem statystycznym. Moment ten jest bardzo istotny ze względu na problem doboru oprogramowania, które będzie funkcjonowało w przedsiębiorstwie. Doświadczenie wskazuje, że głównym problemem użytkownika pakietu Stratgraphics nie jest technika obsługi i angielska terminologia, lecz prawidłowa identyfikacja merytorycznej zawartości poszczególnych procedur. Stratgraphics jest „przyjaznym" pakietem statystycznym z rozbudowaną stroną graficzną wysokiej jakości.

Jak wynika z praktyki, każde przedsiębiorstwo prowadząc swoją działalność, prędzej czy później zakłada bazę danych. W kolejnym etapie stara się zgromadzone dane analizować w różnorodny sposób między innymi pod względem statystycznym. Jest to bardzo istotne ze względu na dobór oprogramowania, które będzie funkcjonowało w przedsiębiorstwie. Wybór powinno ograniczyć się do oprogramowania z liczną dostępną literaturą w języku polskim i dosyć łatwą techniką obsługi. Doświadczenie wskazuje, że głównym problemem użytkownika pakietu Statgraphics nie jest technika obsługi i angielska terminologia, lecz prawidłowa identyfikacja merytorycznej zawartości poszczególnych procedur. Można więc powiedzieć, że Statgraphics jest „przyjaznym" pakietem statystycznym z rozbudowaną stroną graficzną wysokiej jakości.

Celem niniejszego artykułu jest wskazanie, że pakiet Statgraphics jest narzędziem, dzięki któremu w sposób szybki i przystępny można przedstawić opis matematyczny i statystyczny procesów gospodarczych i ich prognozowanie. Oczywiście przy spełnieniu kilku uwarunkowań związanych z bazą danych.

1. Struktury przestrzenne w identyfikacji zjawisk

Wydaje się często na starcie przedsięwzięcia, że danych jest pod dostatkiem – przede wszystkim w formie statystyk. Kiedy jednak sięgać po nie, rzadko bywają optymalne pod względem oceny i wykorzystania dla celów danego przedsięwzięcia. Istnieje w tym względzie przeważnie bezład danych, chociaż w przedsiębiorstwach, instytucjach i administracji publicznej występują permanentnie dane dotyczące informacji, konsultacji, analiz itp. Przemiana tych danych w abstrakcyjne twory zwykle dokonywana jest za pomocą wielowymiarowego skalowania. Postępowanie takie prowadzi do analizowania każdego obiektu jako punktu w przestrzeni wielowymiarowej o współrzędnych odpowiadających wartościom charakterystyk tych obiektów. Uogólniając te przemiany, otrzymujemy strukturę przestrzenną, która jest uporządkowanym i skończonym zbiorem elementów w przestrzeni cech. Struktura ta może przyjmować różną postać, zależną od liczby wymiarów i przyjętych skal wartości:

- jednowymiarowa dwuwartościowa (jedna cecha i dwie wartości), na przykład zakup towaru lub niezakupienie tego towaru,
- jednowymiarowa wielowartościowa (jedna cecha o wielu wartościach), na przykład konsument, jego wiek, wielkość zakupu, kwota zakupu, liczba zakupionych artykułów,
- dwuwymiarowa dwu- lub wielowartościowa (dwie cechy o dwóch lub wielu wartościach), na przykład zakup jako efekt ceny lub ilości,
- przestrzenna trójwymiarowa, gdzie elementy mogą tworzyć dowolne konfiguracje przestrzenne,
- przestrzenna cztero- i więcej wymiarowa, gdzie obraz fizycznej przestrzeni znika i w grę wchodzi interpretacja analityczna lub ograniczenie liczby cech i ich wartości oraz o przyporządkowanie ich badanym elementom. Na przykład badanie konsumpcji indywidualnej, uzależnia się ją od poziomu dochodów, cen, gustów, upodobań itp.

W każdej strukturze przestrzennej mamy do czynienia z niezależnością cech, czyli ortogonalnością współrzędnych. W przypadku cech zależnych następuje odkształcenie struktury wielowymiarowej i spłaszczenie obszaru identyfikowalności. Elementy znajdujące się w takim obszarze charakteryzują się nadmiarem informacji, natomiast elementy położone poza tym obszarem charakteryzują się

niedoinformowaniem. W przypadku identyfikacji elementów w ramach danej struktury przestrzennej mamy zatem do czynienia tylko w przypadku cech niezależnych, których geometrycznym obrazem jest tzw. przestrzeń ortogonalna o współrzędnych prostokątnych. Aby osiągnąć wymienioną przestrzeń, konieczne staje się zastąpienie większej liczby cech zależnych mniejszą liczbą cech (czynników) niezależnych. Zazwyczaj liczba tych czynników, które dają przestrzenie ortogonalne, nie przekracza trzech lub czterech, co zbliża te przestrzenie do trójwymiarowych przestrzeni realnych, lecz o nieidentycznym zakresie współrzędnych.

Dla osiągnięcia identycznego zakresu współrzędnych i nadania poszczególnym wymiarom jednakowej wagi interpretacyjnej dokonuje się tzw. rotacji ortogonalnej, polegającej na lepszym dopasowaniu osi czynnikowych do ortogonalnej przestrzeni cech, która jest po prostu przemianowaniem nieoznaczonych cech na czynniki lepiej odpowiadające wspólnym wpływom reprezentowanych przez nie cech.

Tego rodzaju postępowania sprawiają, że abstrakcyjna przestrzeń cech upodabnia się do przestrzeni fizycznej lub geograficznej.

2. Źródła badań

Warunkiem prowadzenia badań jest posiadanie dostatecznie szerokich informacji statystycznych dotyczących zarówno jej wielkości, jak i struktury oraz czynników ją kształtujących.

Źródła danych możemy podzielić na:

- pierwotne,
- wtórne (wydawnictwa statystyczne, inne opracowania zawierające wyniki badań innych instytucji).

Gromadzone w ten sposób dane umożliwiają prowadzenie bardzo ważnego nurtu badań, mianowicie szerokich porównań statystycznych w czasie i przestrzeni. Po wykorzystaniu źródeł wtórnych, w zależności od potrzeb sięgamy do źródeł pierwotnych:

- ewidencja i sprawozdawczość przedsiębiorstw,
- personel fachowy i operatywny przedsiębiorstw,
- inne w zależności od potrzeb.

Często wykorzystywanym źródłem informacji w badaniach przestrzennych konsumpcji jest sprawozdawczość rynkowa, obejmująca dane dotyczące:

- dostawy towarów na zaopatrzenie rynku w ujęciu wartościowym,
- sprzedaży detalicznej towarów,
- obrotów.

Trzeba sobie zdawać sprawę, że wartości te są dość grubym szacunkiem i nie obejmują informacji o ruchach towarowych w większości handlu nieuspołecznionego. Jednak obecnie sprawozdawczość rynkowa umożliwia obserwację zmian w strukturze konsumpcji, obliczanie współczynników elastyczności cenowej konsumpcji, konstruowanie wielu modeli przyczynowo-opisowych itp.

Ważnym źródłem informacji w badaniach statystycznych są liczne biuletyny wydawane przez regionalne biura doradcze.

3. Komputerowe banki danych w „magazynie danych"

Prowadzenie złożonych badań, w których zachodzi proces korzystania z tego samego zestawu informacji na różnych etapach analizy lub częste powtarzanie tego samego cyklu obliczeń przy różnych parametrach, jest ułatwione dzięki temu, że można korzystać z komputerowego banku danych. Obecnie, gdy prawie wszystkie obliczenia wykonywane są za pomocą komputerów, odpowiednio skonstruowany bank informacji w zdecydowany sposób wpływa na efektywność i dostępność do danych. Wszystkie dane są do dyspozycji użytkownika.

Zasoby tych banków danych, gdyby je uporządkować i zintegrować oraz umieścić w odpowiednim „magazynie danych", mogłyby służyć szybkiemu wejściu w operacyjną strukturę danych. Umożliwiłoby to zarządzanie ich zapasami, powstałyby bowiem wielkie możliwości ich wykorzystania poprzez ocenę ad hoc, wyłącznie dla określonego celu. Trafność wyboru byłaby zaś tym skuteczniejsza, im bardziej dostęp użytkownika do wszystkich danych, według swobodnych kryteriów i oceny, umożliwiłby wykorzystanie dla zadania odpowiedniej ilości sformalizowanych danych.

Konsolidacja i nagromadzenie danych w magazynie danych dla celów rozwoju systemów mikroregionalnych i związanych z nimi podstawowych decyzji strategicznych, mogłyby wpłynąć na dynamiczny rozwój „Mikroregionalnego Strategicznego Systemu Informacyjnego".

System ten powinien służyć potrzebom statystyki dla administracji publicznej, wszelkim innym przedsięwzięciom gospodarczym, wszystkim zainteresowanym podmiotom, będąc niezależnym od każdego innego systemu zarządzania oraz banków danych, które aktualnie egzystują.

Magazyn danych jest obszerną koncepcją, co pozwala zarządzać zapasami danych w taki sposób, że powstają nowe widoki i możliwości wykorzystania. Integracja nagromadzonych danych w celach szerokich przedsięwzięć w systemie zachowania danych, pozwala użytkownikowi na wkraczanie do wszystkich danych. Kryteria wyboru i oceny mogą tu być swobodnie definiowane, ponieważ restrykcje zostają pokonane przez normalizację danych i zależną od nich zastosowalność wstępnej strukturyzacji. Mimo to magazyn danych nie oznacza

w potocznym rozumieniu Software lub Software standard, bazuje on raczej na obrazie każdorazowego przedsięwzięcia, zewnętrznych i wewnętrznych stosunkach oraz na strukturach rozwoju. Standardowy Software nie może więc nie tylko nie nadawać się, on jest wprost nie pożądany.

Stopniowa standaryzacja danych Mikroregionalnego Strategicznego Systemu Informacyjnego umożliwiłaby wprowadzanie do projektowania i dokumentowania nowoczesnych (*Case-Tools*) rozwiązań usługowych, które powinny być przystosowane do potrzeb klientów i realizowane wraz z rozwojem innych oferentów.

Taka koncepcja usług musi zatem bazować na analizie informacyjnych wymogów klientów. Na tej bazie można by więc projektować ujęcia i funkcje użytkowników, a poprzez zintegrowaną koncepcję testową przygotowywać symulacje oraz wspierające działania logistyczne.

Z powyższych danych mogliby korzystać zarówno planiści, jak również decydenci transakcji, jeśli potrzebowaliby danych do nowych planów rozwojowych.

Reasumując, magazyn danych powinien być logiczną podstawą Mikroregionalnego Strategicznego Systemu Informacyjnego, który wyposażony w odpowiednio skupione informacje dotyczące ukierunkowanego rozwoju mikroregionu mógłby spełniać należycie swoją rolę.

W obecnej sytuacji działania wolnorynkowego, trudnością staje się pozyskiwanie pełnych, kompletnych informacji z działalności wielu przedsiębiorstw, co jest istotnym problemem w procesie budowy baz danych. Tworzące się bazy mogą być niekompletne, a co zatem idzie niewiarygodne, powstaje problem szacowania brakujących informacji.

W statystyce na szczęście można spotkać wiele procedur (interpolacyjnych) wykorzystywanych do szacowania brakujących danych, głównie dotyczą one szeregów czasowych.

Programy użytkowe w zasadzie nie są elementami banków danych ani magazynu danych, ale tworzą odrębny moduł systemu przetwarzania. Jednakże uzyskanie takiego modelu uzależnione jest głównie od liczby rodzajów programów, które współpracować będą w sposób bezpośredni z daną bazą. Takim programem między innymi może być pakiet statystyczny **Statgraphics.**

4. Analiza współzależności

W wielu przypadkach wzajemne powiązania i współzależności są wyraźnie widoczne. Nie trzeba żadnego pogłębionego rozważania ani stosowania szczególnych metod i środków, ażeby dostrzec związek, jaki istnieje między ilością opa-

dów atmosferycznych, a wielkością plonów w rolnictwie, między stanem techniki a wydajnością pracy itp.

Nie zawsze jednak ujawnienie współzależności poszczególnych zjawisk jest tak proste, a jej stwierdzenie wymaga zastosowania niezwykłych środków.

W szczególności staje się niezbędne posłużenie specjalnymi środkami statystyki matematycznej, jeżeli analiza współzależności zjawisk ma być ściślejsza i ma być określony jej stopień.

W tym zakresie pomocny jest pakiet statystyczny **STATGRAPHICS**. Zakres tematyczny procedur **Statgraphics'a**, które pojawią się po zainicjowaniu pracy z pakietem, zawiera bloki tematyczne, takie jak:

1) zarządzanie danymi i narzędzia systemowe,
2) kreślenie wykresów i statystyka opisowa,
3) analiza wariancji i regresji,
4) analiza szeregów czasowych,
5) procedury zaawansowanej statystyki,
6) procedury matematyczne.

Ogół procedur pakietu, których jest około 150, podzielono tu na 22 grupy (A do V), które są pojedynczymi pozycjami głównego menu.

Dane do pakietu można wprowadzać bezpośrednio, używając wewnętrznego edytora, w formacie swobodnym lub w postaci arkusza kalkulacyjnego albo importować praktycznie z wielu innych programów (Lotus 123, dBASE i wszystkie mające pliki tworzone w kodzie ASCII) za pomocą funkcji import.

Wprowadzone dane można dzielić, łączyć z innymi, dodawać nowe rekordy, kasować rekordy, pliki, zmienne, czyli pełny jest zakres działania na danych.

Dane statystyczne to zbiór informacji o cechach statystycznych jednostek pewnej zbiorowości lub szeregi czasowe. Informacje te są numeryczne, jeśli badana cecha jest mierzalna lub jakościowe, jeśli cecha jest niemierzalna.

Wprowadzone zmienne (cechy statystyczne) są to dane miesięczne za rok 1993 z województwa poznańskiego i przedstawiają (ODR Sielinko – Kalkulacje rolnicze i WUS Poznań) :

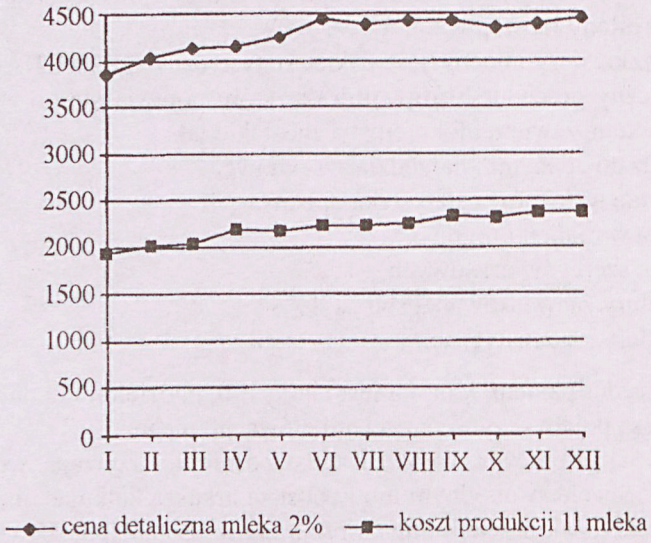
– koszt bezpośredni produkcji 1l mleka 4% – cecha niezależna,
– cena detaliczna 1l mleka 2 i 3% – cecha zależna.

Wartości cechy mierzalnej zaobserwowane dla n jednostek populacji statystycznej lub n okresów zapisuje się w pakiecie w postaci n-elementowego wektora – w edytorze danych jest to wektor kolumnowy.

Wektory i macierze określa się tutaj jako zmienne (*variables*), których nazwy służą do identyfikacji odpowiednich danych i ich deklarowania w procedurach obliczeniowych.

Każda zmienna należy do odpowiedniego zbioru danych, a zbiór danych tworzy dowolną liczbę zmiennych. Zmienne pochodzące z jednego badania lub będące przedmiotem wspólnej analizy grupuje się w jednym zbiorze. Zbiory

danych muszą być nazwane, a nazwa może się składać maksymalnie z ośmiu znaków alfanumerycznych. Przy przetwarzaniu w pakiecie pełną identyfikację zmiennych (cech statystycznych) dokonujemy przez podanie nazwy pliku i nazwy zmiennej, na przykład GRUPA.*wzrost,* przy czym sam plik z danymi figuruje pod nazwą GRUPA.ASF.



Rys. 1. Wykres cech statystycznych

Źródło: Opracowanie własne

Pakiet **Statgraphics** obejmuje wiele procedur, które wykonywane są według jednolitych reguł. Dotyczą one między innymi następujących operacji:

- deklarowania danych i założeń do analizy na ekranie wprowadzania danych,
- realizacji procedury (uruchomienie, wywołanie i wykonywanie dodatkowych opcji procedury),
- zapamiętywania wyników analizy w postaci zmiennych,
- drukowania i zapamiętywania ekranów tekstowych i graficznych,
- zapamiętywania ekranów do późniejszego odtwarzania (slaidy).

Po wybraniu procedury należy podać niezbędne informacje do wykonania procedury w polach, które nazywamy aktywnymi. Rodzaj wprowadzanych informacji jest zawsze określony przez program za pomocą napisów obok pola aktywnego.

Ogół metod statystycznych można podzielić na metody opisu statystycznego (*Deskriptive methods*), które służą wyłącznie do opisowej charakterystyki zbiorowości lub zjawisk oraz na metody wnioskowania statystycznego (*Statistical*

inference), które zajmują się stawianiem wniosków o populacji generalnej na podstawie próby losowej. Dział metod opisowych wykorzystuje jedynie metody opisu dotyczące rozkładu tylko jednej cechy w zbiorowości. Podstawą obliczeń i operacji są zawsze dane indywidualne.

Analiza regresji stanowi ważną część wszystkich metod badania współzależności zjawisk. Używa się ją tak do zmiennych ilościowych, jak i wyrażonych wartościowo zmiennych jakościowych. W analizie regresji wartości zmiennej zależnej (*dependent variable*) tłumaczy się za pomocą modelu matematycznego, zwanego modelem regresji. Jego integralną częścią jest funkcja regresji, która przedstawia zależność wartości średnich warunkowych zmiennej zależnej względem jednej zmiennej niezależnej (*simple regression*) lub względem kilku zmiennych niezależnych (*multiple regression*). Rozrzut wartości zmiennej zależnej wokół średnich warunkowych tłumaczony jest tutaj za pomocą zmiennej losowej nazywanej składnikiem losowym (*error term*).

Do podstawowych zagadnień analizy regresji należy dobór postaci modelu regresji, oszacowanie jego parametrów na podstawie danych empirycznych, ocena przyjmowanych założeń i wnioskowanie o jego parametrach. Utworzony „prawidłowy" model może być wykorzystany do przewidywania wartości zmiennej zależnej przy ustalonych wartościach zmiennych niezależnych. Wszystkie te procesy rozwiązywane mogą być w ramach omawianego pakietu **Statgraphics,** a uściślając w dziale *Regression Analysis*.

W ramach procedury regresji z jedną zmienną (*Simple Regression*) dokonujemy obliczeń modeli regresji z jedną zmienną niezależną. Związane one są z szacowaniem parametrów wybranej postaci modelu, oceny jakości modelu, wnioskowania o jego parametrach.

Na podstawie wspomnianych wcześniej danych (rys. 1), gdzie:

– cena detaliczna mleka 2% jest zmienną (cechą) zależną,

– koszt bezpośredni produkcji 1l mleka 2% jest zmienną (cechą) niezależną,

oszacujemy parametry modelu regresji kosztów bezpośrednich produkcji 1l mleka względem ceny detalicznej 1l mleka 2%.

Dane statystyczne zostały zapisane w pliku o nazwie A, a cechy statystyczne zwane dalej zmiennymi zależnymi i niezależnymi odpowiednio:

– mleko2,

– kosztmleka.

Po wypełnieniu powyższymi danymi tablicy procedury *Simple Regression* i uruchomieniu procedury otrzymujemy następujące obliczenia:

Na podstawie powyższych wyników możemy stwierdzić, że oszacowana na podstawie próby funkcja regresji ma postać

$$Y = 1548{,}48 + 1{,}23149X$$

gdzie 403,248 i 0,182295 stanowią standardowe błędy oceny parametrów.

Wynik estymacji liniowej regresji zmiennej mleko2 względem kosztmleka

Regression Analysis – Linear model: Y=a+bX					
Dependent variable: A.mleko2%			Independent variable: A.kosztmleka		
Parameter	Estimate	Standard Error	T Value	Prob. Level	
Intercept	1548.48	403.248	3.84002	.00326	
Slope	1.23149	0.182295	6.75549	.00005	
Analysis of Variance					
Source	Sum of squares	Df.	Mean Square	F-Ratio	Prob. Level
Model	339964.99	1	339964.99	45.6	.00005
Residual	74493.923	10	7449.392		
Total (Corr.) 414458.92 11					
Correlation Coefficient = 0.905683					
Stnd. Error of Est. = 86.3099			R - squared = 82.03		

Źródło: Opracowanie własne.

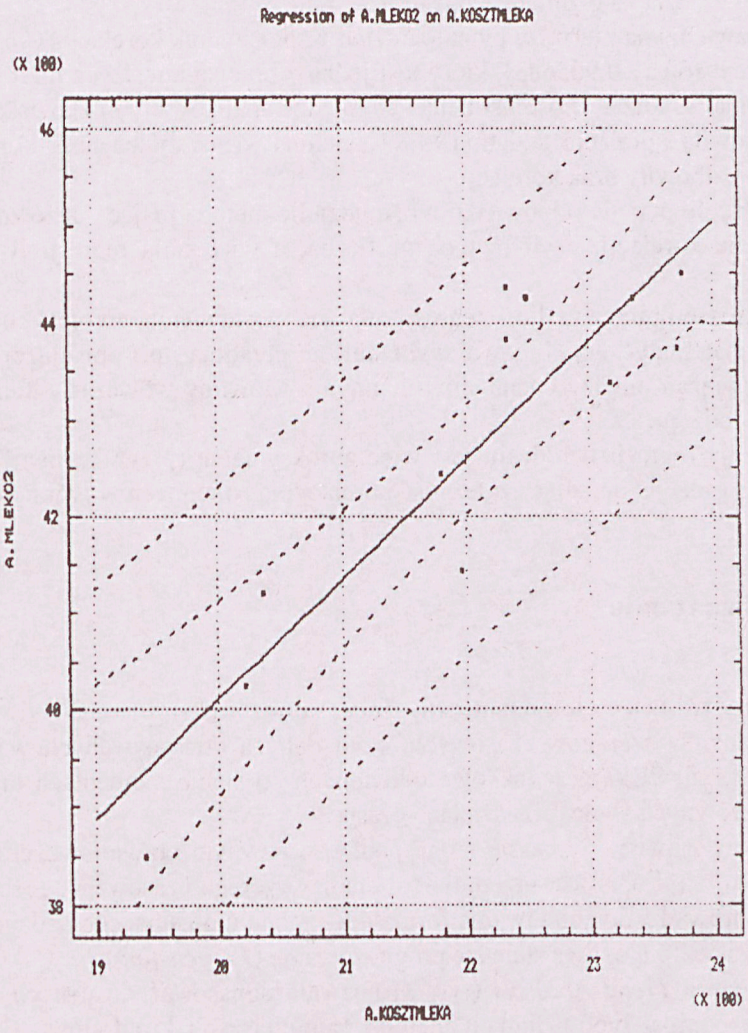
Obliczona dla współczynnika kierunkowego funkcji regresji wartość statystyki t - Studenta z pięcioma stopniami swobody zachodzi = 0,00005, czyli hipoteza o braku wpływu zmiennej niezależnej na zmienną Y ($H_0 : \alpha = 0$) jest prawie niemożliwa.

Wyniki analizy wariancji omawiane są w ramach modelu regresji. Poprzez dopasowanie linii prostej do danych empirycznych całkowita suma kwadratów odchyleń zmiennej zależnej od średniej (414458,92) podzielona została na sumę kwadratów odchyleń wyjaśnioną regresją liniową względem zmiennej niezależnej (339964,99) oraz sumę kwadratów niewyjaśnioną tą regresją (74493,923). Średni kwadrat odchyleń niewyjaśnionych regresją (7449,392) jest wariancją reszt. Dodatni pierwiastek z tej wielkości (86,3099) jest odchyleniem standardowym reszt (σ modelu) mówiącym o przeciętnej wielkości odchyleń zaobserwowanych wartości zmiennej zależnej od wartości „teoretycznych". Miarą dopasowania modelu do danych empirycznych jest współczynnik determinacji R^2 (82,03 %). Procedura *Simple Regression* przewiduje szacowanie trzech typów modeli nieliniowych sprowadzalnych przez odpowiednie transformacje do postaci liniowej. Wybór analitycznej postaci funkcji regresji II rodzaju nie jest sprawą łatwą. Decyzję o klasie funkcji należy podejmować nie tylko na podstawie wstępnej analizy materiału statystycznego, ale również na podstawie źródeł pozastatystycznych.

Współpraca z pakietem **Statgraphics** umożliwia bardzo szybkie „wpasowanie" modelu do danych empirycznych, rozpatrując jedynie współczynnik determinacji.

Funkcja regresji jest to analityczny wyraz przyporządkowania średnich wartości zmiennej zależnej konkretnym wartościom zmiennych niezależnych.

Posługiwanie się empiryczną krzywą regresji, jako przybliżeniem nieznanej krzywej regresji I rodzaju, okazuje się w praktyce niewystarczające. Wykres empirycznej krzywej regresji pozwala wprawdzie na wyrobienie sobie poglądu co do przybliżonego kształtu funkcji regresji I rodzaju, nie dostarcza jednak informacji o jej analitycznej postaci oraz o wartościach jej parametrów. Z tego względu empiryczna krzywa regresji nie może być narzędziem przewidywania,



Rys. 2. Regresja II rodzaju

Źródło: Opracowanie własne

jaką wartość przyjmie zmienna zależna przy ustalonych wartościach zmiennych uznanych za niezależne, którą uważa się zwykle za ostateczny cel analizy regresji.

Mówiąc o korelacji zjawisk, mamy na myśli sytuacje, w których wahania w jednym z nich wywołują tendencję do ich odpowiedniego kształtowania się w drugim zjawisku. Jeśli tak się dzieje, to znaczy, że zjawiska te znajdują się ze sobą w korelacji.

Celem analizy korelacji jest udzielenie odpowiedzi na pytania:

1) Jak wielki jest stopień powiązań analizowanych zjawisk?
2) Jaki wpływ wywiera cecha traktowana jako zmienna niezależna na cechę traktowaną jako zmienna zależna?

Odpowiedzi na pierwsze pytanie udzieli współczynnik korelacji (*Correlation Coefficient* równy 0,905683), który jest jedną z precyzyjniejszych miar określenia stopnia współzależności istniejących między analizowanymi zjawiskami.

Całkowita korelacja dodatnia to +1, ujemna –1. Współczynnik korelacji 0 oznacza całkowity brak korelacji.

Na drugie pytanie odpowiedzi może udzielić metoda mająca duże znaczenie w analizie korelacji, czyli metoda graficzna, a więc linia regresji II rodzaju (rys. 2).

Wykres empiryczny linii regresji ułatwia postawienie hipotezy na temat funkcji matematycznej (liniowa, wykładnicza, parabola itp.) opisującej mechanizm powiązań między badanymi zmiennymi. Mówimy wówczas o funkcji regresji II rodzaju.

Funkcja regresji II rodzaju jest więc aproksymantą (przybliżeniem) funkcji regresji I rodzaju, opisującą zależność korelacyjną zmiennych w próbie losowej.

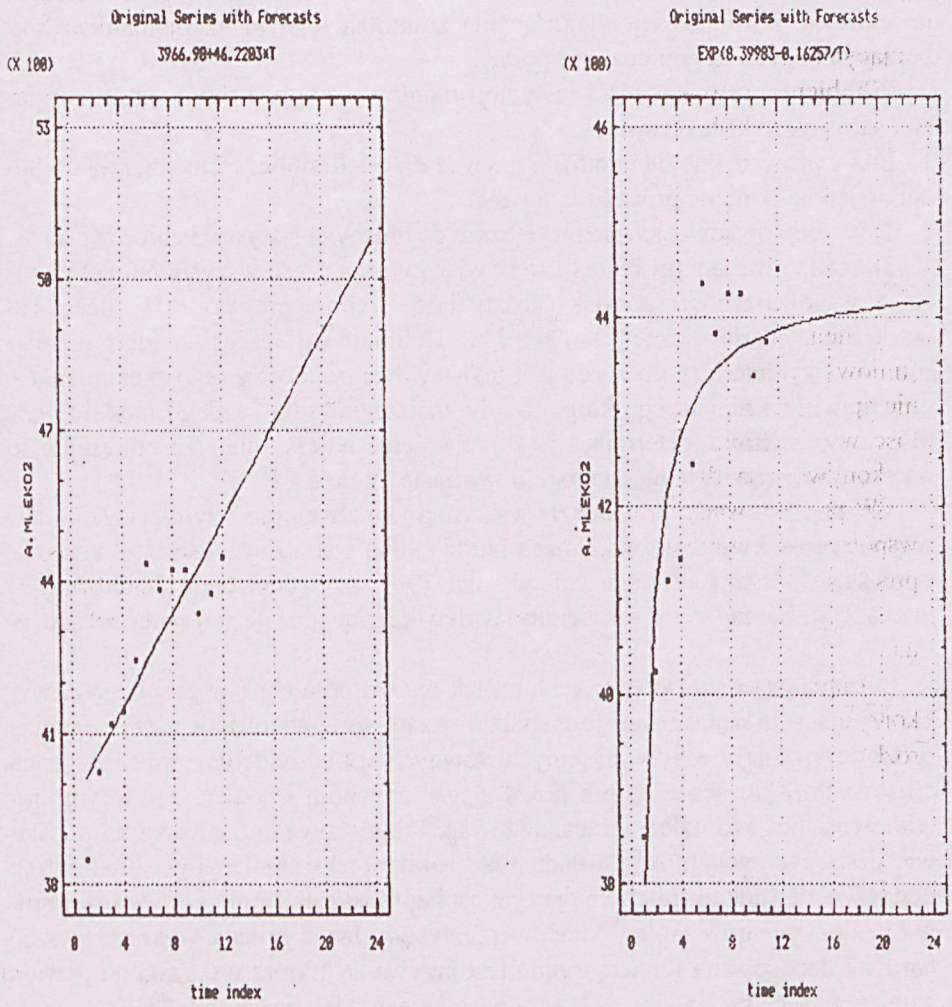
5. Analiza trendu

Procedury tworzące blok tematyczny *Times series procedures* mają zastosowanie do analizy szeregów czasowych, czyli danych otrzymywanych w wyniku obserwacji zjawiska w jednakowo oddalonych od siebie momentach lub w kolejnych, równych sobie, przedziałach czasu.

Zmiany zjawiska w czasie mogą podlegać pewnym prawidłowościom, których wykrycie i opis jest przedmiotem analizy szeregu czasowego. Jeżeli użyta metoda prowadzi do zbudowania formalnego modelu opisującego badane zjawisko w czasie, to możliwe staje się prognozowanie (*Forecasting*).

Procedura *Trend Analisys* (rys. 3) pozwala dopasować do danych szeregu czasowego kilka typów funkcji metodą najmniejszych kwadratów, a więc tą samą metodą, co w analizie regresji. Dopasowane funkcje opisują trend zjawiska, które jest ekstrapolowane na okresy przyszłe.

Wybór funkcji w ramach analizy trendu nie jest sprawą łatwą. Przyjmujemy funkcję liniową, gdy przyrosty absolutne w kolejnych okresach są w przybliżeniu stałe; funkcję wykładniczą, gdy stałe są przyrosty względne zjawiska. Krzywą logistyczną wykorzystujemy wówczas, gdy można przewidzieć wystąpienie poziomu nasycenia zjawiska. Ponownie, dopasowanie funkcji do danych empirycznych można przeprowadzić bardzo szybko, po oszacowaniu parametrów funkcji trendu. Wszystkie wyniki podawane są w jednostkach wyjściowych badanych cech statystycznych.



Rys. 3. Analiza trendu

Źródło: Opracowanie własne

6. Podsumowanie

Dla większości zagadnień związanych z produkcją i dystrybucją niezbędny jest matematyczno-statystyczny opis zjawisk gospodarczych. Informacja statystyczna jest tym cenniejsza, im obejmuje więcej danych – jeżeli przedstawiona jest w wielu przekrojach, za pomocą liczb i wykresów, a przede wszystkim jeżeli jest szybka i aktualna.

Dlatego też prace statystyczne należy wspomagać techniką komputerową, która ma bardzo dużo możliwości przetwarzania danych zarówno pod względem ilościowym, jak i jakościowym, łącznie z automatycznym uzupełnianiem brakujących danych za pomocą interpolacji.

Problem ten w większości instytucji mógłby rozwiązać z powodzeniem pakiet statystyczny **Statgraphics.**

Raz wprowadzone dane można używać do wielu obliczeń, wybierając tylko odpowiednie, kolejne procedury, a więc:

1) metody opisowe, służące wyłącznie do opisowej statystyki zbiorowości,

2) analizę wariancji i regresji stanowiącą ważną część wszystkich metod badania współzależności zjawisk (koszty bezpośrednie produkcji 1l mleka 4% względem ceny detalicznej 1l mleka 2%). Dobranie najlepszego modelu regresji (liniowy, wykładniczy itp.) – co jest podstawową trudnością tego zagadnienia – nie sprawia najmniejszego kłopotu, gdy analizujemy dla każdego modelu wartość współczynnika determinacji (R^2). Pamiętać należy tutaj, że obliczenia te wykonywane są, dla jednego modelu, w czasie około 5 - 10 s !

W prezentowanym przykładzie wskazano jedną z najprecyzyjniejszych miar, współczynnik korelacji, określającą bardzo silną współzależność 0,91 kosztów produkcji 1l mleka w ciągu całego roku 1993 względem ceny detalicznej 1l mleka 2% i bardzo silne wywieranie wpływu cechy niezależnej na cechę zależną.

Ostatnią przedstawioną procedurą, tak bardzo potrzebną w gospodarce, wykorzystującą te same dane, jest analiza trendu (prognozowanie), czyli analiza rozkładu badanej cechy w kolejnych, równych sobie, oddalonych przedziałach czasowych. Zjawiska w czasie mogą ulegać zmianom i podlegać pewnym prawidłowościom i to należy przeanalizować. Prognozowanie zjawiska na podstawie dopasowanych funkcji trendu ma charakter ekstrapolacyjny. Analityczne dopasowanie funkcji trendu mierzy jakości aproksymacji i bliższe wartości prognoz odczytujemy z tablicy końcowej statystyk. Jak wynika z wykresu (rys. 3) bardziej dopasowaną funkcją trendu jest krzywa „S", która wskazała na pewien stopień wysycenia zjawisk, w tym przypadku ceny detalicznej mleka 2%.

Pamiętajmy, że dane zawarte w tym opracowaniu zostały zebrane z jednego mikroregionu, a dane z innego mikroregionu mogłyby posłużyć do kolejnych obliczeń, scaleń lub porównań, bo duszą statystyki jest porównanie.

Reasumując, pakiet statystyczny **Statgraphics** jest w stanie w bardzo krótkim czasie przeprowadzić kompleksową analizę danego zjawiska, przy zastosowaniu różnych kryteriów – czasu, jakości i zależności zachodzących między nimi.

Literatura

[1] **Abt S.;** Systemy logistyczne w gospodarowaniu. Teoria i praktyka logistyki, Wydawnictwo Akademii Ekonomicznej, Poznań 1995
[2] **Abt S., Przybył M.:** Der Nahverker in City – Logistik – Aspekten, Frankfurter Institut für umweltorientierte Logistik e. V. , referat, Frankfurt n.Odrą 1995
[3] **Mynarski S.;** Badania przestrzenne rynku i konsumpcji. Przewodnik metodyczny, PWN, Warszawa 1992
[4] **Podgórski J.;** Statystyka z komputerem, MIKOM, Warszawa 1995

UTILIZATION OF PACK STRATGRAPHICS TO ANALYSISE OF QUALITATIVE TRANSFORMATIONS IN TIME

Aim of present article is indication, that pack Statgraphics is tool thanks to which in manner quick and accessible one can introduce description mathematical and of statistical processes economic and their forecasting. Of course of realization several of conditionings connected with base given.

Firms driving one's own activity founds base given and then old asssembled given to analyse in various manners among other things in respect of statistical. This is very essential from regard on problem of selection of software, which will work in firm. Experience points, that main problem of user of pack Statgraphics is not tevchnics ofservice and English nomenclature, but correct identification essential contents of each procedures. Statgraphics is "friendly" statistical pack with complex graphic side high qualities.