

*Jacek Bialek**

AUTORSKI PROGRAM KOMPUTEROWY STAT_STUD JAKO NARZĘDZIE WSPOMAGAJĄCE NAUCZANIE STATYSTYKI

1. WPROWADZENIE

W ostatnim czasie obserwujemy dynamiczny rozwój możliwości wykorzystywania programów komputerowych w procesie nauczania przedmiotów ilościowych, w szczególności statystyki. I choć początkowo możliwości te były ograniczone jedynie do multimedialnych prezentacji prowadzonych podczas zajęć, dziś w powszechnym użyciu są takie pakiety statystyczne jak SPSS, Statistica, Gauss czy środowisko R. Korzystanie z takiego oprogramowania nie tylko sprawia, iż zajęcia stają się atrakcyjniejsze dla studentów, ale również, poprzez wizualizację często skomplikowanych rachunkowo problemów, pozwalają studentom lepiej zrozumieć istotę omawianych zagadnień. Nikt właściwie nie ma wątpliwości, co do doniosłej roli, jakie tego typu oprogramowanie odgrywa we współczesnym nauczaniu statystyki i innych przedmiotów ilościowych (por. Kończak 2008, s. 63–73), Kufel 2008, s. 75–85). Co więcej, w dobie globalizacji, a co za tym idzie rosnącej międzynarodowej wymiany informacyjnej, towarzyszący jej rozwój narzędzi i technik informatycznych zmusza niejako do wprowadzania tego typu rozwiązań również i w dydaktyce.

Należy jednak zaznaczyć, iż często, mimo szczerych chęci wprowadzenia tego typu oprogramowania do zajęć, uczelni nie stać na zakup odpowiednich licencji, a już na pewno odpowiednio częstych ich aktualizacji. Ale też nie zawsze jest taka konieczność – niekiedy bowiem, profesjonalny program „potrafi” o wiele więcej, niż wymagają tego treści programowe danego przedmiotu ilościowego. Być może zatem, stawiając przed sobą cele wyłącznie dydaktyczne, a nie badawcze i naukowe, zakup oprogramowania o tak szerokich możliwościach jest niekonieczny.

Jednym ze sposobów ominięcia omawianych wyżej barier jest oczywiście korzystanie z programów typu freeware (darmowych bez możliwości ingerencji

* Dr inż., Katedra Metod Statystycznych, Uniwersytet Łódzki.

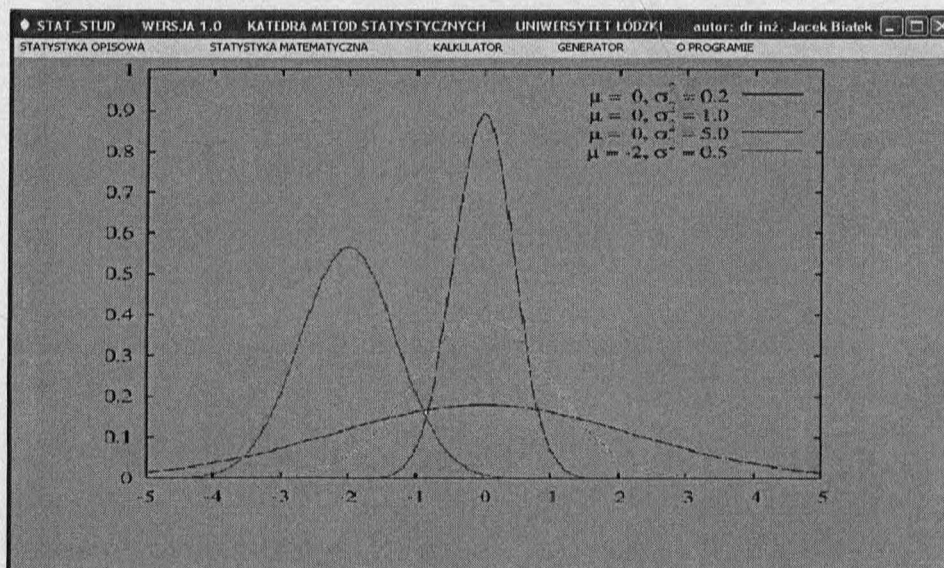
w kod) bądź na licencji GPL (General Public License). Tę drugą opcję stanowią oprogramowania typu Open Source, czyli na Powszechnej Licencji Publicznej GNU z możliwością ingerencji użytkowników w kod źródłowy. Takim właśnie programem jest np. ekonometryczny GRET, popularyzowany również i w Polsce (por. Kufel 2007). Jednak może się zdarzyć, iż programy tego typu nie zawierają wszystkich metod i narzędzi, jakie w ramach prowadzonych zajęć chcielibyśmy omówić. Oczywiście, w przypadku programów na „licencji otwartej” (Open Source) jest możliwość modyfikacji kodu i dopisania nowych, odpowiadających nam procedur i funkcji. Jednakże wymaga to specjalistycznej wiedzy w zakresie języka programowania, w jakim stworzono dane oprogramowanie, no i oczywiście czasu. Dlatego innym wyjściem, w przypadku znajomości któregoś z języków programistycznych, jest napisanie takiego programu od początku. Jest to oczywiście odpowiedzialne i czasochłonne przedsięwzięcie, ale ukierunkowane z góry na cele czysto dydaktyczne, w szczególności na te metody, które omawia się zgodnie z treściami nauczania danego przedmiotu ilościowego. Poza satysfakcją, napisanie takiego programu i jego rozpowszechnienie może sprawić, iż jednostki dydaktyczne, których nie stać na zakup drogiego, komercyjnego oprogramowania, będą miały mimo to możliwość przeprowadzenia demonstracji komputerowych na zajęciach. Co więcej, sami studenci, poznając kolejne partie materiału, a później ucząc się do egzaminów, pracując na domowych komputerach nie będą musieli używać pirackiego oprogramowania (bo jest taka pokusa w dobie Internetu) lecz właściwie darmowego, przyjaznego środowiska programistycznego. Właśnie ta argumentacja sprawiła, iż zamierzeniem autora od roku 2007 było napisanie programu wspomagającego nauczanie statystyki. I tak, w 2008 roku, powstał program STAT_STUD ver. 1.0.

2. PROGRAM STAT_STUD

Program STAT_STUD został napisany w 2008 roku, a jego głównym celem miało być od początku wspomaganie procesu nauczania statystyki (stąd zresztą jego nazwa, od „statystyka dla studentów”). Planuje się, iż zostanie on jeszcze w 2009 roku wydany na płycie dołączonej do książki pt. „Statystyka dla studentów z programem STAT_STUD”, której drugim współautorem jest dr Adam Depta. Idea twórcy programu jest taka, aby był on: funkcjonalny, prosty w obsłudze i właściwie darmowy (licencję na program będzie posiadał każdy, kto zakupi go wraz z książką). Oczywiście jego możliwości są ograniczone jedynie do podstawowych zagadnień omawianych na przedmiocie „statystyka” i pod tym względem nie może konkurować z żadnym z wymienionych wcześniej programów. Jednakże, ten niewielkich rozmiarów programik (główny plik nie przekracza 3 MB na dysku komputera), pozwala rozwiązać większość zagadnień jakie omawia się w ramach statystyki.

Program STAT_STUD został napisany w środowisku Turbo C++, co powoduje, iż może być (nawet komercyjnie) rozpowszechniany bez zakupu dodatkowych licencji.

Co więcej, został tak skompilowany, aby jego użytkowanie nie wymagało od użytkownika instalowania dodatkowych bibliotek. W programie tym wprowadzamy dane ręcznie, a w przypadku szeregu szczegółowego mamy możliwość wczytania danych z pliku CSV. Program składa się z następujących modułów: *STATYSTYKA OPISOWA*, *STATYSTYKA MATEMATYCZNA*, *KALKULATOR* I *GENERATOR*. Okno dialogowe programu przedstawia rys. 1.



Rys 1. Okno dialogowe w programie STAT_STUD

Źródło: Opracowanie własne.

Ponieważ cele programu mają charakter czysto dydaktyczny, dlatego wprowadzono w nim kilka udogodnień dla studentów-użytkowników. Po pierwsze, w przypadku błędnego wprowadzenia danych, bądź próby np. korzystania z modelu przy niespełnionych jego założeniach, program natychmiast o tym informuje. A zatem użytkując program, student coraz lepiej powinien kojarzyć konkretną metodę statystyczną, z właściwymi jej założeniami i ograniczeniami. Po drugie, program zawiera również pewną dozę teorii – dla niektórych zagadnień (choćby przy estymacji) student może nie tylko oszacować, ale również poznać matematyczną postać przedziału ufności. Szacując wartości krytyczne dla statystyk testowych, można zapytać o liczbę stopni swobody, można nimi manipulować obserwując ich wpływ, możemy również obserwować wpływ poziomu ufności na szerokość przedziału ufności. Po trzecie, w przypadku bardziej

zaawansowanych metod oprogramowanych w STAT_STUD, pod przyciskiem INFO użytkownik znajdzie informację o ich twórcy, a niekiedy publikacji, w której się po raz pierwszy opisano dany algorytm (patrz np. *Kalkulator testu normalności Shapiro-Wilka*). Po czwarte wreszcie, używając modułu *STATYSTYKA MATEMATYCZNA*, użytkownik może zarówno wygenerować wartość kwantyla określonego rzędu dla wybranego rozkładu prawdopodobieństwa (nie ma więc potrzeby odczytywania go z tablic statystycznych), jak i poprosić o postawienie wniosku statystycznego dotyczącego weryfikowanych hipotez. Wyniki uzyskiwane w poszczególnych modułach możemy raportować, zapisując je do pliku o formacie *doc*, bądź *xls* (w przypadku *GENERATORA*).

3. OPIS MODUŁÓW PROGRAMU STAT_STUD

Moduł *STATYSTYKA OPISOWA* składa się z następujących sekcji: *Analiza struktury*, *Analiza korelacji*, *Analiza regresji* i *Analiza dynamiki*. Dodatkowo, by ułatwić prowadzenie ćwiczeń ze statystyki, *Analiza struktury* podzielona jest na składowe: *SZEREG SZCZEGÓŁOWY*, *SZEREG ROZDZIELCZY PUNKTOWY*, *SZEREG ROZDZIELCZY PRZEDZIAŁOWY*. Moduł związany z analizą struktury zawiera większość miar, jakie się tu omawia z podziałem na klasyczne i pozycyjne. Poniżej zamieszczono rozwiązanie przykładowego zadania wykonane w STAT_STUD.

The screenshot shows a dialog box titled "SZEREG SZCZEGÓŁOWY" with the following elements:

- WPROWADZANIE DANYCH:**
 - NR PORZĄDKOWY: 16
 - WARTOŚĆ: [empty field]
 - AKCEPTUJ
 - PRZELICZ
 - ANULUJ
 - WCZYTAJ DANE ?
- ANALIZA STRUKTURY - MIARY KLASYCZNE:**
 - ŚREDNIA ARYTM.: 2,460
 - ODCHYLENIE STD.: 0,752
 - ODCHYLENIE PRZEC.: 0,579
 - WARIANCJA: 0,565
 - WSP. ZMIENNOŚCI / STD.: 0,306
 - WSP. ZMIENNOŚCI / PRZEC.: 0,235
 - WSP. ASYMETRII: 0,457
 - KURTOZA: 2,724
 - EKSCESS: -0,276
- ANALIZA STRUKTURY - MIARY POZYCYJNE:**
 - MEDIANA: 2,400
 - KWARTYL 1: 1,900
 - KWARTYL 3: 2,900
 - DOMINANTA: 2,400
 - ODCHYLENIE ĆW.: 0,500
 - WSP. ZMIENNOŚCI / ĆW.: 0,208
 - WSP. ASYMETRII: 0,000
- DANE:** A vertical list of values: 2,4, 2,5, 3,9, 2,4, 1,7, 2,1, 2,0, 3,2, 2,4, 1,1, 1,9, 1,8, 2,9, 3,9, 2,7.
- Buttons:** RAPORT, ZAMKNIJ.
- Options:** NORMALNA PRECYZJA, WYSOKA PRECYZJA.

Rys 2. Okno dialogowe sekcji *Analiza struktury* w programie STAT_STUD

Źródło: Opracowanie własne.

W ramach *Analizy korelacji* w programie przewidziano możliwość analizy zarówno danych opisanych *szeregiem korelacyjnym* (miary Pearsona i Spearmana wraz z rangowaniem) jak i dane ujęte w *tablicy korelacyjnej*. Wprowadzanie danych do tablicy korelacyjnej jest tu bardzo intuicyjne, program zaś liczy wszystkie omawiane na studiach miary oparte na statystyce chi-kwadrat, stosunki korelacyjne, a także współczynnik korelacji liniowej Pearsona. Obrazuje to rys. 3.

Rys 3. Okno dialogowe sekcji *Analiza korelacji* w programie STAT_STUD

Źródło: Opracowanie własne.

Dodatkowo, co odróżnia program od większości innych, w sekcji *Analiza dynamiki*, obok zakładki *Analiza trendu* figuruje również zakładka *Indeksy agregatowe*. Mamy tu możliwość wyznaczenia indeksu wartości dla danego agregatu, jak również cenowych i ilościowych agregatowych indeksów Laspeyresa, Paaschego i Fishera. W dowolnej chwili użytkownika programu, można przełączyć tzw. *normalną precyzję* wyników (3 miejsca po przecinku) na *wysoką precyzję* (15 miejsc po przecinku).

Moduł **STATYSTYKA MATEMATYCZNA** składa się dwóch zasadniczych sekcji: *Estymacja przedziałowa* i *Testowanie hipotez statystycznych*. W ramach estymacji przedziałowej, student ma możliwość oszacowania przedziałów ufności dla średniej, odchylenia standardowego i wariancji, wskaźnika struktury, współczynnika korelacji liniowej i współczynników regresji liniowej. Każda z tych sekcji zawiera dokładny opis założeń towarzyszących danemu modelowi danych, czy danej metodzie. Co więcej, mamy możliwość wykonywanie operacji obliczeniowych na kilku, jednocześnie aktywnych oknach dialogowych. Sprawia to, iż użytkownik może dokonać porównania wyników pochodzących z różnych modeli oraz przenosić wyniki uzyskane już w jednym z modułów do kolejnego. Poniżej umieszczono obliczenia przeprowadzone w ramach zakładki *Estymacja przedziałowa* -> *Przedział ufności dla średniej*:

PRZEDZIAŁY UFNOŚCI DLA ŚREDNIEJ			PRZEDZIAŁY UFNOŚCI			
PRZEDZIAŁY UFNOŚCI DLA ŚREDNIEJ			MODEL I	MODEL II	MODEL III	
średnia z próby =	20		90 %	[18,3540 ; 21,6452]	[18,2957 ; 21,7043]	[18,3540 ; 21,6452]
sigma =	5	ANULUJ	95 %	[18,8365 ; 21,4035]	[18,5871 ; 21,4429]	[18,8365 ; 21,4035]
liczba obs. n =	100	PRZEŁICZ	99 %	[18,7121 ; 21,2879]	[18,6402 ; 21,3198]	[18,7121 ; 21,2879]
MODEL II			90 %	[18,8368 ; 21,1632]	[18,8117 ; 21,1883]	[18,8368 ; 21,1632]
średnia z próby =	20		95 %	[19,0200 ; 20,9800]	[19,0029 ; 20,9971]	[19,0200 ; 20,9800]
odchylenie std z próby s =	5	ANULUJ	99 %	[19,1776 ; 20,8224]	[19,1656 ; 20,8344]	[19,1776 ; 20,8224]
liczba obs. n =	100	PRZEŁICZ	90 %	[19,3682 ; 20,6408]	[19,3517 ; 20,6483]	[19,3682 ; 20,6408]
MODEL III (eksperymentalny)						
Ciepło w populacji ma dwa razy tyle, że temperatura jest jednaki				RAPORT		
średnia z próby =	20			ANULUJ WSZYSTKO		
odchylenie std z próby s =	5	ANULUJ				
liczba obs. n =	100	PRZEŁICZ		ZAMKNIJ		

Rys 4. Okno dialogowe sekcji *Przedział ufności dla średniej* w programie STAT_STUD

Źródło: Opracowanie własne.

Sekcja *Testowanie hipotez statystycznych* składa się z dwóch części: *Testy parametryczne* (test dla średniej, test dla dwóch średnich, test dla wariancji, test dla dwóch wariancji, test dla wskaźnika struktury, test dla dwóch wskaźników struktury, test dla współczynnika korelacji, test dla stosunku korelacji, test dla współczynników regresji liniowej) i *Testy nieparametryczne* (test serii losowości, test serii zgodności, test niezależności chi-kwadrat, test zgodności chi-kwadrat, test par Wilcoxona, test normalności Shapiro-Wilka, test Kołmogorowa-Smirnowa). A więc użyteczność programu jest tu dość wysoka, trudno

bowiem znaleźć niektóre spośród wymienionych testów nieparametrycznych w dostępnych pakietach. Użytkowanie tego modułu jest równie proste i intuicyjne jak pozostałych, dla przykładu zaprezentowano wyniki obliczeń dla testu par Wilcoxona (test rangowanych znaków, por. Domański 2001) przeprowadzone w STAT_STUD.

TEST PAR WILCOXONA (TEST RANGOWANYCH ZNAKÓW)

Ho: F1 = F2 TEST ZGODNOŚCI DLA DWÓCH ROZKŁADÓW CIĄGŁYCH
 H1: F1 <> F2 DLA ZMIENNYCH POŁĄCZONYCH

WPROWADZANIE KOMPLETNYCH DANYCH INFO RAPORT ZAMKNIJ

NR. PORZĄDKOWY obserwacja Xi obserwacja Yi

7 AKCEPTUJ

OBSZAR KRYTYCZNY ? PRZELICZ ANULUJ

obserwacja Xi	obserwacja Yi	Xi-Yi	ABS (Xi-Yi)	Ri	Ri +	Ri -
22	22	0,0000	0,0000	-	-	-
22	25	-3,0000	3,0000	2,5	-	2,5
11	8	3,0000	3,0000	2,5	2,5	-
12	11	1,0000	1,0000	1	1	-
40	35	5,0000	5,0000	4	4	-
56	49	7,0000	7,0000	5	5	-

suma rang różnic dodatnich T+ 12,5000 suma rang różnic ujemnych T- 2,5000

statystyka testowa T = 2,5000 obszar krytyczny (alfa=0.05) (-inf; 0]

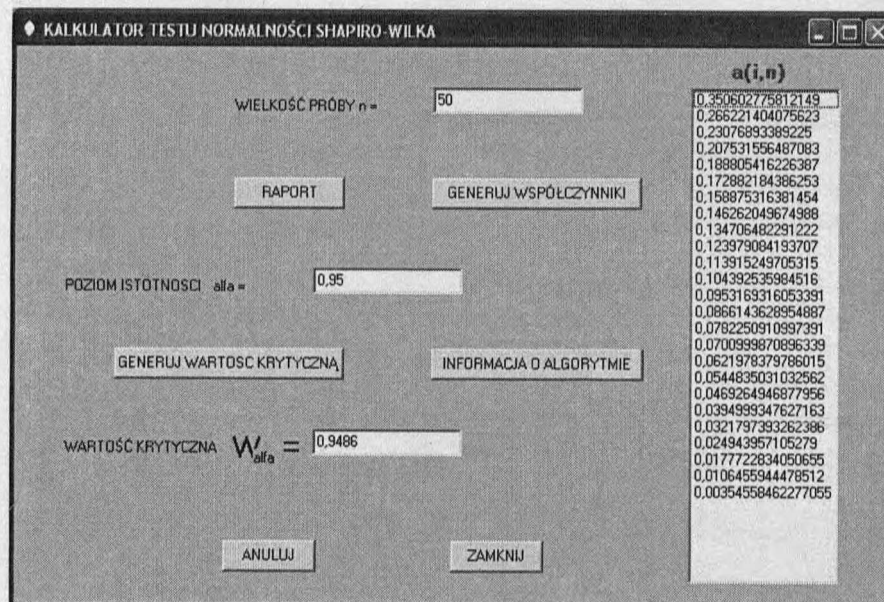
WERYFIKUJ Nie ma podstaw do odrzutu hipotezy Ho przy danym poziomie istotności alfa.

Rys 5. Okno dialogowe sekcji *Test par Wilcoxona* w programie STAT_STUD

Źródło: Opracowanie własne.

Co ważne dla studentów – program krok po kroku przeprowadzi weryfikację postawionych hipotez, rozłoży obliczenia na poszczególne etapy, określi obszar krytyczny dla testu oraz dokona (opcjonalnie) weryfikacji hipotez, stawiając wniosek dotyczący słuszności weryfikowanej hipotezy dla zadanego poziomu istotności.

Moduł **KALKULATOR** przeznaczony jest do zagadnień takich jak: *minimalna liczebność próby* (przy estymacji wskaźnika struktury), szacowanie *wartości krytycznych* rozkładów oraz pozwala wyznaczyć kwantyle dowolnego rzędu dla rozkładu normalnego o zadanych parametrach. To faktycznie potrafi większość dostępnych pakietów. Zaletą jednak tego modułu jest fakt, iż umożliwia on również wyznaczenie współczynników występujących w teście normalności Shapiro-Wilka, według dość nowoczesnych algorytmów pochodzących od Roystona (por. Royston 1982, s. 115–124; 1995, s. 547–551), a dokładniejszych od współczynników pochodzących z oryginalnej pracy Shapiro i Wilka (1965, s. 591–611). Wyniki szacunków współczynników dla $n = 50$ elementowej próby przedstawiono na rys. 6.



Rys 6. Okno dialogowe sekcji *Kalkulator testy normalności Shapiro-Wilka* w programie STAT_STUD

Źródło: Opracowanie własne.

Moduł **GENERATOR** stworzony został z myślą o tych studentach bądź pracownikach naukowych, którzy – np. dla celów symulacyjnych – potrzebują wygenerować określoną liczbę wartości dla zmiennych losowych pochodzących z zadanego rozkładu. Zaimplementowane generatory dotyczą rozkładów: *równomiernego, normalnego, lognormalnego, t-Studenta, chi-kwadrat, F-Snedecora i trójkątnego*. Program korzysta z generatora liniowego o okresie 2^{96} (dla rozkładu równomiernego) opisanego w pracy Wieczorkowski, Zieliński (1997),

generatora Fibonacciego (do losowania liczb całkowitych z przeliczalnego zbioru), generatora ROU dla rozkładu normalnego (Ratio of Uniforms Method – por. Kinderman, Monahan 1977, s. 257–260) oraz związków pomiędzy wymienionymi rozkładami. Wygenerowane wartości można zapisać w pliku o formacie *xls*, co umożliwia przeprowadzenie dalszych analiz w programie Excel (np. budowanie histogramu, przeprowadzenie testów zgodności, etc.).

4. ZAKOŃCZENIE

Jak wspomniano, program STAT_STUD powstał z myślą o studentach uczących się statystyki. Program pozwala śledzić skomplikowane niekiedy rachunki krok po kroku, czuwa nad poprawnością wprowadzanych danych i ich zgodnością z założeniami modelowymi, a także zawiera wsparcie teoretyczne dla wnioskowania statystycznego. Atutami programu są jego wysoka funkcjonalność (co pokazano w ramach opisu poszczególnych modułów), prosta obsługa i bogaty *help*. Program zawiera również dość unikatowe narzędzia, jak obszerny generator rozkładów, generator współczynników dla testu Shapiro-Wilka, sekcje *Indeksy agregatowe*, czy *Tablica korelacyjna*. Nowością jest również możliwość wprowadzania danych w postaci szeregów rozdzielczych, zarówno punktowych jak i przedziałowych. Autor wyraża głęboką nadzieję, iż program STAT_STUD znajdzie uznanie zarówno wśród studentów zgłębiających tajniki statystyki, jak i dydaktyków, którzy ten przedmiot prowadzą.

LITERATURA

- Domański Cz. (red, 2001), *Metody statystyczne. Teoria i zadania*, WUŁ, Łódź.
- Kinderman J., Monahan J. (1977), *Computer Generation of Random Variables Using the Ratio of Uniform Deviates*, ACM Trans. Math. Soft. 3.
- Kończak G. (2008), *Rola i znaczenie symulacji komputerowych we wspomaganiu procesu nauczania statystyki*, Folia Oeconomica 217, Acta Universitatis Lodziensis, Łódź.
- Kufel T. (2007), *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, PWN, Warszawa.
- Kufel T. (2008), *E-projekty w nauczaniu metod ilościowych z wykorzystaniem oprogramowania Gretl*, Folia Oeconomica 217, Acta Universitatis Lodziensis, WUŁ, Łódź.
- Royston P. (1982), *An Extension of Shapiro and Wilk's W Test for Normality to Large Samples*, Applied Statistics 31.
- Royston P. (1995), *Remark AS R94: A Remark on Algorithm AS 181: The W Test for Normality*, Applied Statistics 44.
- Shapiro S., Wilk M. (1965), *An Analysis of Variance Test for Normality (Complete Samples)*, Biometrika 52.
- Wieczorkowski R., Zieliński R. (1997), *Komputerowe generatory liczb losowych*, WNT, Warszawa.

*Jacek Bialek***THE ORIGINAL COMPUTER PROGRAMM STAT_STUD
AS A TOOL FOR SUPPORTING A TEACHING OF STATISTICS**

The power of teaching by using computer programs is well known. But some of commercial programs are very expensive and out of a range of many universities. In this paper we propose a original computer program STAT_STUD, which is almost free, very functional and easy to use. The program includes the following modules: descriptive and mathematical statistics, calculator and generator. In the paper we discuss the mentioned modules and present the merits of the program.