

*Grzegorz Kończak**

ROLA I ZNACZENIE SYMULACJI KOMPUTEROWYCH WE WSPOMAGANIU PROCESU NAUCZANIA STATYSTYKI

1. WPROWADZENIE

W ostatnich kilkunastu latach nastąpił dynamiczny rozwój możliwości wykorzystania komputerów i odpowiedniego oprogramowania w procesie dydaktycznym. Początkowo była to możliwość przygotowania wydruków komputerowych i różnorodnych materiałów pomocniczych do zajęć, następnie możliwości prowadzenia wykładów z pomocą przygotowanych prezentacji oraz wykorzystania pakietów biurowych i specjalistycznego oprogramowania na ćwiczeniach w pracowniach komputerowych. Obecnie w procesie nauczania przedmiotów ilościowych powszechnie wykorzystywane są takie pakiety statystyczne jak SPSS, Statistica oraz R a także elementy pakietu Office – arkusz kalkulacyjny Excel, edytor tekstu Word i program prezentacyjny Power Point.

Szczególne możliwości w zakresie wykorzystania komputerów w dydaktyce występują w nauczaniu statystyki. Poza możliwością szybkiego przeprowadzenia złożonych obliczeń, w nauczaniu statystyki, bardzo pomocna okazuje się możliwość przeprowadzania z pomocą komputera symulacji rzeczywistych doświadczeń jak np. wielokrotne rzuty monetą, kostką i dużo bardziej złożonych wielokrotnie powtarzanych eksperymentów. Kilkanaście lat temu, aby wykonać proste symulacje komputerowe konieczne było przygotowanie specjalnego programu. Obecnie nawet dość złożone projekty symulacyjne można przygotować bez pisania programów, a jedynie wystarczy posłużyć się wbudowanymi funkcjami arkusza kalkulacyjnego Microsoft Excel.

W artykule przedstawiono analizę możliwości wykorzystania symulacji komputerowych do wspomagania procesu nauczania przedmiotów statystyka, statystyka opisowa i statystyka matematyczna. Wskazano programy, które są wykorzystywane

* dr, Akademia Ekonomiczna w Katowicach.

podczas zajęć prowadzonych w Akademii Ekonomicznej w Katowicach do przygotowania i zaprezentowania symulacji komputerowych. Przedstawiono wybrane przykłady symulacji komputerowych, które są wykorzystywane na różnych etapach procesu dydaktycznego, do prezentacji i wyjaśnienia różnych zagadnień statystycznych.

2. SYMULACJA KOMPUTEROWA I JEJ ZNACZENIE W ROZPOZNAWANIU I ROZUMIENIU PRAWIDŁOWOŚCI ZJAWISK MASOWYCH

Wielokrotne wykonywanie identycznych doświadczeń często było drogą do poznawania praw rządzących naturą. U podstaw odkrywania reguł rachunku prawdopodobieństwa było zainteresowanie grami hazardowymi. W XVII wieku de Mere obserwując gry w kości i aktywnie uczestnicząc w tych grach na tyle wypracował intuicję, że grając w kości potrafił wskazywać zdarzenia nieco bardziej niż inne prawdopodobne (nie posługując się pojęciem prawdopodobieństwa), co pozwoliło mu wygrać fortunę (P.L. Bernstein, 1997). Tenże de Mere zwrócił się do B. Pascala z prośbą o pomoc w rozwiązaniu problemu podziału wygranej w niedokończonej grze w kości. Próby rozwiązania tego zagadnienia często są postrzegane jako pierwsze kroki w poznawaniu i rozwoju metod rachunku prawdopodobieństwa. Wielu wybitnych naukowców rzuciło tysiące razy monetą by zrozumieć, potwierdzić lub lepiej rozpoznać rządzące naturą prawidłowości. O. Lange i A. Banasiński (1968) przytaczają, że doświadczenia takie przeprowadzali przyrodnik G. Buffon (4040 rzutów monetą) oraz statystyk K. Pearson (24 tysiące rzutów monetą). Na różnych uczelniach przeprowadzano podobne doświadczenia zmierzające do oceny częstości otrzymywania ustalonej strony monety. W XVIII wieku francuski przyrodnik G. Buffon obliczył prawdopodobieństwo zdarzenia, że igła o długości l rzucona losowo na siatkę linii równoległych odległych o a przetnie linię. Na tej podstawie M.P.S. Laplace zaproponował doświadczalną metodę szacowania liczby π . Jednak zarówno przy rzutach monetą jak i w doświadczeniu z igłą dla oszacowania liczby π dla uzyskania zadowalających wyników należy wykonać bardzo dużą liczbę doświadczeń. Obecnie, wykorzystując możliwości techniczne, rzeczywiste doświadczenia można zastąpić komputerową symulacją. Jeszcze kilkanaście lat temu przeprowadzenie prostej symulacji komputerowej wymagało napisania specjalnego programu. Nadal jest to niezbędne przy przeprowadzaniu złożonych symulacji, ale jednocześnie wiele prostych symulacji o dużych walorach dydaktycznych można przygotować wykorzystując podstawowe funkcje arkusza kalkulacyjnego. Zapoznanie studentów z problematyką symulacji komputerowych jest rów-

nocześnie zachętą do opracowywania przez nich własnych projektów symulacyjnych.

Symulacja to numeryczna procedura pozwalająca na przeprowadzenie eksperymentu z wykorzystaniem komputera. Szczególnym rodzajem symulacji komputerowej jest symulacja Monte Carlo – metoda symulacyjna wykorzystująca serię doświadczeń statystycznych. I.N. Kowalenko i in. (1989) określają Monte Carlo jako metodę obliczeniową, która jest oparta na interpretacji probabilistycznej szukanych wielkości i wykorzystaniu realizacji doświadczeń losowych w celu oszacowania tych wielkości. T. Hodgson i M. Burke (2000) podkreślają duże znaczenie symulacji komputerowych w rozwiązywaniu złożonych problemów, gdy trudne lub wręcz niemożliwe jest otrzymanie dokładnego rozwiązania analitycznego. Zauważają jednocześnie, że studenci często lepiej rozumieją pewne zagadnienia jeśli zostały one również przedstawione symulacyjnie a nie tylko od strony teoretycznej. Ważną rolę zastosowania symulacji w nauczaniu statystyki podkreśla również J.D. Mills (2003). Zaznacza jednocześnie, że uzupełnienie wykładu teoretycznego przykładami symulacyjnymi może być bardzo pomocne dla studentów na różnych kierunkach studiów.

Z metodami symulacji nieodłącznie związane są liczby losowe. Pierwsze tablice liczb losowych opublikował w 1927 roku L.H.C. Tippett, a nieco później podobne tablice przedstawili R.A. Fisher i F Yates. C.R. Rao (1994) podkreśla, że wytwarzanie liczb losowych to obecnie wielomiliardowy przemysł światowy wykorzystujący poważne badania naukowe i najnowsze szybkie technologie komputerowe. Zwiększające się możliwości szerokiego zastosowania komputerów i użycia niezawodnych generatorów liczb losowych stworzyły nowe możliwości w zakresie symulacji komputerowych oraz doprowadziły do powstania nowoczesnych metod statystycznych jak np. metody analizy podrób bootstrap czy jackknife.

Wraz z dynamicznym rozwojem możliwości komputerów w zakresie przetwarzania dużych ilości informacji w ostatnich dziesięcioleciach następował ciągły wzrost znaczenia technik symulacji komputerowych w badaniach naukowych. Symulacyjnie możemy porównywać szybkości różnych procedur obliczeniowych, oceniać obciążenia i błędy średniokwadratowe złożonych estymatorów, konstruować przedziały ufności dla szacowanych parametrów populacji, rozpoznawać własności testów statystycznych. Metody symulacji są podstawą metod analizy podrób jak bootstrap oraz jackknife. W dalszej części jednak skoncentrujemy się na możliwościach wykorzystania symulacji w procesie dydaktycznym, a w szczególności w nauczaniu treści przedstawianych na przedmiotach statystyka, statystyka opisowa i statystyka matematyczna.

3. SYMULACJE KOMPUTEROWE - NARZĘDZIA DO WYKORZYSTANIA

Wiele prostych symulacji komputerowych można przeprowadzić bez wykorzystywania specjalistycznych programów. W najprostszym przypadku do przeprowadzenia symulacji komputerowych wystarczy wyłącznie dostęp do sieci internetowej oraz dowolna przeglądarka stron internetowych. Projekty do przeprowadzenia symulacji „on-line” zamieszczone na stronach internetowych zwykle mają charakter dydaktyczny. Wśród różnych pomocy dydaktycznych dostępnych na stronach internetowych Kończak G. (2007) wymienia również takie projekty, które mają charakter symulacji komputerowej.

Nieco większe możliwości w zakresie przeprowadzenia symulacji komputerowych uzyskujemy wykorzystując powszechnie dostępny arkusz kalkulacyjny Microsoft Excel (może to również być inny arkusz kalkulacyjny). Możliwość generowania, z wykorzystaniem wbudowanych funkcji, wartości z takich rozkładów jak jednostajny, normalny, chi-kwadrat i wiele innych pozwala nawet osobom nie znającym języków programowania szybko opracować nawet dość złożone projekty symulacyjne. Mogą to być zarówno symulacje o charakterze dydaktycznym, jak również pozwalające rozwiązać rzeczywiste problemy statystyczne. Przykład wykorzystania funkcji arkusza kalkulacyjnego Excel do losowania podrób z pobranej próbki, a następnie do estymacji wybranych parametrów i weryfikacji hipotez przedstawia D. Christie (2004). Autor ten zwraca jednocześnie uwagę na nie najlepsze własności generatora liczb losowych wbudowanego w arkusz kalkulacyjny Excel. Fakt ten może prowadzić do zastrzeżeń co do stosowania tego generatora w badaniach naukowych, ale jednocześnie nie dyskryminuje tego rozwiązania w zakresie dydaktycznym. J.R. Evans (2000) przedstawia możliwości wykorzystania w dydaktyce projektów symulacyjnych w arkuszu kalkulacyjnym oraz omawia dodatek do arkusza kalkulacyjnego Excel – Crystal Ball (<http://www.crystalball.com>). Z wykorzystaniem tej aplikacji, również osoby niezbyt biegle posługujące się funkcjami arkusza kalkulacyjnego, bardzo szybko mogą opracować nawet złożone projekty symulacyjne. W obszernych arkuszach wynikowych przygotowywanych przez Crystal Ball można otrzymać symulacyjne oceny różnorodnych charakterystyk jak np. wartość oczekiwana czy przedział ufności dla nieznanego szacowanego parametru populacji.

Znacznie większe możliwości opracowania projektów symulacyjnych daje wbudowany w arkusz kalkulacyjny Excel język VBA. Szczególnie możliwość wykorzystania pętli, instrukcji warunkowych oraz zdefiniowania własnych funkcji i procedur obliczeniowych stwarzają z tego środowiska dobre i powszechnie dostępne narzędzie do opracowywania symulacji komputerowych. Ze względu na dostęp do arkusza kalkulacyjnego na uczelniach te właśnie narzędzia (Excel i VBA) mogą być stosowane w projektach symulacyjnych o charakterze

dydaktycznym. Przykładowy program symulacyjny, wspomagający wprowadzenie zagadnienia regresji liniowej, napisany w języku VBA przedstawiają G.T. Jones i in. (2004). Student ma możliwość określenia parametrów liniowej funkcji regresji a następnie wygenerowania obserwacji zgodnie z przyjętym modelem. Na podstawie otrzymanych wartości szacowane są parametry modelu. Użytkownik ma możliwość porównania wartości parametrów modelu regresji liniowej oraz otrzymanych ich ocen.

Dostępny na licencji *Open Source* program R jest bardzo interesującym narzędziem do przeprowadzania symulacji komputerowych. Użytkownik ma możliwość wykorzystania generatorów wartości zmiennych losowych o różnych rozkładach, instrukcji warunkowych i pętli. W Internecie dostępne są również witryny pozwalające na obsługę programu on-line. W takim przypadku użytkownik nie musi na lokalnym komputerze instalować programu R lecz po wprowadzeniu komend do okna przeglądarki internetowej wyświetlane są wszystkie rezultaty. Możliwości wykorzystania programu R on-line zapewnia m.in. system EMILeA-stat (K. Cramer K. i in., 2004). Pakiet podstawowy może być rozszerzany o pakiety zwiększające możliwości programu. W szczególności istnieje możliwość korzystania z pakietów, które posiadają specjalne funkcje wspomagające proces przeprowadzania analiz symulacyjnych. Za szczególnie przydatne do zaznajomienia się z ideą symulacji i opracowania projektów symulacji komputerowych należy uznać pakiety *rv* oraz *TeachingDemos*. Pierwszy z nich pozwala na definiowanie wektorów losowych o różnych rozkładach. Na tych wektorach można wykonywać różne operacje np. wyznaczać symulacyjnie charakterystyki funkcji zmiennych losowych. Drugi z wymienionych pakietów zawiera wiele przykładów, które mogą być wykorzystane jako pomoc przy wyjaśnianiu takich pojęć jak np. centralne twierdzenie graniczne, moc testu, konstrukcja przedziałów ufności czy metody analizy podrób.

Możliwości przeprowadzenia symulacji komputerowych zapewniają również takie programy jak SPSS i Statistica, choć nie są to programy ukierunkowane na analizy symulacyjne. Oba wspomniane pakiety dają użytkownikowi możliwość korzystania z funkcji pozwalających generować wartości z wielu różnych rozkładów. Zarówno w SPSS jak i w programie Statistica istnieje możliwość programowania w języku VBA, a więc wykorzystania np. instrukcji warunkowych i pętli.

4. WYBRANE PRZYKŁADY PROJEKTÓW SYMULACYJNYCH WYKORZYSTYWANE W NAUCZANIU STATYSTYKI

Procedury generowania liczb losowych są wykorzystywane do rozwiązywania różnorodnych problemów. R. Wieczorkowski i R. Zieliński (1997) wyróżniają trzy kategorie zagadnień, gdzie wykorzystuje się liczby losowe:

- problemy związane z przeprowadzaniem badań reprezentacyjnych,
- zadania numeryczne rozwiązywane metodą Monte Carlo,
- badanie różnych zjawisk i procesów za pomocą ich komputerowej symulacji (modelowania).

Zazwyczaj mówiąc o symulacji mamy na myśli problem związany z przeprowadzeniem wielu doświadczeń, jednak w szczególnych przypadkach proces symulacji może być wykonany jeden raz lub kilkakrotnie. W celu uzyskania wystarczająco dobrych ocen parametrów zazwyczaj niezbędne jest jednak przeprowadzenie setek lub tysięcy doświadczeń.

Przykłady symulacji komputerowych, które mogą być pomocne w przedstawianiu i objaśnianiu problemów statystycznych mogą być wykorzystywane na różne sposoby:

- w formie prezentacji danego zagadnienia na wykładzie, jako uzupełnienie prezentowanych treści teoretycznych,
- na ćwiczeniach w pracowni komputerowej,
- poprzez udostępnienie plików z przygotowanymi projektami do przeprowadzenia symulacji, jako pomoc do indywidualnej pracy studenta.

Poza wymienionymi możliwościami wykorzystania przykładów symulacji komputerowych w nauczaniu statystyki można zachęcać studentów do samodzielnych poszukiwań rozwiązań różnych rzeczywistych problemów metodami symulacyjnymi. To rozwiązanie może być szczególnie przydatne w ramach specjalnego przedmiotu, którego tematyka poświęcona jest metodom symulacji komputerowych lub w ramach spotkań koła naukowego.

W tabeli 1. przedstawiono wybrane przykłady symulacji komputerowych, które są wykorzystywane na zajęciach ze statystyki w Katedrze Statystyki Akademii Ekonomicznej w Katowicach. Prezentowane na wykładach przykłady są jedynie ilustracją omawianego zagadnienia i uzupełnieniem przekazywanych wiadomości teoretycznych. Wprowadzają one jednocześnie studentów w problematykę symulacji komputerowej. Część z wymienionych w tabeli 1 przykładów jest zamieszczona na dydaktycznej stronie Katedry Statystyki Akademii Ekonomicznej w Katowicach pod adresem <http://stat.ae.katowice.pl>.

Tabela 1

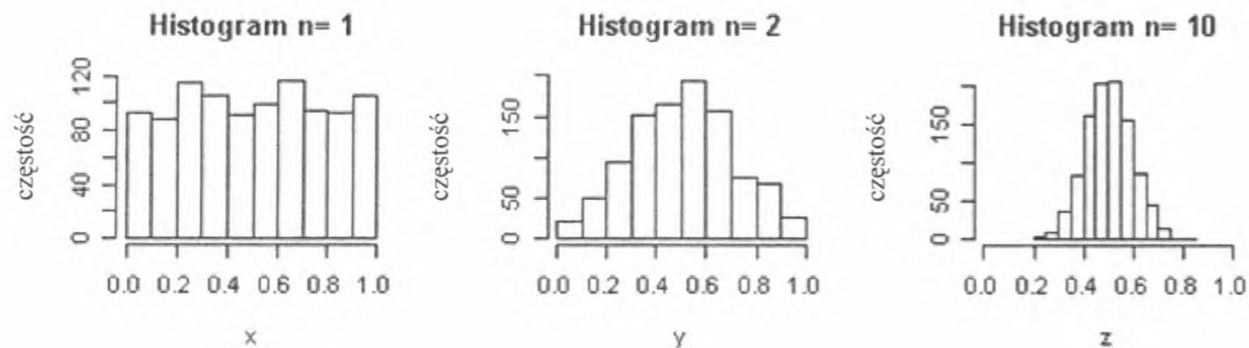
Wybrane przykłady zagadnień teoretycznych prezentowanych z wykorzystaniem przykładów symulacyjnych

Zagadnienie	Symulacja komputerowa
Rozkłady zmiennych losowych.	Generowanie wartości z zadanego rozkładu. Sporządzenie histogramu. Ocena parametrów rozkładu. Metody generowania wartości z zadanego rozkładu.
Rozkłady statystyk z próby	Konstrukcja histogramu obserwacji, średniej i wariancji z próby.
Wartość oczekiwana zmiennej losowej	Wartość oczekiwana liczby oczek w jednym rzucie kostką.
Prawo wielkich liczb	Częstość występowania „orzelka” w rzucie monetą. Szacowanie pola figury. Doświadczenie Buffona z igłą.
Metoda bootstrap	Ocena przedziałowa parametrów, weryfikacja hipotez statystycznych.
Wyznaczanie rozkładów złożonych statystyk, wyznaczanie kwantyli	Rozkład indeksów ekonomicznych z próby.
Analiza wariancji	Test analizy wariancji. Symulacja testu przy dla różnych danych wejściowych.
Symulacja rzeczywistych procesów	Symulacja notowań giełdowych. Symulacja procesów demograficznych.

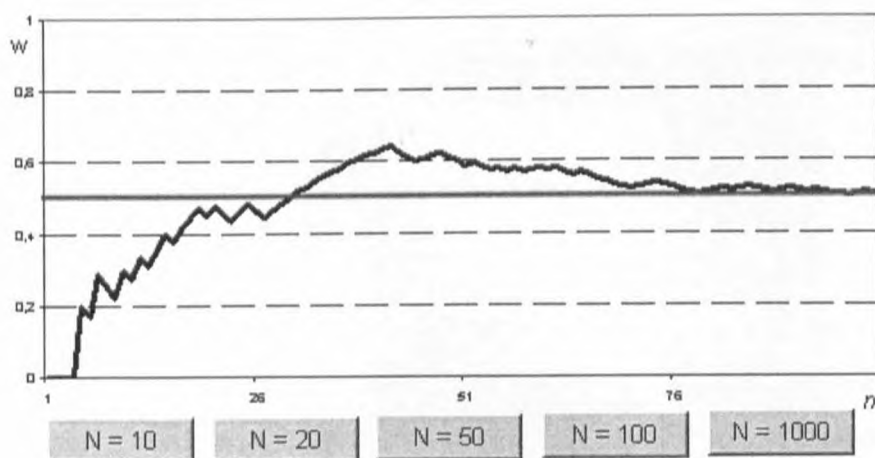
Na rys 1. przedstawiono opracowany w programie R projekt symulacyjny, który może być wykorzystany jako ilustracja centralnego twierdzenia granicznego. Użytkownik ma możliwość wybrania postaci rozkładu zmiennej losowej. Dla wybranej zmiennej losowej na podstawie przeprowadzonych symulacji wykreslane są histogramy dla średniej z próby dla prób o liczebnościach $n = 1, 2$ oraz 10 .

Na rys. 2 przedstawiono prezentację symulacyjną, która podobnie jak poprzedni przykład może być wykorzystana do zilustrowania centralnego twierdzenia granicznego. Dla różnych n ($n = 10, 20, 50, 100, 1000$ – liczba doświadczeń) użytkownik ma możliwość obserwacji częstości względnej występowania „orzelka” w rzucie monetą.

Na rys. 3 przedstawiono projekt symulacji pozwalającej na wyznaczenie rozkładu indeksu agregatowego cen według formuły Laspeyresa. W tym przykładzie do przeprowadzenia symulacji wykorzystano dodatek Crystal Ball firmy Oracle. Posługując się tym dodatkiem bez potrzeby pisania programu, a nawet bez konieczności wprowadzania formuł do komórek arkusza kalkulacyjnego można przygotować nawet złożone projekty symulacyjne. Dla danych wejściowych Crystal Ball pozwala na wybór spośród 22 dostępnych typów rozkładów zmiennych losowych m.in. zero-jedynkowy, równomierny, jednostajny, trójkątny, normalny i logarytmiczno-normalny. Dla wybranego rozkładu użytkownik ma możliwość podania wartości parametrów. Program wykonuje zadaną liczbę symulacji i w żądanej przez użytkownika formie przedstawia obszerne rezultaty przeprowadzonej symulacji.



Rys. 1. Symulacyjna ilustracja centralnego twierdzenia granicznego



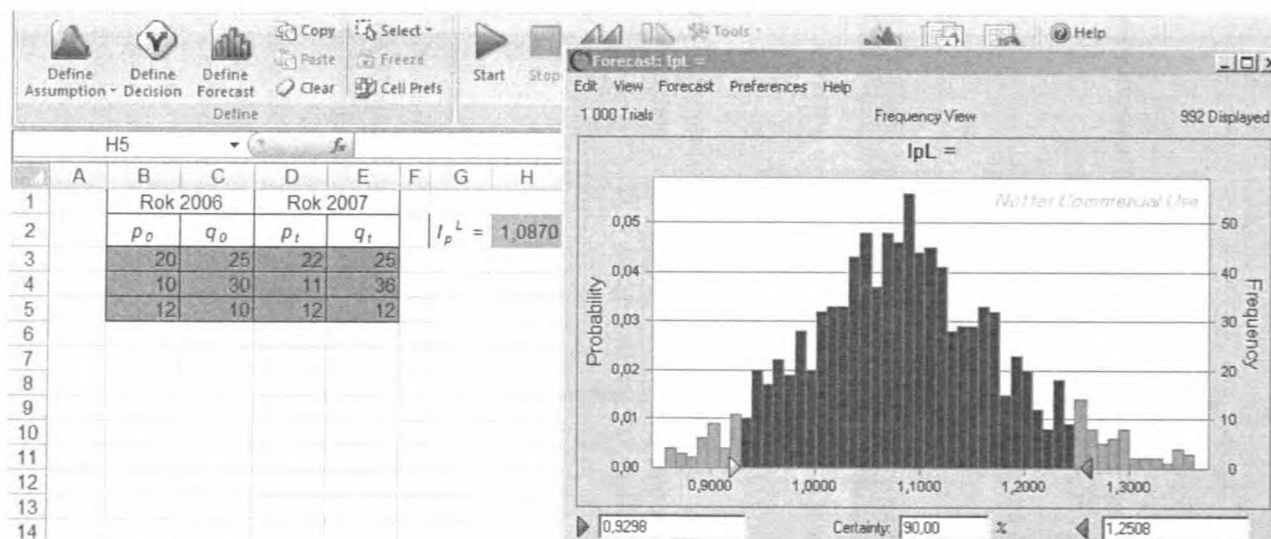
Rys. 2. Symulacja rzutu monetą – częstość względna wyrzucenych „orłów” w zależności od liczby doświadczeń

Przedstawione projekty symulacyjne są tylko wybranymi przykładami. Głównym celem przy opracowaniu tych przykładów jest ułatwienie studentom zrozumienia prezentowanych zagadnień teoretycznych.

5. ZAKOŃCZENIE

W ostatnich latach nastąpił znaczny wzrost możliwości wykorzystania komputerów nie tylko w badaniach naukowych ale również w procesie dydaktycznym. Do przedmiotów, gdzie możliwość wykorzystania komputera i odpowiedniego opracowania daje szczególne możliwości ubogacenia procesu dydaktycznego należy zaliczyć statystykę.

Proste przykłady symulacji komputerowych mogą być uzupełnieniem prezentowanych na wykładzie treści teoretycznych. Przygotowane pliki z określoną symulacją mogą być udostępniane studentom do umożliwienia im indywidualnego ćwiczenia i zarazem poznawania idei symulacji Monte Carlo. Jednak głębsze wprowadzenie w istotę zagadnień symulacyjnych jest możliwe jedynie w ramach przedmiotu specjalnie poświęconego tym zagadnieniom. Wobec rosnących możliwości wykorzystania analiz symulacyjnych nie tylko w statystyce, zdaniem autora, warto przedstawiać studentom przynajmniej w ramach przedmiotów do wyboru problematykę symulacji komputerowych.



Rys. 3. Symulacyjnie uzyskany rozkład indeksu agregatowego z wykorzystaniem Crystal Ball (wersja trial)

Wprowadzanie metod symulacyjnych w procesie dydaktycznym pozwala studentom lepiej zrozumieć omawiany materiał. Jeśli student pozna możliwości symulacji, to zapewne zachęci go to do poszukiwania rozwiązań złożonych problemów z wykorzystaniem metod symulacyjnych.

LITERATURA

- Bernstein P.L. (1997) *Przeciw bogom. Niezwykłe dzieje ryzyka*, WIG Press. Warszawa.
- Christie D. (2004) Resampling with Excel, *Teaching Statistics*, vol. 26, no. 1, s. 9–14.
- Cramer K., Kamps U., Zuckschwerdt Ch. (2004) St-apps and EMILeA-stat: Interactive visualizations in descriptive statistics, w: *Proceedings in Computational Statistics 2004*. Physica – Verlag Heidelberg – New York.
- Evans J.R. (2000) Spreadsheets as a Tool for Teaching Simulation, *INFORMS Transactions on Education*, vol. 1, no 1, s. 27–37.
- Hodgson T., Burke M. (2000), On Simulation and the Teaching of Statistics, *Teaching Statistics*, vol. 22, no. 3, s. 91–96.
- Jones G.T., Jones R.H., Jones K. (2004) A VBA-based Simulation for Teaching Simple Linear Regression, *Teaching Statistics*, vol. 26, no. 2, s. 36–41.
- Kończak G. (2007) Systemy Web Teaching – przegląd rozwiązań stosowanych w różnych krajach. *Acta Universitatis Lodzensis, Folia Oeconomica* vol. 205, s. 195–204.
- Kowalenko I.N., Kuzniecowa N.J., Szurienkow W.M. (1989) *Procesy stochastyczne. Poradnik*. Państwowe Wydawnictwo Naukowe. Warszawa.
- Lange O., Banasiński A. (1968) *Teoria statystyki*. Wydawnictwo Ekonomiczne. 1968.
- Mills J.D. (2003) A Theoretical Framework for Teaching Statistics, *Teaching Statistics*, vol. 25, no. 2, s. 56–58.
- Rao C.R. (1994) *Statystyka i prawda*, Wydawnictwa Naukowe PWN. Warszawa.
- Wieczorkowski R. Zieliński R. (1997) *Komputerowe generatory liczb losowych*. Wydawnictwa Naukowo-Techniczne. Warszawa.

Grzegorz Kończak

THE USE OF MONTE CARLO METHODS IN THE TEACHING OF STATISTICS

The power of computer simulation as a problem-solving is well known. The Monte Carlo study can be used in parameter estimation, hypothesis testing and resampling methods, such as bootstrap and jackknife.

The educational role of computer simulation is considered in the paper. The use of Monte Carlo study can promote a deep conceptual understanding of statistics. The programs, which can be used for preparation the simulation projects and the examples of educational computer simulation are described in the paper.