Joanna Krawiec

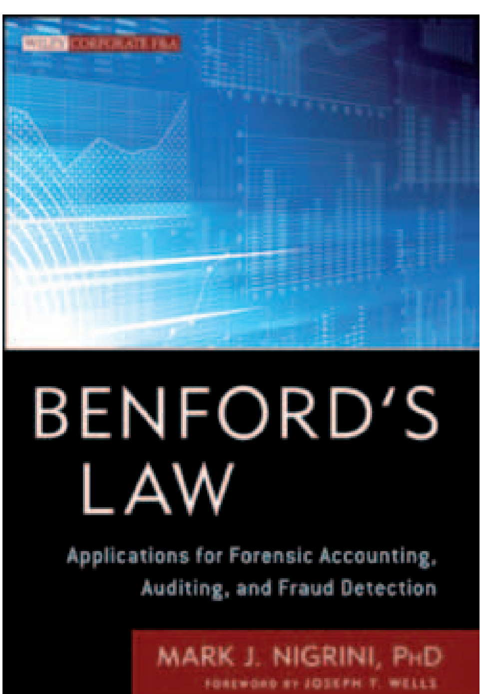
Prawo Benforda jako narzędzie wykrywania manipulacji finansowych

Mark Nigrini
Benford's Law, Applications for Forensic Accounting, Auditing and Fraud Detection
John Wiley & Sons, Hoboken, New Jersey 2012, pp. 330

Standardy rewizji finansowej stawiają przed biegłymi rewidentami wciąż nowe, zmieniające się wymagania.

International Standards on Auditing, opracowane i publikowane przez IFAC (International Federation of Accountants), rosnący nacisk kładą na procedury analityczne na zbiorach zagregowanych danych i efektywną ocenę ryzyka. Szczególnym rodzajem ryzyka, wyjątkowo trudnym w ocenie, jest ryzyko istotnego zniekształcenia sprawozdań finansowych przedsiębiorstwa wskutek oszustwa.

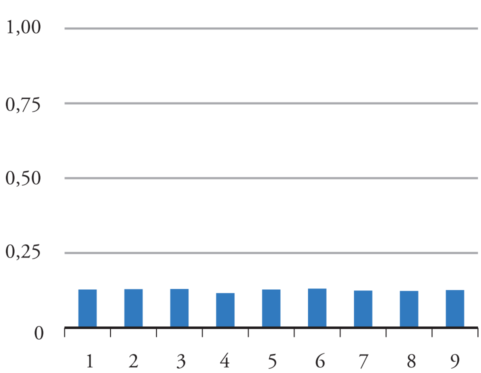
Większość procedur audytowych polega na rozmowach z kierownictwem, z księgowymi i na przetwarzaniu dostarczonych przez nich danych. Jeśli chodzi o ocenę poziomu ryzyka oszustwa, to zadawanie pytań wprost jest bezcelowe, a analizowanie dużych zbiorów danych „na chybił trafił" jest nieskuteczne. Mimo to, ISA 240[3] wymaga od biegłych rewidentów tego, by na każdym etapie badania sprawozdania finansowego brali pod uwagę możliwość popełnienia oszustwa lub błędu, przez osoby odpowiedzialne za zarządzanie przedsiębiorstwem, i aktywnie odnosili się do tej kwestii. Jednocześnie ISA 315[4] stanowi, że narzędziem służącym ocenie ryzyka, począwszy już od fazy planowania badania powinny być procedury analityczne, i szczególnie podkreśla możliwość ich zastosowania do wykrywania nadzwyczajnych transakcji, włączając w to oszustwa[4]. Dlatego rozwój narzędzi analitycznych, służących przetwarzaniu danych finansowych w celu wykrywania celowych, bezprawnych działań pracowników jest tak ważny. Mark Nigrini w swojej najnowszej książce prezentuje takie właśnie narzędzia.



Prawo Benforda – matematyka czy numerologia?

Metody analizy danych finansowych, którym Nigrini poświęcił około 20 lat badań[8], opierają się na pewnym matematycznym fenomenie. Zjawisko, odkryte przypadkiem przez Franka Benforda w 1938 roku[1], dotyczy rozkładu znaczących cyfr w zbiorach liczb losowych. Jakie jest prawdopodobieństwo, że pierwszą cyfrą liczby losowej jest 1? A jakie jest prawdopodobieństwo, że jest to 7? Odpowiedź zależy od tego, z jakimi zbiorami liczb losowych mamy do czynienia. Okazuje się, że liczby losowe liczbom losowym są nierówne. Zbiory liczb losowych są utworzone przez generatory liczb losowych mają inne właściwości, niż zbiory liczb losowych, które powstały w przyrodzie. Intuicja podpowiada, że prawdopodobieństwa występowania poszczególnych cyfr na pierwszym i kolejnych miejscach powinny być sobie równe t.j. 1/9 = 11,(1)%. Innymi słowy, jest tyle samo liczb, rozpoczynających się na każdą z dziewięciu cyfr[7)]. Taki intuicyjny rozkład mają znaczące cyfry w zbiorach pochodzących z generatorów liczb losowych. Rysunek 1 przedstawia rozkład pierwszych cyfr w zbiorze 10 000 liczb losowo wygenerowanych w Excelu.

Rysunek 1 **Rozkład pierwszych cyfr (generator liczb losowych)**



Źródło: opracowanie własne.

Wśród 10 000 losowych liczb jest po około 1111 liczb rozpoczynających się cyfrą 1, cyfrą 2, itd., aż do cyfry 9. Generator liczb losowych daje wyniki zgodne z intuicją większości ludzi. W przyrodzie jednak zwykle bywa inaczej.

Benford zupełnie przypadkiem zauważył, że strony tablic logarytmicznych są znacznie bardziej zabrudzone na początkowych, niż na końcowych stronach. Można by powiedzieć, że tak jest z większością książek, i nie ma w tym nic niezwykłego. Jednak tablic logarytmicznych nie czyta się dla przyjemności. Nie odkłada się ich na półkę z powodu mało interesującego wstępu. W tablicach poszukuje się wartości logarytmów liczb, a skoro pierwsze strony są znacznie bardziej zniszczone niż ostatnie, to znaczy, że ludzie znacznie częściej poszukują logarytmów liczb rozpoczynających się cyfrą 1 niż cyfrą 9. Dlaczego tak jest? Czyżby liczb zaczynających się jedynką było jednak w rzeczywistości więcej niż pozostałych?

Zaciekawiony odkrytym przez siebie zjawiskiem, Benford zbadał rozkład pierwszych cyfr w około 300 zbiorach liczb pochodzących z otaczającego nas świata[1)]. Dziś może się to wydawać niewielkim osiągnięciem, bo z pomocą arkusza kalkulacyjnego podobną pracę można by wykonać w ciągu jednego dnia, ale w 1938 roku było inaczej. Benford analizował m.in. takie dane, jak: wielkość zbiorników wodnych, długości rzek, populacje państw, ceny akcji, stałe fizyczne, itp. Wykazał, że rozkład pierwszych cyfr w tak różnych zbiorach danych jest bardzo podobny. By pokazać jak wygląda rozkład Benforda, zrobiłam to, co wielokrotnie zrobił sam Frank Benford i dziesiątki naukowców po nim. Dokonałam analizy zbioru danych, którymi w tym przypadku były średnie miesięczne zarobki w podziale na branże w 132 krajach świata w latach 1983-1999.

Przedstawiony na rysunku 2 rozkład, w którym prawdopodobieństwo, że pierwszą znaczącą cyfrą jest 1, wynosi ok. 30,1 proc. i spada logarytmicznie aż do 4,6 proc. dla cyfry 9, nazywamy rozkładem Benforda. Co ciekawe, im bardziej zróżnicowane są zbiory danych, tym większa jest ich zgodność z tym rozkładem. Gdybyśmy stworzyli jeden zbiór, ze wszystkich powstających w naturze zbiorów danych, to ten mega-zbiór miałby rozkład prawie idealnie zgodny z rozkładem Benforda. Rozkład ten odnajdujemy na każdym kroku w otaczającym nas świecie, podobnie jak np. liczby Fibo-

(1) $P\,(\textit{pierwsza znacząca cyfra} = d) = log_{10}(1+d^{-1})$

(2) $P\,(\textit{druga znacząca cyfra} = d) = \sum_{k=1}^{9} log_{10}(1+(10+d)^{-1})$

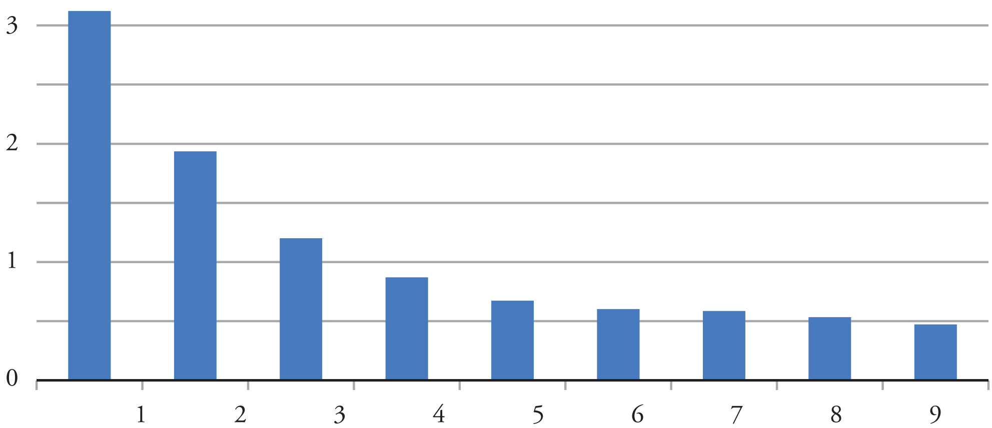
(3) $P\,(D_1=d_{1,\dots,}D_k=d_k) = log_{10}[1+(\sum_{i=1}^{k} d_i \times 10^{k-i})^{-1}]$

nacciego[10]. Te ostatnie mają oczywiście rozkład… Benforda, tak samo jak liczby Lucasa[5], wiele ciągów geometrycznych i innych. Odnajdowanie dziwacznego rozkładu cyfr w pochodzących z natury zbiorach danych nieodparcie kojarzy się z numerologią i brzmi niepokojąco. W rzeczywistości jednak, nie ma z numerologią nic wspólnego, a im dalej zagłębiamy się w matematyczne szczegóły, tym bardziej stają się one zastanawiające.

W 1995 roku Theodore Hill przeprowadził matematyczny dowód następującego twierdzenia: *Jeśli wybierzemy losowo pewną liczbę rozkładów zmiennej losowej, z każdego z tych rozkładów wybierzemy próbę liczb losowych i z tych liczb stworzymy jeden zbiór, to znaczące cyfry liczb tego zbioru mają rozkład Benforda*[2]. Im większa jest liczba różnych rozkładów, np. normalny, gamma, trójkątny, jednostajny, itd., oraz im większa jest liczebność próby, tym zgodność z rozkładem Benforda jest większa. Mówiąc innymi słowy rozkład Benforda jest ostatecznym rozkładem rozkładów. Przeprowadzony przez Hilla dowód pomógł zrozumieć, dlaczego tak często mamy do czynienia z fenomenem znaczących cyfr i uzasadnić jego zastosowanie w wielu dziedzinach nauki, włączając w to wykrywanie oszustw na podstawie danych księgowych. Hill podał wzór nie tylko dla rozkładu pierwszych (1), ale i drugich znaczących cyfr (2) oraz łącznego rozkładu cyfr (3).

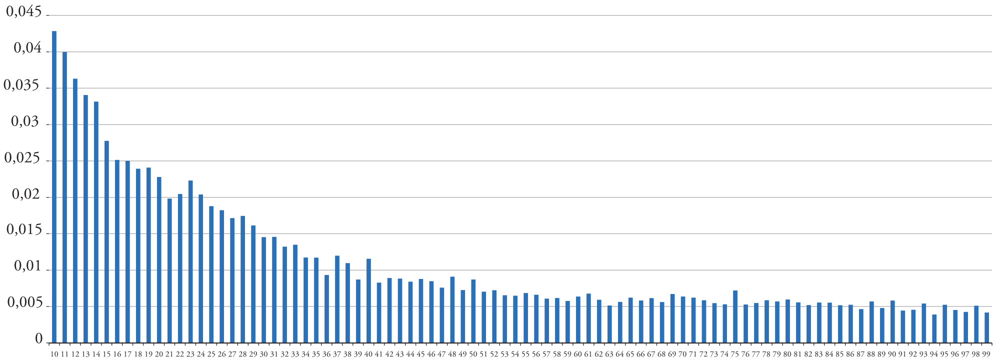
Przykładową ilustracją wzoru (1) jest rysunek 2. Drugie znaczące cyfry mają rozkład podobny jak pierwsze (2), z tym, że krzywa ma znacznie mniejsze nachylenie, czyli zróżnicowanie prawdopodobieństwa dla kolejnych cyfr jest znacznie mniejsze. Można podać wzory rozkładu dla trzeciej, czwartej i kolejnych, aż do dziewiątej cyfry. Zauważylibyśmy wtedy, że zróżnicowanie wartości prawdopodobieństwa bardzo szybko spada, i od czwartej cyfry w górę rozkład staje się praktycznie jednostajny. Wzór (3) jest szczególnie interesujący. Kryje się za nim zaskakujące twierdzenie – prawdopodobieństwa występowania kolejnych znaczących cyfr są od siebie zależne. Np. prawdopodobień-

Rysunek 2 **Średnie zarobki miesięczne na świecie w podziale na branże w latach 1983-1999, pierwsza znacząca cyfra**



Źródło: opracowanie własne na podstawie *Occupational Wages around the World,* Economics Web Institute, http://www.nber.org/oww/.

Rysunek 3 **Średnie zarobki miesięczne na świecie w podziale na branże w latach 1983-1999, pierwsza i druga znacząca cyfra**



Źródło: opracowanie własne na podstawie *Occupational Wages around the World,* Economics Web Institute, http://www.nber.org/oww/.

stwo, że cyfra 3 wystąpi na drugim miejscu jest większe, jeśli pierwszą cyfrą jest 1, niż gdy pierwszą cyfrą jest 2. Albo mówiąc inaczej, więcej jest liczb zaczynających się na 13 niż na 23. Rysunek 3 przedstawia rozkład dwóch pierwszych cyfr, w zbiorze danych na temat średniego wynagrodzenia w podziale na branże w 132 krajach świata w latach 1983-1999.

To wszystko jednak to dopiero wierzchołek góry lodowej. Hill, Nigrini i dziesiątki innych naukowców dowiodło szeregu twierdzeń dotyczących rozkładu Benforda. Najciekawsze z nich to m.in.:

- twierdzenie Nigriniego o tym, że sumy liczb rozpoczynających się określonymi cyframi są równe! [7)],
- prawo Pinkhama o niezależności względem skali, które mówi, że mnożenie przez stałą nie ma wpływu na rozkład Benforda[9)],
- twierdzenie Millera o niezmienności w bazie – zbiór liczb losowych, który ma rozkład Benforda w bazie 10, ma rozkład Benforda w każdej innej bazie[7)],
- twierdzenie Nigriniego i Millera o tym, że zbiór liczb stanowiących różnice pomiędzy liczbami zbioru Benforda, jest zbiorem Benforda,
- twierdzenie Nigriniego o tym, że zbiór liczb stanowiących różnice pomiędzy liczbami zbioru Benforda, który jest zbiorem Benforda, nie jest niezależny względem skali, a zatem nie spełnia prawa Pinkhama[7)].

Wszystkie te twierdzenia mogą być użyteczne w wykrywaniu oszustw księgowych. Na przykład oszustwo, które wiązało się z pomnożeniem wszystkich liczb przez pewną stałą, będzie niewykrywalne przy użyciu podstawowych testów z zakresu analizy znaczących cyfr, ale może być wykryte dzięki zastosowaniu ostatniego z powyższych twierdzeń.

Analiza Benforda – dla laików czy dla wtajemniczonych?

Nigrini w swojej książce daje liczne przykłady zastosowania każdego z wyżej wymienionych twierdzeń. W ciągu wielu lat badań naukowych i swojej praktyki zawodowej, początkowo w firmie audytorskiej, a później w FBI, na podstawie oryginalnych danych i analiz przeprowadzonych przez Benforda, rozwinął solidny aparat narzędziowy służący wykrywaniu różnego rodzaju oszustw – księgowych, podatkowych i ubezpieczeniowych. W książce kilka studiów przypadków zostało szczegółowo przeanalizowanych, a wyniki tej analizy opatrzono licznymi wykresami. Pokazano jak użyteczne może być każde z wyżej wymienionych twierdzeń, i jak oparte na nich narzędzia stosować poprawnie. Nigrini nie jest jednak bezkrytycznym entuzjastą analizy Benforda, i wskazuje jakie są ograniczenia w jej stosowaniu. Analiza znaczących cyfr jest tylko narzędziem, które Nigrini oddaje w ręce użytkowników. Żadne narzędzie, nawet najbardziej doskonałe, nie jest w pełni uniwersalne, a efekty jego zastosowania zależą w dużej mierze od tego, co i jak zrobi z nim użytkownik.

Po pierwsze, nie każdy zbiór liczb losowych, z samej swej natury, może być zbiorem Benforda[7)]. Prawa Benforda nie spełniają zbiory z przedziału liczb ograniczonego w taki sposób, że nie ma w nim co najmniej kilku rzędów wielkości. Nigrini podaje tu przykłady amerykańskich ulg podatkowych. By podać bliższy nam przykład weźmy zbiór wszystkich dopłat do zakupu mieszkań udzielonych w ramach programu „Rodzina na swoim". Ze względu na limity cen i powierzchni dopłaty kształtują się w przedziale 40-80 tysięcy złotych. Dlatego rozkład pierwszych cyfr będzie silnie zaburzony na korzyść cyfr 4, 5, 6, 7 i 8, natomiast cyfry 1, 2, 3 i 9 będą występowały bardzo rzadko. Podobny przypadek stanowią np. niektóre dane klimatyczne. Temperatura powietrza w Warszawie w ciągu roku waha się w przedziale ok. -30 do +36 stopni C, co powoduje, że na pierwszym znaczącym miejscu jest dużo cyfr 1, 2, oraz 3 i stosunkowo mało pozostałych cyfr.

Po drugie, nasuwa się pytanie o wielkość zbioru liczb, który może być poddany analizie znaczących cyfr. W zbyt małym zbiorze rozkład cyfr nie będzie zgodny z rozkładem Benforda. Rozkład Benforda jest rozkładem logarytmicznym, a prawdopodobieństwa występowania kolejnych cyfr mają charakter granic przy liczebności zbioru dążącej do nieskończoności. Zdaniem Nigriniego minimalna liczebność zbioru powinna wynosić 1000 dla celów analizy dwóch pierwszych cyfr, i co najmniej 300 dla analizy pierwszych cyfr[7)]. Z drugiej strony im większy jest zbiór, tym mniejsze odchylenie od wartości teoretycznej będą powodowały pojedyncze anomalie.

Na przykład, jeśli w pewnej spółce w ciągu roku dokonuje się 100 000 zapisów na kontach, z czego 10 to zapisy związane z pewnym oszustwem, analiza znaczących cyfr nie może skutecznie takiego oszustwa wykryć. Prawdopodobieństwo, że pierwszą cyfrą w zbiorze Benforda jest 1, wynosi log 2 = 30,102%. Zatem w podanym wyżej zbiorze 30 102 zapisy powinny się rozpoczynać cyfrą 1. Prawdopodobieństwo, że pierwszą cyfrą jest 2, wynosi log 3/2 = 17,609%, a zatem 17 609 zapisów będzie miało pierwszą cyfrę 2. Prawdopodobieństwa maleją logarytmicznie, aż do log 10/9 = 4,575% dla cyfry 9. Załóżmy, że wskutek oszustwa zmieniono 10 zapisów (5 po stronie winien i 5 po stronie ma) z wartości 1000 zł do wartości 9999 zł – np. dlatego, że 10 000 zł jest progiem, powyżej którego transakcje pewnego pracownika wymagałyby dalszej autoryzacji. Wtedy zamiast 4576 zapisów rozpoczynających się cyfrą 9, będzie ich 4586, natomiast zapisów rozpoczynających się cyfrą 1 będzie o 10 za mało, czyli „tylko" 30 092. Najlepszą miarą zgodności z rozkładem Benforda jest MAD – średnie bezwzględne odchylenie. W powyższym przykładzie wskutek oszustwa zmieni się ono o 2*(10/100 000)/9 = 0,0000(2), co jest w granicach błędu statystycznego i nie może być podstawą do wnioskowania o braku zgodności z rozkładem.

Po trzecie, jeśli już dysponujemy zbiorem danych, który ze swej natury może być zgodny z rozkładem Benforda i liczebność zbioru jest wystarczająca, rodzi się pytanie, jakie dane powinniśmy właściwie rozpatrywać w konkretnym przypadku. W przypadku danych księgowych i zastosowania prawa Benforda do wykrywania oszustw, czy powinny być to wszystkie zapisy na kontach? A może tylko zapisy na kontach wynikowych? Albo tylko lista faktur zakupowych, czy też dane na temat wynagrodzeń pracowników? Jakie liczby należy rozpatrywać jako jeden zbiór, a jakie należałoby testować osobno? Na każde z tych i wielu innych pytań użytkownik zaprezentowanych w książce Nigriniego narzędzi musi sobie odpowiedzieć sam. Skuteczność analizy Benforda będzie w dużej mierze zależała od doboru danych, a to z kolei stawia pewne wymagania przed jej użytkownikami.

Analiza znaczących cyfr nie jest „samograjem", standardową procedurą taką samą w każdym przypadku. Wręcz przeciwnie, wymaga oceny ryzyka i przewidywania możliwości wystąpienia konkretnego rodzaju oszustwa po to, byśmy wiedzieli, czy analizować dane np. zakupowe, sprzedażowe, płacowe, podatkowe, czy jeszcze inne. Dodatkowo, inne manipulacje dotyczą liczb mniejszych, a inne większych od zera, np. zwykle przychody i aktywa są zawyżane, podczas gdy koszty i pasywa zaniżane. Tu też jednak nie wolno generalizować, bo np. koszty podatkowe mogą być zawyżane w celu uniknięcia płacenia podatków. Wiele tego rodzaju rozważań należy przeprowadzić osobno w każdym przypadku, a powyżej podano tylko nieliczne przykłady. Rozważania te wymagają nie tylko znajomości reguł dotyczących rozkładu cyfr, ale także wyobraźni i doświadczenia w badaniu danych finansowych. Nie znaczy to, że bez ogromnej wiedzy i doświadczenia analizy Benforda nie da się skutecznie stosować. Jak napisał we wprowadzeniu do książki Joseph T. Wells: *Nie musisz umieć zbudować zegarek, żeby wiedzieć, która jest godzina*. W dodatku, analiza cyfr jest na tyle wciągająca, lub jak napisał w swojej książce Nigrini, wręcz „uzależniająca", że każdy, kto zacznie ją stosować, szybko zgłębi dalsze jej tajniki, bo trudno się od tego powstrzymać.

Książka Marka Nigriniego wprawdzie nie jest pozbawiona skomplikowanych wzorów i matematyczny aspekt zagadnienia jest złożony, ale zawiera praktyczne opisy zastosowania matematycznych modeli zrozumiałe nawet dla laików. Można byłoby wręcz zarzucić, że sposoby prowadzenia testów i oceny zgodności wyników tych testów z rozkładem Benforda są wyjaśniane tak, jakby czytelnik nie miał pojęcia ani o podstawach statystyki, ani o programach takich jak Excel czy Access. Choć niektórym na pewno będzie to przeszkadzać, to generalnie zapewne przyczynia się do większej dostępności i popularności tej książki.

Bibliografia:

1. Benford F., *The law of anomalous numbers,*"Proceedings of the American Philosophical Society", 1938, March, 78 (4), pp. 551–572.
2. Hill T.P., *A Statistical Derivation of the Significant-Digit Law,* "Statistical Science", 1995, 10(4), pp. 354-363.
3. *ISA 240, The Auditor's Responsibilities Relating to Fraud in an Audit of Financial Statements*, IFAC.
4. *ISA 315 (Redrafted), Identifying and Assessing the Risks of Material Misstatement Through Understanding the Entity and Its Environment.*
5. Lucas E., *Theorie des nombres*, Gauthier-Villars, Paris 1891.
6. Newcomb S., *Note on the frequency of use of the different digits in natural numbers,* "American Journal of Mathematics", 1881, 4(1), pp. 39-40.
7. Nigrini M., *Benford Law. Applications for Forensic Accounting, Auditing and Fraud Detection*, John Wiley & Sons Inc., Hoboken, New Jersey 2012, pp. 23, 65-69, 74, 97-102, 113-115, 160.
8. Nigrini M., *The Detection of Income Tax Evasion Through an Analysis of Digital Frequencies*, rozprawa doktorska, University of Cincinnati, OH, USA, 1992, pp. 62-65.
9. Pinkham R.S., *On the Distribution of First Significant Digits*, "Annals of Mathematical Statistics", 1961, 32(4), pp. 1223-1230.
10. Sigler L.E., *Fibonacci's Liber Abaci.* Springer-Verlag, New York 2002.

Mgr **Joanna Krawiec**, doktorantka Kolegium Nauk o Przedsiębiorstwie, SGH.