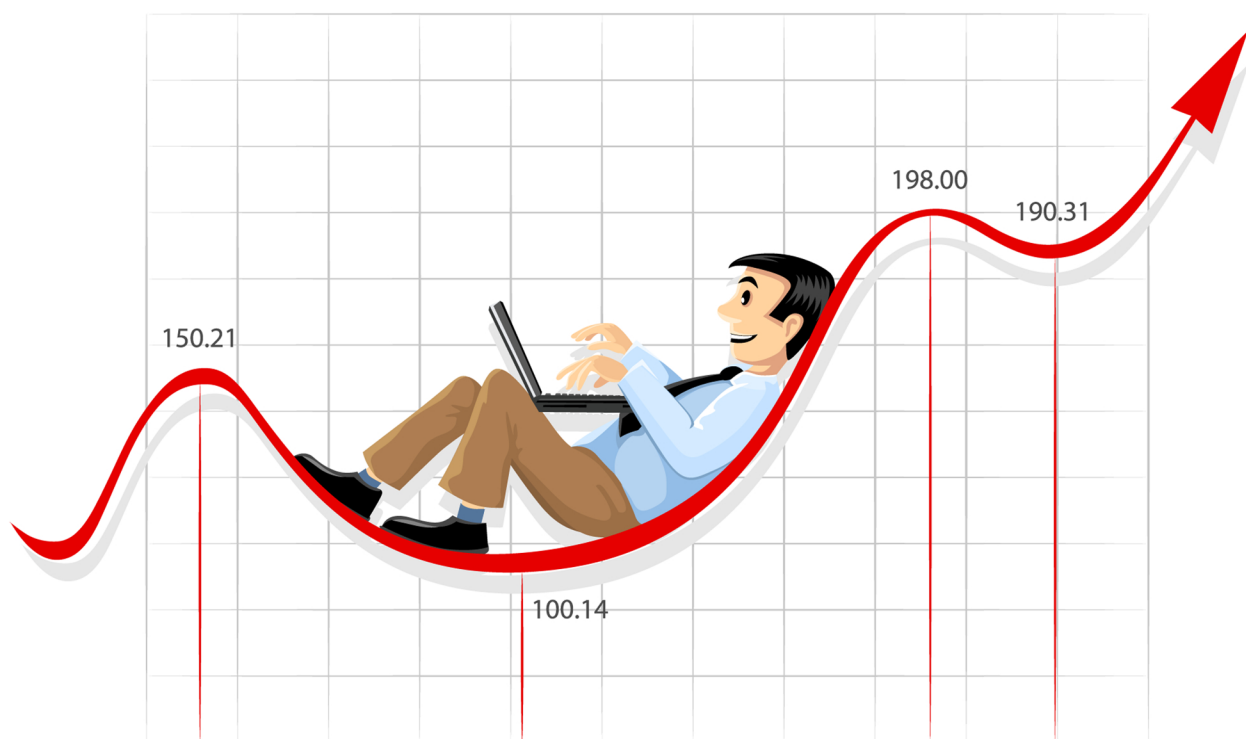




Co nowego w nauce?

Dane powiedzą ci całą prawdę



Siedlce 2013



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Projekt współfinansowany ze środków Unii Europejskiej
w ramach Europejskiego Funduszu Społecznego



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Co nowego w nauce?

Dane powiedzą ci całą prawdę

„Człowiek – najlepsza inwestycja”

Siedlce, 12 kwietnia 2013

Projekt współfinansowany ze środków Unii Europejskiej w ramach
Europejskiego Funduszu Społecznego

**Bezpłatne materiały szkoleniowe związane
z Projektem Uniwersytetu Przyrodniczo-Humanistycznego w Siedlcach
realizowanym w partnerstwie z Miastem Siedlce
"Praktyki pedagogiczne – kompetentnie, twórczo, przyjemnie"**

Program Operacyjny Kapitał Ludzki
Priorytet III. Wysoka jakość systemu oświaty,
Działanie 3.3 Poprawa jakości kształcenia
Poddziałanie 3.3.2. Efektywny system kształcenia i doskonalenia nauczycieli

Nr projektu: WND-POKL.03.03.02-00-041/10

Nr umowy dofinansowania: UDA-POKL.03.03.02-00-041/10-01

Okres realizacji projektu: 1.09.2010 r. – 20.10.2014 r.

Wartość projektu: 3 833 175,00 zł

Redakcja materiałów: dr Ryszard Kowalski, mgr Olga Szykarczyk

Biuro Projektu

Uniwersytet Przyrodniczo-Humanistyczny w Siedlcach, Instytut Biologii

ul. Bolesława Prusa 12, 08-110 Siedlce, p. 39

e-mail: praktyki@uph.edu.pl, www.praktyki.uph.edu.pl

tel./fax. 25 6431380



Konsekwentny, skromny, pogodny...

Wspomnienie o mgr Eugeniuszu Sabacie – współtwórcy uczelni wyższej w Siedlcach, nauczycielu akademickim i wychowawcy wielu roczników studentów matematyki.

(1921-2013)

Eugeniusz Sabat urodził się w 1921 roku w Zagórzcu (dzisiejsza dzielnica Kłobucka). Matematykę studiował na Uniwersytecie Łódzkim, gdzie później pracował jako adiunkt. Po założeniu rodziny od roku 1957 zamieszkał w Siedlcach. Był inspektorem szkolnym w Kuratorium Oświaty. Matematyki uczył w Liceum Ogólnokształcącym nr 2 im. Królowej Jadwigi i w Liceum Pedagogicznym. Jednocześnie pracował w Studium Nauczycielskim w Ciechanowie.

Oboje z żoną Wandą, nauczycielką historii w Liceum Pedagogicznym, byli oddani sprawom oświaty i nauczania. Dobrze znali dynamicznie rozwijające się w tamtym okresie środowisko szkolne Siedlec i powiatu siedleckiego (mówiło się wówczas, że „*Siedlce to miasto szkół*”). Żywo też interesowali się polityką oświatową prowadzoną przez ówczesne władze państwowe.

Wiadomość, że Ministerstwo Oświaty i Szkolnictwa Wyższego zamierza likwidować studia nauczycielskie (tzw. SN-y) i w ich miejsce powoływać trzyletnie samodzielne wyższe szkoły nauczycielskie, która dotarła do Siedlec w 1968 r., stała się dla Niego wyzwaniem. Postanowił podjąć starania, by właśnie w Siedlcach – stolicy Podlasia - w miejsce

istniejącego Studium Nauczycielskiego utworzyć pierwszą w tym rejonie szkołę wyższą. Z konsekwencją i determinacją prowadził rozmowy, pisał pisma, organizował narady, przekonywał przeciwników realizacji swojego pomysłu. Trzeba bowiem wiedzieć, że były też wysoko postawione osobistości przeciwnie utworzeniu w Siedlcach uczelni wyższej kształcącej nauczycieli. W staraniach wspierali Go wówczas dyrektor Liceum Pedagogicznego i Studium Nauczycielskiego Eugeniusz Gorecki, jego zastępca Mieczysław Dąbrowski oraz część władz partyjnych PZPR i administracyjnych miasta.

Była już wiosna 1969 roku, a Ministerstwo nie podjęło jeszcze decyzji w sprawie utworzenia uczelni w Siedlcach. Wszystkie dotychczasowe starania okazały się bezowocne. W dniu 6-go maja, przewodnicząc trzyosobowej delegacji, Pan Eugeniusz Sabat pojechał do Ministerstwa Oświaty i Szkolnictwa Wyższego. Po przeprowadzonych rozmowach otrzymał od dyrektora departamentu Wacława Wojtyńskiego wstępną decyzję na uruchomienie w Siedlcach wyższej uczelni. Mimo, iż czasu na organizację materialnego i kadrowego zaplecza szkoły wyższej było bardzo mało, panowie Sabat, Gorecki i Dąbrowski „zwerbowali” wymaganą przepisami kadrę naukowo-dydaktyczną, w tym pierwszego rektora uczelni – Józefa Kozłowskiego oraz docent Alinę Suszko-Purzycką. Jednocześnie zajmowali się tworzeniem warunków technicznych i administracyjnych przyszłej uczelni. Wszystko toczyło się od tej pory szalonym tempem. Już 1-go lipca 1969 roku przeprowadzono pierwszą rekrutację studentów na kierunkach: matematyka z fizyką, chemia z fizyką oraz nauczanie początkowe z wychowaniem fizycznym. Rok później utworzono kierunek biologia z chemią, a po dwóch latach – filologię polską. W krótkim czasie w uczelni zdołano utworzyć trzy wydziały: Matematyczno-Przyrodniczy, Pedagogiczny i Humanistyczny. Kadra naukowo dydaktyczna uczelni stopniowo powiększała się. Z nowym ośrodkiem akademickim zaczęli wiązać się ludzie, których działalność naukowa ceniona była w kraju, a nawet na świecie i którzy swoją osobowością i prężnością przyczynili się do jego rozkwitu. Do nich należy zaliczyć matematyka, śp. prof. Lesława Szczerbę. To z Jego inicjatywy Siedlecka Uczelnia od 1995 roku, jako jedyna wówczas

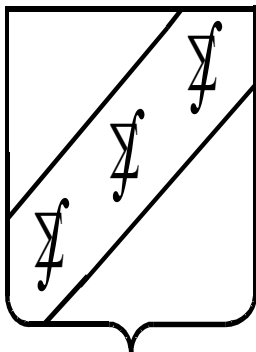
w Polsce, rozpoczęła kształcenie osób niepełnosprawnych. On również, wspólnie z matematykiem Markiem Kordosem, pracownikiem Uniwersytetu Warszawskiego, w 1988 roku utworzyli Ośrodek Kultury Matematycznej. Instytucja ta ma zasięg ogólnopolski. Zrzesza pracowników naukowo-dydaktycznych z wielu polskich ośrodków akademickich.

Pan Eugeniusz Sabat z zadowoleniem i dumą patrzył na rozwijającą się uczelnię. Z nieukrywaną radością, przyjmował nowych pracowników naukowych. Sam, zawsze skromny i opanowany, z poczuciem dobrego humoru, realizował swoje powołanie wykładowcy matematyki. Przez jakiś czas pełnił funkcję prodziekana ds. studiów zaocznych, a także kierownika Zakładu Matematyki. Z pasją pracował jako nauczyciel akademicki dbając o wysoki poziom nauczania i dając własnym przykładem wzór nauczycielskiego postępowania wielu rocznikom studentów matematyki – przyszłym nauczycielom tego przedmiotu. Wiedza, kultura osobista i takt w stosunku do podległych mu pracowników uczelni oraz studentów pozostaną na zawsze w pamięci tych, którym dane było Go poznać.

Ja miałam szczęście być Jego uczennicą w Liceum Pedagogicznym oraz asystentką w Zakładzie Matematyki, którym kierował.

Celina Kadej

OKM



Ośrodek Kultury Matematycznej (OKM) obchodzi w bieżącym roku XXV rocznicę swojego istnienia. Powołano go z inicjatywy pracowników uczelni w Siedlcach. Zrzesza pracowników naukowych, nie tylko matematyków, z różnych ośrodków akademickich i naukowych. Jest to jedyna na świecie tego typu instytucja. Swoją działalność skupia na organizowaniu Szkół Matematyki Poglądowej oraz wydawaniu zeszytów „Matematyka, Społeczeństwo, Nauczanie” -

stanowiących materialny efekt pracy OKM. W styczniu tego roku odbyła się w już 50-ta Szkoła Matematyki Poglądowej.

Pierwsze Szkoły Matematyki Poglądowej poświęcone były przede wszystkim popularyzacji różnych działów matematyki wyższej w środowisku matematyków, propagowaniu matematyki w środowiskach nie matematyków oraz podnoszeniu jakości kształcenia nauczycieli matematyki. Obecnie w tematyce Szkół wyraźnie dominują cztery opcje: matematyczna (propagowanie wśród pracowników wyższych uczelni tego co nowe w matematyce), ogólnokształcąca (przekazywanie wiedzy z historii i filozofii matematyki), społeczna (przekazywanie propozycji dydaktycznych, które można wykorzystać w pracy z uczniami na różnych poziomach nauczania matematyki oraz aplikacyjna (ukazywanie zastosowania matematyki w rozwiązywaniu problemów z różnych dziedzin życia)

Jednym z cenionych wykładowców Szkół organizowanych przez OKM jest Pan Doktor Andrzej Dąbrowski, pracownik naukowy Uniwersytetu Wrocławskiego. W ramach tegorocznego seminarium aktualizującego wiedzę „Co nowego w nauce?” wygłosi wykład i poprowadzi dyskusję na temat „*Dane powiedzą ci całą*



prawdę”. To wspaniały zbieg okoliczności, że wykład ten odbywa się w roku, ogłoszonym **Międzynarodowym Rokiem Statystyki**. W ten sposób, działania podejmowane w projekcie „Praktyki pedagogiczne – kompetentnie, twórczo, przyjemnie” wpisują się w tę międzynarodową inicjatywę, która zwraca uwagę na ogromne znaczenie matematyki w życiu każdego człowieka.

Celina Kadej, koordynator kierunkowy

Piśmiennictwo:

Matusak P. /red/: Od Wyższej Szkoły Nauczycielskiej do Akademii Podlaskiej. wyd. Wyższa Szkoła Rolniczo-Pedagogiczna w Siedlcach 1999.

OKM: Matematyka Społeczeństwo Nauczanie nr 50 wyd. UPH Siedlce 2013.



Dr Andrzej Dąbrowski
Uniwersytet Wrocławski

Studia matematyczne ze specjalnością zastosowania matematyki ukończył na Uniwersytecie Wrocławskim. Zastosowania matematyki były i są jego pasją. Pracę magisterską napisał na temat wartości informacji, a pracę doktorską na temat okresowych szeregów czasowych. Na temat teorii informacji napisał później książkę. Przez cały czas naukowej aktywności publikował prace ze statystyki i rachunku prawdopodobieństwa. Współpracował i współpracuje z różnymi instytucjami, jako konsultant statystyczny - od Szkoły Orłąt w Dęblinie, poprzez porady dla koncernu Forda do projektowania urządzeń diagnostycznych w medycynie. Ma w swoim dorobku także patent europejski dotyczący projektowania dróg.

Drugą jego pasją jest popularyzacja matematyki. Wykładał o matematyce od gimnazjum do Uniwersytetu Trzeciego Wieku. Pisał też artykuły popularne do różnych czasopism. Jest członkiem komitetu redakcyjnego czasopisma popularnego *Delta* i uczestnikiem prawie 50 Szkół Matematyki Poglądowej. Fotografia jednoznacznie świadczy o tym, że jako matematyk ma także przyrodniczą duszę.

Dane powiedzą ci całą prawdę...

Andrzej Dąbrowski

*„Dane są jak ludzie.
Wystarczy je przycisnąć,
a same powiedzą prawdę.”
Ronald Coase, ekonomista*

Dane

Na przełomie XVI i XVII wieku dokonał się w nauce znaczący przełom. Związany jest on głównie z nazwiskiem Galileusza (Galileo Galilei 1564-1642), który krytycznie analizował dzieła Arystotelesa, szczególnie dotyczące praw ruchu.

Arystoteles twierdził, że prędkość swobodnie spadającego ciała jest proporcjonalna do przebytej drogi. Galileusz znalazł w tym rozumowaniu wewnętrzne sprzeczności i ogłosił, że prędkość jest proporcjonalna do czasu. Co więcej, twierdził, że Arystoteles mylił się mówiąc, iż cięższe ciała spadają szybciej niż ciała lżejsze.

W uzasadnieniu swoich tez Galileusz po raz pierwszy w historii użył doświadczenia zamiast teoretycznego, opartego na logice uzasadnienia. W biografii, napisanej przez jego ucznia Vincenzo Vivianiego, podany jest opis eksperymentu, w którym Galileusz zrzucał kule o różnej masie z Krzywej Wieży w Pizie i notował czasy upadku. W taki sposób na scenie nauki pojawiły się dane eksperymentalne, które w zamierzeniu Galileusza miały poprzeć jego tezę. Galileusz został twórcą fizyki doświadczalnej, a

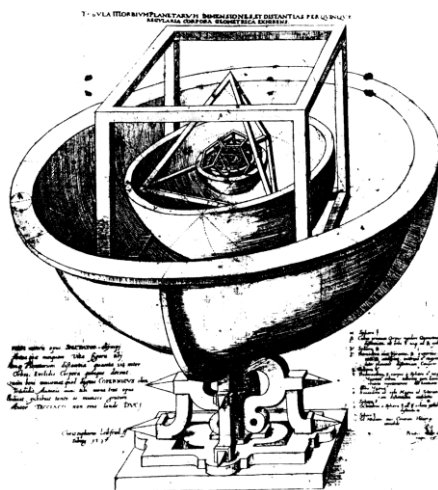
w nauce pojawił się silny konkurent metody dedukcyjnej – metoda indukcyjna, oparta na danych doświadczalnych.

Wraz z danymi pojawiły się błędy pomiarowe. Rzeczywiste czasy upadku kul z wieży były różne, a jednak Galileusz uznał, że powodem różnic były błędy pomiarowe, a nie istotnie różne prędkości spadania kul o różnej masie. To ważne dla metody indukcyjnej rozróżnienie czy mamy do czynienia z błędami, czy też z istotnymi różnicami, jest podstawowym problemem statystyki. W pełni naukowo satysfakcjonujące rozwiązanie nadeszło dopiero 300 lat później, w latach 30. XX wieku.

Wcześniej w nauce pojawiały się dane, nie pochodzące z eksperymentu tylko z obserwacji. Astronomowie od zawsze notowali położenia ciał niebieskich i na ich podstawie formułowali hipotezy związane z ruchem planet i gwiazd. W XV wieku mistrzem obserwacji był Tycho Brahe (1546-1601), który zbudował dwa obserwatoria na wyspie Ven, niedaleko Kopenhagi. Brahe, używając najlepszych na owe czasy instrumentów astronomicznych (nie dysponował teleskopem wynalezionym przez Galileusza) dokonał licznych pomiarów położenia Słońca, Księżyca i innych planet.

Asystentem Brahe, na rok przed śmiercią, został Johannes Kepler (1571-1640), którego zadaniem było wyciągnąć z tych danych ciekawe informacje o planetach. Używając współczesnego języka, Kepler musiał dokonać analizy tych danych lub, jak się czasem nazywa, „wydobyć wiedzę z danych”. Poświęcił na to pół życia, szukając jakichś prawidłowości. Wykonał tysiące obliczeń, rysunków, a nawet próbował różne modele geometryczne. Słynny jest rysunek pochodzący z pracy *Tajemnica kosmograficzna*, gdzie umieścił orbity planet

w koncentrycznych sferach, a na każdej z nich opisał kolejne bryły platońskie wpisujące się w sferę dla kolejnej planety. I to wszystko zgadzało się z danymi astronomicznymi!



Rys. 1 Sfery 6 planet opisane na i wpisane w 5 brył platońskich. Najbardziej zewnętrzna sfera, opisana na sześcianie, obejmuje orbitę Saturna. J. Kepler: *Mysterium Cosmographicum* 1596.

Jego praca została uwieńczona trzema, jak to dziś nazywamy, *Prawami Keplera*. Spośród nich najbardziej zadziwiające jest trzecie prawo, które powiada, że kwadrat obiegu planety wokół Słońca, liczony w latach ziemskich, jest równy trzeciej potędze średniej odległości od Słońca wyskalowanej tak, że odległość Ziemi od Słońca wynosi 1. Wynik jest o tyle zadziwiający, że został uzyskany metodą prób i błędów, a jego potwierdzenie przyszło wraz z mechaniką Newtona (1642-1727). Prace Keplera, przy użyciu współczesnych metod analizy danych, można wykonać w znacznie krótszym czasie. Autor tej pracy pokazuje uczniom

jak w prosty sposób można odkryć trzecie prawo Keplera w czasie nie przekraczającym 20 min (wraz z wyjaśnieniem metody obliczeń).

Tab. 1 Odległości planet od Słońca (w jednostkach astronomicznych) i rok planetarny (w latach ziemskich) oraz III Prawo Keplera. Planety wyróżnione nie były znane Keplerowi.

	Odległość (d)	Rok (t)	d^3/t^2
MERKURY	0,39	0,24	1,03
WENUS	0,73	0,62	1,01
ZIEMIA	1,00	1,00	1,00
MARS	1,53	1,88	1,01
JOWISZ	5,20	11,87	1,00
SATURN	9,54	29,48	1,00
URAN	18,98	84,07	0,97
NEPTUN	30,01	164,90	0,99
PLUTON	39,29	247,85	0,99

Pionierskie prace Keplera są prototypem współczesnego postępowania z danymi. O wiele łatwiej jest zebrać dane nawet w gigantycznych ilościach niż wydobyć drzemiącą w nich wiedzę. Danych dostarczają automaty pomiarowe, skanery, kamery cyfrowe, komputery. Zostawiamy o sobie informacje w szpitalach, bankach, firmach telefonicznych, sieciach społecznościowych (Facebook, Twitter,...). Ilość danych wytworzonych od zarania ludzkości produkuje się co dwa dni. Światowej sławy futurolog John Naisbitt powiada *Toniemy w informacjach, a łakniemy wiedzy.*

Opracowaniem efektywnych metod analizy danych zajmuje się statystyka. Jej osiągnięcia umożliwiły rozwój wielu dziedzin nauki, opartych o dane i doświadczenia. Niemożliwy byłby postęp w medycynie, biologii (szczególnie w genetyce), naukach społecznych, a nawet w archeologii, bez użycia metod statystycznych. Historia jak statystyka zrewolucjonizowała naukę XX wieku została opisana w bardzo poczytnej książce, wydanej w Stanach Zjednoczonych w 2001 roku, *The Lady Testing Tea. How Statistics Revolutionized Science In Twentieth Century*.

Co mówią dane?

*„Nagromadzenie danych
to nie jest jeszcze nauka.”*

Galileusz

Kepler uzyskał swoje wyniki nie posługując się żadną ściśle zdefiniowaną metodą. Na dłuższą metę nie dawało to perspektywy otrzymywania wartościowych wyników uzyskanych z obserwacji czy coraz częściej przeprowadzanych eksperymentów. Osiągnięcie sukcesu było rodzajem sztuki – raz przy sprzyjającym zbiegu okoliczności dawało się otrzymać ciekawy i sprawdzalny innymi metodami wynik, innym razem nie udawało się tego uzyskać. Nauka potrzebowała metody, która pozwoliłaby, tak jak metoda dedukcji w matematyce, wiarygodnie i jednoznacznie odpowiadać na pytania dotyczące rzeczywistości opisanej przez dane.

Pierwszym problemem do rozwiązania była eliminacja błędów pomiarowych. Przyczyną tych błędów były zapewne niedokładne

przyrządy lub pomyłki mierzących. Okazało się jednak, że jest znacznie więcej nieprzewidywalnych przyczyn błędów pomiarowych. Trzeba było zmierzyć się z tym problemem i powiedzieć, co to jest błąd pomiaru.

W roku 1632 Galileusz wydał dzieło *Dialog o dwóch najważniejszych systemach świata: ptolemeuszowym i kopernikańskim*. Rozwazał w nim metody ustalenia odległości nowo odkrytej gwiazdy od Ziemi na podstawie 78 pomiarów wykonanych w 13 różnych miejscach. W dziele tym podał charakterystykę błędów pomiarowych:

- błędy są nieuniknione i przypadkowe,
- małe błędy są bardziej częste niż duże,
- są symetryczne (tak samo często pomiar jest nieoszacowany, jak przeszacowany),

Krzywa przedstawiająca częstość występowania błędów w zależności od ich wartości powinna, według Galileusza:

- mieć maksimum w 0,
- być symetryczna w 0 i malejąca na prawo od 0 i na lewo od 0.

Mając na uwadze, że zmierzona wielkość jest sumą stałej, choć nieznaney, prawdziwej wartości pomiaru i błędu, można odczytać prawdziwą wartość pomiaru z krzywej, opisującej częstości występowania pomiarów. Byłaby to wielkość mieszcząca się w środku symetrii takiego wykresu, na dodatek w punkcie odpowiadającym jednemu maksimum.

W I połowie XVIII w. zaczęto szukać wzoru ustalającego położenie prawdziwej wartości pomiaru na podstawie wszystkich znanych pomiarów. Większość badaczy uważała, że dobrą oceną prawdziwego jest średnia arytmetyczna. Angielski matematyk Thomas Simpson (1710-1761) w pracy przedstawionej Royal Society w roku 1755

uzasadnił, że średnia pomiarów rzeczywiście koncentruje swoje wartości w pobliżu prawdziwego pomiaru.

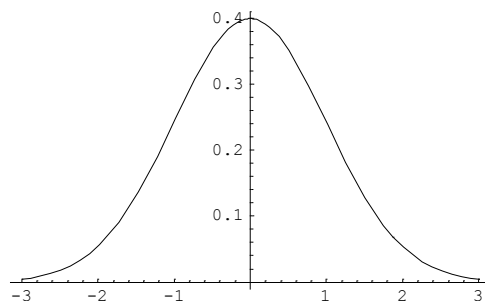
Jednak i ta średnia była skażona błędami pomiarowymi. Warto zainteresować się więc kształtem krzywej rozkładu średniej arytmetycznej. Krzywa jest skoncentrowana wokół prawdziwej wartości pomiaru i co jest bardzo użyteczne, tym bardziej skoncentrowana im więcej wykonano pomiarów. Stąd praktyczny wniosek, że duża liczba pomiarów zbliża średnią arytmetyczną do prawdziwej wartości pomiarów. Ten fakt nazwano *prawem wielkich liczb*.

Jeszcze jedno spostrzeżenie okazało się przełomowe, mianowicie to, że niezależnie od tego, jaka jest krzywa błędów dla pojedynczego pomiaru, krzywa błędów dla średniej arytmetycznej ma w przybliżeniu ten sam kształt. Kształt ten stabilizuje się, gdy średnia oblicza się dla dużej (zazwyczaj 50 i więcej) liczby pomiarów. Znacznie to upraszcza ocenę prawdziwej wartości pomiaru.

Fakt ten zaprezentował w dniu 9 kwietnia 1810 roku w Akademii Paryskiej Pierre Laplace (1749 - 1827) francuski matematyk, astronom, geodeta i fizyk przedstawiając krzywą błędów dla średnich arytmetycznych, wyrażoną wzorem,

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

która od swojego kształtu nazywa się krzywą dzwonową lub krzywą Gaussa.



We wzorze tym stała μ jest prawdziwą wartością pomiaru, a σ jest oszacowaniem błędu tej średniej, zwanego błędem standardowym.

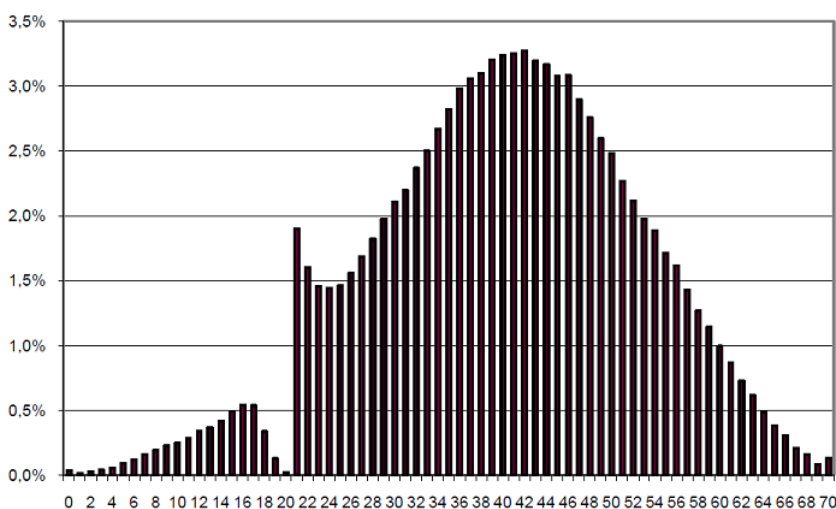
Okazało się, że mierząc obiekty biologiczne: rośliny, zwierzęta czy ludzi można zaobserwować podobne zjawisko. Na przykład, zestawiając pomiary masy ciała jakiegoś gatunku zwierząt w ustalonym wieku otrzymamy również krzywą Gaussa. Tym razem rolę błędów pomiarowych pełni inny czynnik - zmienność osobnicza. Stałą μ jest w tym przypadku masa ciała typowego przedstawiciela gatunku, a σ jest oszacowaniem tej zmienności osobniczej. Przeniosło to metodologię pomiarów na całkiem inne pole – biologii.

Zwolennikiem teorii, zresztą opartej na szeroko przeprowadzonych obserwacjach ze spisu powszechnego Belgii, przeprowadzonego w roku 1829, że pomiary antropometryczne podlegają rozkładowi Gaussa był matematyk, astronom i statystyk belgijski Adolphe Quételet (1796-1874), twórca bardzo popularnego wskaźnika szczupłości, znanego obecnie pod nazwą wskaźnika BMI. Pogląd Quételeta, że cechy opisujące różne zjawiska biologiczne mają rozkład Gaussa został powszechnie przyjęty wśród naukowców XIX wieku i często jest bazą rozumowań współczesnych.

W 1844 roku Quételet zadziwił wszystkich odkryciem, niestety po fakcie, oszustwa popełnionego przez rekrutów do armii francuskiej. Analizując rozkład wzrostu 10 tysięcy rekrutów zauważył, że liczba rekrutów o najniższym wzroście była większa o około 2000 od liczby, która powinna się pojawić zgodnie z krzywą Gaussa. Zakładając, że wzrost rekrutów podlega rozkładowi Gaussa było niezwykle mało prawdopodobne, aby mogła się pojawić tak duża liczba niskich rekrutów. Prawdopodobnie

zaniżyli oni swój wzrost poniżej poziomu wymaganego przez armię. Rozumowanie to przypominało znaną w matematyce metodę dowodu niewprost. Był to jeden z pierwszych przypadków zastosowania nowego typu rozumowania na podstawie danych nazwanego później testowaniem hipotez.

Gdyby żył Quételet, zwróciłby uwagę na wyskok ponad krzywą Gaussa w pobliżu 30% (próg punktów, od którego matura jest zdana) w rozkładzie ocen na maturze z polskiego, poziom podstawowy. Ten fakt każe przypuszczać, że około 1,5% populacji maturzystów „siłą woli” przekroczyło ten magiczny 30% próg zdawalności. Sytuacja ta powtarza się w wynikach matury z lat 2011 i 2012. Co ciekawe, takiego zjawiska nie ma w wynikach matury z matematyki.



Rys. 2 Wyniki matury z polskiego. Poziom podstawowy. Źródło: Fundacja Naukowa SmarterPoland.pl [www. Smarterpoland, wpis z 14 marca 2013 *Czy poloniści są mniej obiektywni a matematyki jest za mało w liceum?*]

W końcu XVIII wieku rozważano problem, czy komety należą do układu słonecznego. Pierre Laplace odpowiedział negatywnie na to pytanie w roku 1812. Laplace twierdził, że orbity komet przecinają się pod przypadkowym kątem do ekliptyki, co jest sprzeczne z ich przynależnością do Układu Słonecznego. Twierdzenie to oparł na obserwacji, że średni kąt przecięcia orbit komet z ekliptyką w przybliżeniu pokrywa się z oczekiwanym kątem przy założeniu przypadkowości orbit. Wcześniej, w roku 1796, w podobny sposób Laplace uzasadnił hipotezę mgławicową przypisywaną Emanuelowi Swedenborgowi i Immanuelowi Kantowi. Hipoteza ta głosi, że planety Układu Słonecznego powstały z płaskiego dysku utworzonego z materii, która odrywała się stopniowo od masy w centrum tworzącej Słońce. Laplace pokazał, że ekliptyki planet nie są losowe i koncentrują się w pobliżu jednej płaszczyzny, wyznaczającej położenie dysku, z którego powstały planety.

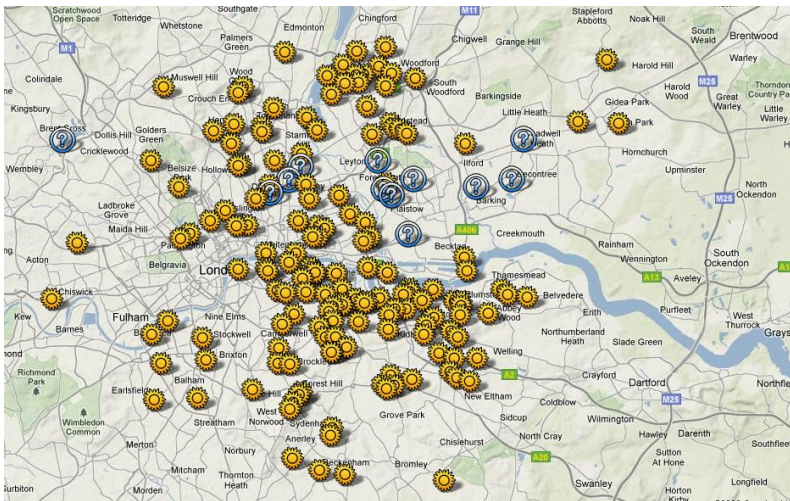
W obu przypadkach Laplace posłużył się metodą argumentacji podobną do rozumowania niewprost. Gdyby komety należały do Układu Słonecznego to kąt przecięcia ich orbit z ekliptyką winien być nieprzypadkowy. Skoro jest przypadkowy – to nie należą do Układu. Gdyby hipoteza mgławicowa nie była prawdziwa, to ekliptyki planet musiałyby być losowe – a nie są. Jest to argument za przyjęciem hipotezy mgławicowej.

Sposób postępowania, jaki zastosowali Quételet i Laplace, w którym dostateczne nagromadzenie faktów stanowi silny argument na rzecz rozważanej hipotezy, a niezwykle mało argumentów jest przeciw hipotezie, jest charakterystyczny dla rozumowania statystycznego. Krótko

mówiąc, statystyka pozwala ocenić czy można już zakrzyknąć: *To nie może być przypadek!* czy: *To czysty przypadek!* Czasami wypracowanie tej oceny wymaga pomysłowości i wyrafinowanych metod matematycznych.

Problem testowania, a więc rozstrzygania, która z postawionych hipotez jest w świetle zebranych danych prawdziwa, rozwiązywali na przełomie lat 20 i 30 XX wieku Polak Jerzy Sława Neyman (1894-1981) i Anglik Egon Pearson. Nauka zyskała precyzyjne narzędzie badawcze i wnioskowanie na podstawie danych zostało zaakceptowane w różnych dziedzinach aktywności.

Jesienią 1944 roku Londyn został zaatakowany przez nową, niezwykle groźną broń niemiecką. Rakiety V2 ważyły 13 ton i rozwijały czterokrotną prędkość dźwięku (4800 km/godz). Ulokowane głównie na wybrzeżu Holandii, średnio w odległości 250 km od Londynu, lecąc na wysokości 80 km docierały do stolicy Wielkiej Brytanii po 3 minutach od startu. Obrona przeciwlotnicza nie mogła więc podjąć skutecznej obrony. Żadna z rakiet nie została zniszczona a miejsca upadku rakiety były trudne do przewidzenia. Aby zorganizować skuteczną ewakuację należałoby przewidzieć, gdzie upadną rakiety. Czy jest to możliwe? Zapadła decyzja o weryfikacji, czy miejsca upadku rakiet tworzą nieprzypadkowe skupienia.



Rys. 3 Mapa upadku pocisków V2, zrekonstruowana na stronie internetowej Londynu http://londonist.com/2009/01/london_v2_rocket_sitesmapped.php

Sprawdzenie, czy obserwacje pojawiają się całkowicie losowo było jednym z pierwszych zagadnień statystycznych, rozwiązanych przez Karla Pearsona (1857-1936), ojca Egona. Skonstruował on test χ^2 , który pozwala ocenić stopień zgodności danych z postulowanym rozkładem prawdopodobieństwa. Za pomocą testu χ^2 Pearson, w roku 1900, rozstrzygnął problem czy kulka na kole ruletki w Monte Carlo zatrzymuje się przypadkowo.

Obszar Londynu o rozmiarach 12 na 12 km został podzielony na 576 kwadratów o boku 500 m. Policzono liczbę rakiet V2, które spadły na każdy kwadrat.

Tab. 2 Rozmieszczenie rakiet V2 w 576 kwadratach

liczba rakiet	0	1	2	3	4	7
liczba kwadratów	229	211	93	35	7	1

Łącznie spadło na ten obszar 537 rakiet. Średnio na kwadrat przypadło więc $537/576 \approx 0,93$ rakiety. Gdyby rakiety spadały przypadkowo, to ich rozmieszczenie w kwadratach podlegałoby rozkładowi Poissona z wartością oczekiwaną 0,93.

Tab. 3 Oczekiwane liczby rakiet w kwadratach przy założeniu losowego ich rozmieszczenia.

liczba rakiet	0	1	2	3	4	≥ 5
oczekiwana liczba kwadratów	226,7	211,4	98,5	30,6	7,1	1,6

Te dwie tabele są zadziwiająco zgodne. Można oszacować, że prawdopodobieństwo pojawienia się takiej tablicy, przy założeniu losowości, wynosi 0,88. Jest więc dostatecznie duże, aby przyjąć, że pociski spadały przypadkowo na Londyn.

Inną metodę testowania można zilustrować na przykładzie eksperymentu medycznego, w którym badano wpływ stresu na odporność przeciwko chorobie Heinego-Medina. W eksperymencie 23 małpy podzielono na dwie grupy: 11 małp (grupa kliniczna) poddano silnemu stresowi, 12 małp (grupa kontrolna) nie poddano żadnej presji (są to tzw. badania kliniczno-kontrolne). Wszystkie małpy zakażono wirusem choroby 24 godziny po zakończeniu sesji wywoływania stresu u małp w grupie klinicznej. Wszystkie małpy zachorowały, niektóre zmarły. Tabela wyników tego doświadczenia ma postać:

	Przeżyło	Zmarło	Razem
Grupa kliniczna	7	4	11
Grupa kontrolna	1	11	12
Razem	8	15	23

Czy stres zwiększa szansę przeżycia po zakażeniu wirusem Heinego-Mediny? W grupie klinicznej jest prawie dwukrotna przewaga małych, które przeżyły nad tymi, które zmarły. W grupie kontrolnej natomiast przewaga małych, które zmarły nad tymi, które przeżyły jest aż jedenastokrotna. Wydaje się, że wniosek jest jednoznaczny. Jednak należy przekonać sceptyków, że taki wynik, gdy stres nie ma wpływu na zdolność przeżycia, jest rzadkim, graniczącym z niemożliwością, wydarzeniem. Dokładnie mówiąc, taki wynik, a nawet gorszy (gdy w grupie kontrolnej umierają wszystkie zwierzęta) może teoretycznie się pojawić z niezwykle małym prawdopodobieństwem.

Prawdopodobieństwo to łatwo się oblicza. Na jednej stronie 11 kartek papieru umieścić można symbol grupy klinicznej, na 12 kartkach symbol grupy kontrolnej. Jeżeli stres nie ma wpływu na przeżycie to symbole przeżycia i zgonu nie są zależne od symboli grup. Kartki można więc potasować i na ich odwrocie napisać 8 razy symbol przeżycia i 15 razy symbol zgonu. Jaka jest szansa, że symbol przeżycia znajdzie się tylko na jednej kartce z grupy kontrolnej? Jaka jest szansa, że symbol przeżycia nie znajdzie się na żadnej kartce z grupy kontrolnej? To pierwsze prawdopodobieństwo wynosi:

$$\frac{\binom{12}{1}\binom{11}{7}}{\binom{23}{8}} \approx 0,0081$$

Drugie prawdopodobieństwo jest jeszcze mniejsze:

$$\frac{\binom{12}{0}\binom{11}{8}}{\binom{23}{8}} \approx 0,0003$$

Szansa otrzymania obserwowanego lub gorszego wyniku, gdy stres nie ma wpływu na przeżycie, wynosi zaledwie 0,0084. Taki wynik uznawany jest w nauce za dowód, że stres, występujący bezpośrednio przed zakażeniem wirusem Heinego-Medina, sprzyja przeżyciu. Do lekarzy teraz należy wyjaśnienie biologiczne tego fenomenu. Test uzyskany takim sposobem, zaproponowany przez wybitnego angielskiego statystyka Ronalda Fishera (1890-1962) nazywa się *dokładnym testem Fishera*.

Podobna była historia z odkryciem bozonu Higgosa. Od dwóch lat prowadzono intensywne eksperymenty w wielkim zderzaczach hadronów w instytucie CERN w Genewie. Co sekundę wypełniano komputery informacjami o objętości 200 tys. płyt DVD. Wreszcie, 4 lipca 2012 roku, dyrektor CERN ogłosił: *Zaobserwowaliśmy w naszych danych jasne sygnały o pojawieniu się nowej cząsteczki, na poziomie 5 sigma, w rejonie 126 GeV*. Sygnały te oznaczały wykroczenie pomiarów poza pas otaczający pewną krzywą, oznaczający brak bozonu. Dlaczego czekali tak długo, ponosząc nadzwyczajne koszty (przeszło 13 mld \$)?

W fizyce panują niezwykle wysokie kryteria losowości. Granicą, oznaczającą, że taki wynik jest przypadkowy przyjęto poziom 5 sigma, co oznacza, że gdyby bozon nie istniał to przeskoczenie tego pasa zdarzałoby się raz na 3,5 mln pomiarów. I właśnie 4 lipca sygnał był tak mocny, że nieistnienie bozonu stało się praktycznie (mówi się statystycznie) niemożliwe.

Analiza statystyczna danych znalazła swoje miejsce w nauce. Okazuje się, że na przełomie XX i XXI wieku stała się częścią bardzo dochodowego przemysłu. Przemysłu opartego na wiedzy. Powstała nowa gałąź aktywności, którego surowcem są dane. Nazwa tej aktywności *data mining* nawiązuje do wydobywania surowców (*mining* – ang. górnictwo). Celem jest wydobywanie użytecznych informacji z danych, tak jak z olbrzymiej ilości skał wydobywa się garstkę diamentów. Pracownicy tego przemysłu, analitycy danych, muszą doskonale znać statystykę, a znajomość informatyki, struktur baz danych jest niezbędnym narzędziem. W krajach, gdzie rozwijany jest przemysł przyszłości, oparty o wiedzę, analitycy danych są poszukiwaną elitą, znajdującą się w pierwszej piątce najbardziej poszukiwanych specjalności. Hal Varian, główny ekonomista Google'a powiada, że *...w następnej dekadzie to statystyk będzie hitem wśród zawodów* (dosłownie, *next sexy job*).

Wśród firm, oferujących produkt oparty o dane, wyróżnia się właśnie Google. Sposób oceny popularności stron internetowych PageRank, oparty o analizę danych dotyczących wejść na strony internetowe przynosi firmie znaczny dochód.

Nie wszyscy wiedzą, że Google skutecznie i z dużym wyprzedzeniem przewiduje epidemię grypy. W Stanach Zjednoczonych metoda Google'a,

praktycznie bezpłatna, jest konkurencyjna wobec kosztownego systemu *CDC's U.S. Influenza Sentinel Provider Surveillance Network*, pobierającego dane od 1500 lekarzy, którzy w czasie obserwowanych 16 mln rocznie wizyt pacjentów raportowali do centrali CDC przypadki chorych z objawami grypy.

Tymczasem Google wykorzystał dane, którymi i tak dysponuje – zapytania skierowane do wyszukiwarki Google'a. Opracowując model prognostyczny wyodrębnił słowa kluczowe, zawarte w pytaniach w okresie poprzedzającym bezpośrednio epidemię grypy. Wyodrębniono 48 słów kluczowych, najbardziej związanych z epidemią grypy: komplikacje (11), leczenie (8), symptomy (8), czas trwania (4), nietypowe objawy grypy(4), objawy komplikacji (4), antybiotyki (3), typowe lekarstwa na grypę (2), objawy chorób skojarzonych (2), lekarstwa antywirusowe (1), choroby bliskie (1). W statystycznym modelu logitowym powiązано częstość występowania tych słów kluczowych w zapytaniach Google'a z częstością pojawiania się pacjentów z objawami grypy. System świetnie się sprawdza w Stanach Zjednoczonych. Został też skalibrowany do danych w wielu krajach, nie wyłączając Polski (<http://www.google.org/flutrends/intl/pl/pl/#PL>).

Wpływ przemysłu opartego o dane często jest niezauważalny, a stykamy się z nim na co dzień. Kupując książkę przez Internet w sklepie, w którym jesteśmy zarejestrowani, otrzymujemy propozycje kolejnych zakupów, dopasowanych do naszych gustów lub otrzymujemy listę zakupów najczęściej dokonywanych przez osoby, które kupiły książkę właśnie przez nas zakupioną. W banku kredyt udzielany jest na podstawie specjalnej procedury skoringowej, opartej o analizę danych o kliencie.

Wielkie linie lotnicze, gdy lot jest odwołany, w pierwszej kolejności znajdują miejsce w samolocie dla klientów, których interesy są najbardziej zagrożone. W internetowej telewizji MSNBC, w dziale „polecane wiadomości”, analizowanych jest 16 ostatnio oglądanych przez widza wiadomości i polecane są te, najbardziej pasujące do profilu widza. Na północ od Little Rock, stolicy stanu Arkansas mieści się siedziba Axiom Corporation, globalnej firmy marketingowej. Dzień i noc pracuje w jej centrali 23 tysiące serwerów, gdzie analizowane są dane pół miliona klientów z całego świata, średnio 1500 danych o każdym z nich. Klienci są podzieleni na 70 grup o podobnych preferencjach. Oferta marketingowa jest posyłana do klientów według analizy ich potrzeb. A po dokonaniu zakupu, wysyłana jest propozycja zakupu sprzętu uzupełniającego, materiałów eksploatacyjnych itd.

Wizja świata zdominowanego przez dane właśnie się realizuje. Statystyka stała się niezbędnym narzędziem w opanowaniu tego świata. Nie dziwi więc, że rok 2013 ogłoszono Międzynarodowym Rokiem Statystyki.



W ramach obchodów tego Roku wygłaszane są wykłady przybliżające szerokiemu audytorium metody i wyniki uzyskane poprzez statystykę. Temu celowi służył też ten wykład.

Literatura

1. Berthold, M., R., Borgelt, C., Hoppner, F., Klawonn, F., *Guide to Intelligent Data Analysis. How to Intelligently Make Sense of Real Data*, Springer, London, 2010

2. Dąbrowski, A., *Rozkład normalny*, Matematyka w szkole, nr 31, 3-5, GWO, Gdańsk, 2008
3. Dąbrowski, A. *Statystyka użyteczna prawie wszędzie?*, Nr 45, 14-18, Matematyka, Społeczeństwo, Nauczanie Siedlce 2010
4. Diamond, S., *Wszechobecna statystyka*, Wiedza Powszechna, Warszawa 1970
5. Droesbeke, J.J, Tassi, P., *Historie de la statistique*, Press Universitaires, Paris, 1990
6. Kordos, M. *Wykłady z historii matematyki*, Script, Warszawa 2006
7. Salzburg, D., *The Lady Testing Tea. How Statistics Revolutionized Science In Twentieth Century. Henry Holt, New York , 2002*
8. Wang, C., *Sense and Nonsense of Statistical Inference Popular Statistics. Controversy, Misuse, and Subtlety*, Marcel Dekker, New York, 1992

NOTATKI ZE SZKOLENIA

Biuro Projektu
"Praktyki pedagogiczne – kompetentnie, twórczo, przyjemnie"
Instytut Biologii, Uniwersytet Przyrodniczo-Humanistyczny
w Siedlcach

ul. Bolesława Prusa 12 pok. 39

08-110 Siedlce

tel./fax: (25) 643 13 80

www.praktyki.uph.edu.pl

e-mail: praktyki@uph.edu.pl

„Człowiek – najlepsza inwestycja”

