

EWOLUCJA NAUK EKONOMICZNYCH II  
EKONOMIA A PANDEMIA COVID-19 POTRZEBA BIEŻĄCYCH DOSTOSOWAŃ CZY ZMIANY  
PARADYGMATU?  
22 listopada 2022 roku

**Krzysztof Jajuga (Uniwersytet Ekonomiczny we Wrocławiu)**

**Józef Pocięcha (Uniwersytet Ekonomiczny w Krakowie)**

**Mirosław Szreder (Uniwersytet Gdański)**

**Współczesne dylematy metod statystycznych i ekonometrycznych w zastosowaniach społeczno-ekonomicznych**

**Contemporary dilemmas of the use of statistical and econometric methods in socio-economic research**

**Streszczenie**

Opracowanie przedstawia współczesne tendencje i najważniejsze wyzwania, związane ze stosowaniem klasycznych i współczesnych metod ilościowych (przede wszystkim statystycznych) w badaniach społecznych i ekonomicznych.

W pierwszej części jest przedstawiony opis kluczowych przemian w ilościowych badaniach ekonomicznych i społecznych, między innymi związanych z postępem technologicznym. Następnie przeprowadzone jest syntetyczne porównanie statystycznych metod weryfikujących i metod eksploracyjnych.

W drugiej części przedstawiona jest dyskusja o konieczności znalezienia kompromisu pomiędzy łatwo dostępnymi zbiorami danych statystycznych i zaawansowanym oprogramowaniem statystycznym do ich analizy a formalnymi wymogami wnioskowania statystycznego. Jest obecnie krytyka współczesnego rozumienia i posługiwania się takimi kategoriami jak: istotność statystyczna czy p-value, co wynika z niedostatecznego respektowania najważniejszych założeń modelu wnioskowania statystycznego w praktyce badawczej.

W trzeciej części przedstawiono istotę i zasady uczenia statystycznego oraz panoramę statystycznych metod uczenia. W efekcie prowadzi to do sformułowania paradygmatu uczenia statystycznego, dla prowadzenia współczesnych badań statystycznych.

**Abstract**

The paper study presents contemporary trends and the most important challenges related to the use of classical and modern quantitative methods (primarily statistical) in social and economic research.

The first part presents a description of key changes in quantitative economic and social research, including those related to technological progress. A synthetic comparison of statistical verification methods and exploratory methods is then carried out.

The second part presents the discussion about the necessity to find a compromise between easily accessible data sets and advanced statistical software for analyzing them and formal requirements of statistical inference. It is current criticism of understanding and using such notions as statistical significance and p-value. This results from insufficient respect in practice for assumption of mathematical model of inference.

The third part presents the essence and principles of statistical learning and a panorama of statistical learning methods. As a result, this leads to the formulation of a paradigm of statistical learning for conducting modern statistical research.

**Słowa kluczowe:**

wnioskowanie statystyczne, testowanie hipotez, uczenie statystyczne, uczenie maszynowe, postęp technologiczny.

**Key words:**

statistical inference, hypothesis testing, statistical learning, machine learning, technological progress.

**Wstęp**

Opracowanie przedstawia współczesne tendencje i najważniejsze wyzwania, związane ze stosowaniem klasycznych i współczesnych metod ilościowych (przede wszystkim statystycznych) w badaniach społecznych i ekonomicznych. Rozważania podzielone są na trzy części:

- opis kluczowych przemian w ilościowych badaniach ekonomicznych i społecznych;
- główne wyzwania związane ze stosowaniem klasycznej teorii wnioskowania statystycznego;
- wyzwania związane z uczeniem statystycznym (maszynowym) w badaniach społeczno-ekonomicznych.

## 1. Przemiany w badaniach ekonomicznych i społecznych – niektóre współczesne tendencje

### 1.1. Badania w naukach ekonomicznych i społecznych w ostatnich latach

Analiza wyzwań, jakie pojawiają się w zastosowaniach metod ilościowych (matematycznych, statystycznych, ekonometrycznych) w badaniach zjawisk ekonomicznych i społecznych, musi mieć na uwadze przede wszystkim zmiany, jakie zachodzą współcześnie w tych badaniach.

Naszym zdaniem są dwie kluczowe tendencje, które kształtują i prawdopodobnie będą kształtować charakter badań w naukach ekonomicznych, jak również w części dyscyplin zaliczanych do nauk społecznych.

Pierwsza tendencja wynika ze stwierdzenia faktu, że główny nurt nauk ekonomicznych, zwłaszcza klasyczna ekonomia i klasyczne finanse, nie okazały się wystarczająco skuteczne w wyjaśnianiu i prognozowaniu procesów ekonomicznych i finansowych, zwłaszcza w sytuacji gwałtownych zmian w gospodarce, na rynku finansowym, a także w zachowaniu podmiotów (agentów) gospodarczych. Krytyka klasycznych teorii ekonomicznych i finansowych, opartych na założeniu racjonalności człowieka (*homo oeconomicus*), zaczęła być coraz bardziej intensywna po doświadczeniach kryzysu lat 2007-2008.

Naszym zdaniem (por. [Jajuga, 2019]), są co najmniej cztery cechy przemian współczesnych badań ekonomicznych, które mają i będą mieć kluczowy wpływ na kształt badań zjawisk ekonomicznych i społecznych, oczywiście również tych badań, w których zastosowanie mają metody ilościowe.

#### 1. Wzrost znaczenia pozytywnych (deskryptywnych) badań zjawisk ekonomicznych.

Z uwagi na to, że badania opierające się na normatywnych teoriach ekonomicznych mogą nie odzwierciedlać rzeczywistości (oparcie na kontrfaktycznych założeniach, brak odporności metod na dynamiczne zmiany w otoczeniu gospodarczym i społecznym, mniejsza waga przywiązana do rozwiązywania rzeczywistych problemów ekonomicznych), rośnie znaczenie badań deskryptywnych, w szczególności badań empirycznych, próbujących wyjaśnić działania podmiotów (agentów) gospodarczych.

#### 2. Stopień formalizacji nauk ekonomicznych.

Ta cecha ma szczególne znaczenie jeśli chodzi o znaczenie badań za pomocą metod ilościowych. Historia osiągnięć w naukach ekonomicznych, na przykład tych, za które

przyznana została Nagroda pamięci Nobla, wskazuje na dużą formalizację badań, przez stosowanie modeli matematycznych. Nie ulega najmniejszej wątpliwości, iż istotny stopień matematycznej formalizacji badań jest niezbędny, aby uniknąć braku pogłębionej analizy w niektórych badaniach, nazywanych usprawiedliwiająco jakościowymi. Jednak nadmierna formalizacja może prowadzić do zwiększonego ryzyka modelu, czyli nieprzydatności modelu w świecie rzeczywistym. Warto tu zacytować słynne stwierdzenie Roberta Mertona (por. [Merton, 1994]), odnoszące się do nauki o finansach (tłumaczenie: Krzysztof Jajuga):

„Czasem matematyka modeli staje się zbyt interesująca i zapominamy o głównym celu tych modeli. Matematyczne modele w finansach są precyzyjne, ale same modele są tylko przybliżeniem złożonego świata realnego. Dokładność tych modeli jako użytecznego przybliżenia świata rzeczywistego różni się znacznie w czasie i przestrzeni. Dlatego praktyk powinien stosować te modele niezobowiązująco, ostrożnie oceniając ich ograniczenia w każdym zastosowaniu”

Wynika z tego, że model będący efektem badań naukowych, powinien być odporny na zmiany w świecie rzeczywistym oraz przejrzysty dla użytkownika.

### 3. Interdyscyplinarność badań ekonomicznych.

Badania procesów ekonomicznych w coraz większym stopniu wykorzystują dorobek innych dyscyplin, co wynika z następujących przyczyn:

- pewna liczba dyscyplin naukowych pozwala na bardziej pogłębioną analizę zachowania agentów ekonomicznych, są to takie dyscypliny jak: psychologia, neurofizjologia (wpływ funkcjonowania mózgu na zachowanie człowieka), socjologia (wpływ interakcji w sieciach społecznych na zachowanie człowieka), antropologia (wpływ kultury na zachowanie człowieka);
- przemiany zachodzące w globalnej gospodarce stawiają wymagania metodyczne, wymuszające sięganie do innych obszarów wiedzy.

### 4. Metody weryfikacji teorii i metody eksploracji danych.

Metody ilościowe (statystyczne, ekonometryczne) można podzielić na dwie ogólne grupy, ten podział znany jest od dawna:

- metody weryfikujące (potwierdzające, konfirmujące, *confirmatory*), w których dokonuje się weryfikacji istniejących teorii z zastosowaniem dostępnych danych;

- metody eksploracyjne (*exploratory*), w których na podstawie danych próbuje się wykryć prawidłowości, mogące stanowić podstawę nowej teorii.

Pierwsza grupa metod jest rozwijana od bardzo wielu lat na gruncie klasycznej statystyki (a także ekonometrii). Historycznie główną „siłą napędową” stymulującą powstawanie kolejnych metod weryfikujących (z reguły są to znane metody wnioskowania statystycznego) było założenie normalności (i wielowymiarowej normalności) rozkładu statystycznego,

Metody eksploracyjne analizy danych są na większą skalę rozwijane od około sześćdziesięciu lat. W tym obszarze najważniejsze pionierskie prace są autorstwa Johna Tukeya, w szczególności książka „*Exploratory Data Analysis*” (por. [Tukey, 1977]), w której opisane szeroko są metody eksploracyjnej analizy danych. Ponadto pojawia się tam stwierdzenie autora, że zbyt duży nacisk był położony na wnioskowaniu statystycznym, a za mały na wykorzystaniu danych w celu identyfikacji hipotezy, która byłaby następnie testowana klasycznymi metodami wnioskowania statystycznego.

Wzrost znaczenia metod eksploracyjnych wynika z trzech przyczyn:

- nowe procesy ekonomiczne i społeczne nie są (jeszcze) odzwierciedlane w teorii , a zatem nie uzasadniają stosowania metod confirmacyjnych;
- następuje gigantyczny wzrost dostępnych danych, przede wszystkim Big Data, tzn. wielkich, zmiennych i różnorodnych zbiorów danych, których analiza może prowadzić do odkrywania nowej wiedzy;
- postęp technologiczny spowodował, że jest coraz łatwiej efektywnie (szybko i precyzyjnie) analizować dane.

Można tu stwierdzić (syntetycznie i nieco upraszczając), że metody weryfikujące to metody statystyki matematycznej, wywodzące się z podejścia stochastycznego, zaś metody eksploracyjne to metody szeroko rozumianej statystyki opisowej, nie mające u podstaw założeń o stochastycznym charakterze analizowanych zmiennych.

Druga tendencja, która kształtuje i będzie kształtować badania w naukach ekonomicznych wynika z przemian technologicznych. W tym zakresie należy wyróżnić trzy „siły napędowe”:

- wzrost szybkości komputera;
- wzrost liczby dostępnych danych;
- wzrost powiązań między podmiotami w sieciach społecznych.

Te trzy cechy spowodowały, że nastąpiła w pewnym sensie „konfrontacja” i „konkurencja” podejść w badaniach procesów ekonomicznych i społecznych, w których stosowane są metody ilościowe. Te podejścia to:

- stosowanie klasycznych metod statystycznych (lub ekonometrycznych);
- stosowanie metod uczenia maszynowego; upraszczająco, lecz nie zawsze słusznie nazywanych metodami sztucznej inteligencji.

W następnym punkcie pokazujemy, że rozróżnienie tych metod, jakoby jedno pochodziły ze statystyki, a drugie z informatyki, jest błędem.

### **1.2. Statystyka a uczenie maszynowe (sztuczna inteligencja)**

Pojęcie „sztuczna inteligencja” (SI) stało się niezwykle popularne w XXI wieku. Jest to modny termin (*buzzword*), słowo klucz, które charakteryzując metodę, ma legitymizować jej wartość dodaną i wysoką jakość. Niestety, ogromna większość osób używających tego słowa nie do końca rozumie istotę tego pojęcia.

Najbardziej popularne podejście wyróżnia 4 poziomy sztucznej inteligencji. Ta klasyfikacja opiera się na typach maszyn (komputerów), które mogą realizować określone aktywności wykonywane wcześniej przez człowieka, w pewnym sensie na kryterium „podobieństwa” komputera do człowieka. W poniższym zestawieniu mowa jest o maszynach (kojarzy się to z *hardware*), podczas gdy oczywiście chodzi o algorytmy (czyli *software*); maszyna sama nic nie „wymyśli”, o wszystkim decyduje człowiek – twórca algorytmu. Oto te 4 poziomy SI.

1. Poziom 1 – maszyny reaktywne. Ten poziom SI nie potrafi korzystać z informacji z przeszłości, tak aby podejmować optymalne decyzje, jedynie reaguje na otrzymywane bieżące sygnały. Na ich podstawie może konstruować możliwe scenariusze i wybierać optymalne decyzje. Algorytmy maszyn reaktywnych są skonstruowane tylko do rozwiązania konkretnych zadań, bez możliwości uogólnienia. Uważa się, że przykładem reaktywnej maszyny jest *Deep Blue*, komputer, który pokonał w latach 1996-1997 roku Garri Kasparowa.
2. Poziom 2 – maszyny z ograniczoną pamięcią. Ten poziom SI może wykorzystywać informacje z przeszłości do podejmowania decyzji dotyczących niedalekiej przyszłości, przy czym te informacje z przeszłości nie są zapamiętywane przez algorytm, tak aby mogły być wykorzystane przy kolejnych podobnych zadaniach.

Uważa się, że niektóre funkcje autonomicznych samochodów działają według tego rodzaju algorytmów.

3. Poziom 3 – maszyny oparte na teorii umysłu. Ten poziom SI mógłby analizować i rozumieć myśli i emocje człowieka, a zatem konstruować działania emulujące działanie człowieka, przez to mógłby nie być nigdy rozróżniany przez test Turinga. Prace nad algorytmami tego rodzaju trwają, choćby prace nad stworzeniem asystentów głosowych. Uważa się, że jednym z przykładów jest słynny robot *Sophia*, jednak nie ulega wątpliwości, że SI tego poziomu w pełnym wymiarze nie istnieje.
4. Poziom 4 – maszyny z pełną świadomością. Ten poziom SI przypomina w całości zachowanie człowieka (oczywiście jego świadomość, a także podświadomość). Ten poziom oczywiście nie istnieje, nie jest też jasne, po pierwsze, czy jest możliwe stworzenie tego rodzaju algorytmów, po drugie, na ile stanowią one będą zagrożenie dla ludzkości.

Spotykana często jest również druga klasyfikacja, w której wyróżniane są trzy poziomy sztucznej inteligencji, są nimi:

Sztuczna inteligencja – wąskie rozumienie (*Artificial Narrow Intelligence*), koncentracja na wąskim pojedynczym problemie; odpowiada to maszynom reaktywnym i maszynom z ograniczoną pamięcią;

Sztuczna inteligencja – ogólne rozumienie (*Artificial General Intelligence*), takie same możliwości jak człowiek, w zakresie rozwiązywania skomplikowanych zadań, jest to uzależnione od całkowitego poznania jak działa mózg;

Sztuczna superinteligencja (*Artificial Superintelligence*) – możliwości większe niż mózg człowieka (np. kreowanie sztuki, podejmowanie decyzji, relacje emocjonalne).

Osiągnięcia badań nad SI doprowadziły do poziomu sztucznej inteligencji w wąskim rozumieniu.

Algorytmy wykorzystywane w tych systemach od pewnego czasu nazywane są algorytmami uczenia maszynowego. Pojawia się tu naturalne pytanie o to czy algorytmy to są nową jakością (rewolucją) w analizie zjawisk ekonomicznych i społecznych, czy jest to naturalna ewolucja w zakresie znanych od dawna metod.

Warto zastanowić się nad dwoma pojęciami, które determinują rozwój algorytmów SI.

Pierwsze pojęcie to uczenie statystyczne (*statistical learning*). Szersze rozważania na ten temat są w dalszej części tego artykułu. Należy tylko nadmienić, że [Hastie, Tibshirani,

Friedman, 2009] opisują uczenie statystyczne jako wykorzystanie dostępnych danych (zwykle ilościowych lub ujętych w postaci kategorii) w celu określenia prognozy pewnych zmiennych.

Drugie pojęcie to uczenie maszynowe (*machine learning*). W podstawowym ujęciu (por. np. [Shalev-Shwartz, Ben-David, 2014]) jest to automatyczne wykrywanie istotnych wzorców w danych. W tym ujęciu podkreśla się, że uczenie maszynowe nie dąży do konstrukcji automatycznej imitacji procesów zachodzących w mózgu człowieka, ale przetwarza duże zbiory danych (nie tylko numerycznych, czy tekstowych, ale również na przykład plików wideo, plików audio, zdjęć satelitarnych itd.) za pomocą algorytmów matematycznych w celu wykrycia wzorców, których nie może na podstawie tak dużych zbiorów danych (i w tak krótkim czasie) wykryć człowiek.

Zagadnienia realizowane zarówno w ramach uczenia statystycznego, jak i uczenia maszynowego mogą być podzielone na cztery podstawowe klasy (przy czym część metod może realizować więcej niż jedno zadanie, a więc należeć do więcej niż jednej klasy). Oto ich lista:

**Metody dyskryminacji** (uczenie z nauczycielem, uczenie nadzorowane, *discrimination, classification*) polegają na podziale zbioru obserwacji (danych) na dwie lub więcej znanych klas, poprzez wyznaczenie jednej (lub więcej) funkcji dzielących (dyskryminujących). Na podstawie tych funkcji można następnie dokonać przydziału nowych obserwacji do jednej z wyznaczonych klas.

**Metody klasyfikacji** (uczenie bez nauczyciela, uczenie nienadzorowane, *classification, clustering*) polegają na podziale zbioru obserwacji (danych) na dwie lub więcej nieznanymi klas. W tym przypadku nie następuje przydział nowych obserwacji do wyznaczonych klas, lecz powtórny podział na klasy takiego powiększonego zbioru obserwacji.

**Metody redukcji wymiaru** (*dimensionality reduction*) polegają na przekształceniu wielu zmiennych na ich zdecydowanie mniejszą liczbę (to jest właśnie redukcja wymiaru), głównie w celu bardziej przejrzystej interpretacji. Jeśli ten zredukowany wymiar wynosi 2 (ewentualnie 3), wtedy istnieje możliwość wizualizacji danych w przestrzeni dwuwymiarowej (ewentualnie trójwymiarowej).

**Metody regresyjne** (*regression*) polegają na wyznaczeniu funkcji, która przedstawia zależność jednej zmiennej (tzw. objaśnianej) od zbioru innych zmiennych (tzw. objaśniających).



Publikacje do metod uczenia maszynowego zaliczają często właśnie metody realizujące powyżej wymienione zadania. Najczęściej wymieniane grupy metod to: metoda wektorów nośnych, metoda najbliższego sąsiada, sztuczne sieci neuronowe, metody analizy skupień (*cluster analysis*), analiza głównych składowych, drzewa klasyfikacyjne, drzewa regresyjne. Nietrudno zauważyć, że – poza sztucznymi sieciami neuronowymi – są to metody znane (niektóre od bardzo dawna) w statystyce. Z kolei rodowód sztucznych sieci neuronowych jest inny niż metod statystycznych, gdyż pierwszy model neuronu został zaproponowany na gruncie neurofizjologii (por. [McCulloch, Pitts, 1943]), dając początek rozwojowi sztucznych sieci neuronowych. Niemniej jednak, analiza struktury sztucznej sieci neuronowej wskazuje, że w ich przypadku mamy do czynienia ze skomplikowaną funkcją nieliniową przekształcającą wejście w wyjście. Wobec tego można je potraktować jako nieliniową funkcję regresji, rozumianej jako funkcja aproksymacyjna, tak jak regresję rozumie statystyka opisowa, niewprowadzająca podejścia stochastycznego.

Powyższe rozważania pozwalają na sformułowanie następujących konkluzji:

1. Metody analizy danych, które nazywane są też metodami uczenia statystycznego przez statystyków, zaś metodami uczenia maszynowego przez informatyków, są to zasadniczo te same metody.
2. Postęp technologiczny (szybkość komputerów oraz rozwój w zakresie Big Data) spowodował, że:
  - po pierwsze, zastosowanie znanych metod uczenia statystycznego stało się łatwe;
  - po drugie, nastąpił rozwój metod eksploracyjnych, które są coraz bardziej skomplikowane, niemniej jednak są to cały czas metody matematyczne, mające u podstaw klasyczne koncepcje statystyczne, takie jak np. szeroko rozumiana korelacja.
3. Stosowane dziś w naukach (badaniach) ekonomicznych i innych naukach społecznych metody ilościowe są konglomeratem dwóch nurtów badawczych, którymi są:
  - nurt rozwijany od kilkadziesiąt lat na gruncie statystyki i ekonometrii, zarówno w obszarze metod wnioskowania statystycznego, jak i metod eksploracyjnych (te drugie w podejściu opisowym);
  - nurt rozwijany na gruncie połączenia wysiłków naukowców z matematyki stosowanej i statystyki z naukowcami zajmującymi się technologiami

informacyjnymi, co zaowocowało rozwojem metodyki sztucznych sieci neuronowych (i podobnych metod).

4. Z punktu widzenia celów badań naukowych metody eksploracyjne mają dużą zaletę w postaci skuteczności prognostycznej a jednocześnie dużą wadę, którą jest znaczna ograniczoność (lub brak) walorów poznawczych, umożliwiających rozumienie zjawisk, a nie tylko praktyczną skuteczność – dotyczy to w szczególności sztucznych sieci neuronowych, działających na zasadzie „czarnej skrzynki”.

W kolejnym punkcie omówione są problemy odnoszące się do klasycznych metod wnioskowania statystycznego, które dotyczą zarówno „konkurencji” ze strony metod eksploracyjnych, jak i wyzwań w zakresie profesjonalnego stosowania tych metod w praktyce.

## **2. Praktyka wnioskowania statystycznego – współczesne wyzwania**

Postęp technologiczny, o którym wspomniano w poprzednim punkcie, służący upowszechnieniu zastosowań statystyki w wielu dziedzinach nauki, uwypuklił jednocześnie poważne zagrożenia związane z niewłaściwym wykorzystaniem wnioskowania statystycznego. Podstawowe źródła tych zagrożeń należy upatrywać w nadmiernym zaufaniu do statystycznych programów komputerowych, przy jednocześnie słabnącym zainteresowaniu ich użytkowników teorią statystyki, której programy te są podporządkowane. Statystyczne pakiety komputerowe są nie tylko coraz bardziej przyjazne dla użytkownika, ale także coraz mniej wymagające, jeśli chodzi o jego wiedzę i przygotowanie statystyczne.

Posługują się więc nimi kolejne grupy naukowców, a także nienaukowców, którzy słabo albo wcale nie znają fundamentów teorii, z której wywodzą się stosowane w tych programach metody, techniki i formuły analityczne. Przekonanie tej części użytkowników o tym, że niewiele trzeba wiedzieć o statystyce, aby wykonać skomplikowane numerycznie obliczenia i analizy, zostaje dodatkowo wzmocnione promocją oprogramowania i wieloma innymi zabiegami marketingowymi ich producentów i handlowców. W tych warunkach, kiedy z jednej strony reklamowana jest na rynku coraz bardziej zaawansowana oferta oprogramowania statystycznego, a z drugiej rośnie liczba osób korzystających z tego

oprogramowania, którzy często nawet nie deklarują zainteresowania teorią statystyki, dochodzić musi do nieporozumień, przekłamań i naukowych nadużyć.

Inną z przyczyn wskazanych wyżej zagrożeń jest zmiana warunków, w jakich współcześnie wykonywane są badania reprezentacyjne. Odczuwany przez wielu respondentów nadmiar i przesyt badaniami ankietowymi, obejmującymi wiele sfer życia indywidualnego i społecznego, doprowadził do znacznego spadku rangi tych badań i zainteresowania nimi. To z kolei powoduje, że z roku na rok rośnie frakcja respondentów odmawiających współpracy z ankieterami, a ich rezygnacje z poddania się pomiarowi sondażowemu generują kłopotliwy dla statystyków błąd nielosowy.

Niestety, błąd braków odpowiedzi nie jest jedynym błędem, którego wielkość rośnie w ostatnich dekadach. Wzrasta bowiem – co nietrudno uzasadnić – znaczenie błędów nielosowych w ogóle<sup>1</sup>. Wpływ tych błędów na jakość wnioskowania jest współcześnie większy niż błędu losowego. Tymczasem wśród wielu badaczy, zwłaszcza o małym doświadczeniu w zastosowaniach metody reprezentacyjnej, brak jest świadomości zarówno istnienia tej kategorii błędów (nielosowych), jak i ich konsekwencji. W klasycznej teorii wnioskowania bowiem jedynym błędem, który jest brany pod uwagę w konstruowaniu narzędzi estymacji i weryfikacji hipotez, jest błąd losowy (*random error*). Interpretacja wyników wnioskowania odniesiona zostaje jedynie do błędu losowania w powtarzalnych próbach losowych. Taki jest matematyczny (probabilistyczny) model wnioskowania statystycznego.

Statystyk nie powinien jednak abstrahować od praktycznych aspektów realizacji badań próbkowych. Nieuwzględnienie konsekwencji oddziaływania takich błędów nielosowych, jak: błąd pokrycia (*coverage error*), błąd braków odpowiedzi (*nonresponse error*), błąd pomiaru (*measurement error*), błąd przetwarzania danych (*postsurvey processing error*) podważa wiarygodność wnioskowania statystycznego. Mimo to w wielu opracowaniach naukowych można się spotkać z całkowitym pomijaniem oddziaływania tych błędów na uzyskane wyniki.

Kolejnym źródłem kłopotów z wiarygodnością badań próbkowych jest uparte dążenie niektórych badaczy do wykorzystania metod i narzędzi wnioskowania statystycznego w sytuacjach, gdy dysponują oni próbą nieprobabilistyczną – najczęściej próbą celową (*purposive sample, judgmental sample*) lub uzyskaną techniką wyboru według wygody

---

<sup>1</sup> Szerzej na ten temat por. [Szreder, 2015].

(*convenience sampling*)<sup>2</sup>. Takie postępowanie oznacza brak spełnienia najważniejszego założenia wstępnego, na którym oparta jest cała teoria estymacji i weryfikacji hipotez. Konsekwencją tego jest utrata możliwości interpretowania w kategoriach probabilistycznych zarówno właściwości statystyk z próby (estymatorów), wyników estymacji przedziałowej, jak i najważniejszych rozstrzygnięć w procesie testowania hipotez. Sens traci w tych warunkach częstościowa interpretacja prawdopodobieństwa, bo odnieść by ją trzeba do procesu generowania prób, który nie jest losowy. „W takich sytuacjach powinniśmy zaakceptować twardą prawdę, że statystyczne wnioskowanie nie jest możliwe” – stwierdzają słusznie Hirschauer et al. [2021].

Bardziej subtelny problem związany z zastosowaniami wnioskowania statystycznego dotyczy obecnej w nim fundamentalnej kategorii, jaką jest istotność statystyczna, a także sposobów rozstrzygania o niej. Zagadnieniom tym poświęcono w ostatnich latach dużo uwagi w środowiskach statystyków na całym świecie. W szczególności problem ten dotyczy przypisywania wskaźnikowi p-value zbyt dużej roli w procesie testowania hipotez. Krytykuje się w szczególności tendencję do jego dychotomizacji. Rozstrzygnięcie o odrzuceniu testowanej hipotezy wyłącznie na podstawie tego, czy wartość p-value przekracza czy nie przekracza progu 5%, stanowi trudne do zaakceptowania uproszczenie. Trzeba być bowiem świadomym tego, że obliczona wartość p-value odnosi się jedynie do wyniku pojedynczej próby, a test statystyczny nie jest narzędziem, które potrafi inherentną niepewność zawartą we wnioskowaniu przekształcić w pewność decyzji o prawdziwości lub nieprawdziwości hipotezy.

Poleganie na tego rodzaju dychotomicznym rozstrzygnięciu, bez analizy rzeczywistych wartości p-value, pozbawia nas niemal zupełnie możliwości rozróżnienia pomiędzy statystycznymi wynikami a naukowymi konkluzjami – stwierdza Goodman [1999]. Uzyskana w badaniu wartość p-value jest zwykle wynikiem działania kombinacji czynników losowych (*random variation*), ale także czynników wyrażających konsekwencje odstępstw od przyjętych założeń statystycznego modelu wnioskowania.

---

<sup>2</sup> Jedną z form wyboru próby według wygody, a więc techniką nazywaną także wyborem przypadkowym (por. [Szreder, 2022]), jest zyskująca na popularności autooselekcja respondentów, zwłaszcza w ankietach internetowych. Odwoływanie się do internetowych paneli respondentów jest tylko jednym z przejawów ogólniejszej tendencji, polegającej na rezygnowaniu z prób losowych na rzecz prób uzyskiwanych technikami nieprobabilistycznymi. Powody tego są różne i najczęściej dotyczą kwestii praktycznych. Często są nimi: brak lub niemożność sporządzenia operatu losowania, wysokie koszty utworzenia operatu, niższe niż w przypadku prób losowych koszty dotarcia do respondentów, a zwykle także krótszy czas niezbędny do dokonania pomiaru sondażowego (obserwacji statystycznej).

W szczególności, mała wartość p-value może oznaczać nieprawdziwość hipotezy zerowej, ale może też oznaczać, że źle określono niektóre matematyczne założenia modelu, albo że obserwacje nie były generowane przez proces losowy, lub też że dały znać o sobie wspomniane wcześniej błędy nielosowe. Z kolei duża (większa od 5%) wartość p-value nie dowodzi prawdziwości hipotezy zerowej. Informuje jedynie o stopniu zgodności wyników próby z wartością parametru populacji, którego dotyczy hipoteza zerowa. W odpowiednim przedziale ufności mieści się bowiem wiele (nieskończenie wiele) innych wartości tego parametru, z którymi mogą być zgodne obserwacje próbkowe<sup>3</sup>.

Negatywne skutki korzystania z metod wnioskowania statystycznego w okolicznościach, gdy nie są spełnione ważne założenia teorii wnioskowania, lub z powodu daleko idących uproszczeń w interpretacji wyników wnioskowania (zwłaszcza w przypadku weryfikacji hipotez), są widoczne w różnych dziedzinach nauki i życia. Czasami jest je łatwo dostrzec w szeroko komunikowanych wynikach badań społecznych, ekonomicznych czy politycznych. Niekiedy jednak są one widoczne dopiero po pewnym czasie i zwykle najpierw dyskutowane bywają w środowiskach statystyków.

Jednym z najwyraźniej dostrzegalnych przejawów kłopotów związanych z właściwym stosowaniem wnioskowania statystycznego jest upadek prestiżu sondaży politycznych, w szczególności wyborczych. Dotyczy to zarówno badań sondażowych realizowanych w Polsce, jak i w innych krajach. W szczególności w Polsce ich systematycznie słabnąca jakość – będąca skutkiem ignorowania ważnych założeń modelu wnioskowania statystycznego, pośpiechu i niestaranności – doprowadziła do stanu, w którym indywidualny sondaż nie jest już prawie przez nikogo traktowany jako wiarygodny. Może i lepiej, że coraz więcej osób nie wiąże tego typu sondaży w ogóle z badaniem statystycznym i z nauką. Tymczasem, gdy tego samego typu badanie (sondażowe) w odniesieniu do procesu wyborczego jest w sposób profesjonalny przygotowane i rzetelnie wykonane, daje ono wyniki wiarygodne i dokładne.

Dowodzą tego liczne prognozy rezultatu wyborczego sporządzone na podstawie badania reprezentacyjnego realizowanego w dniu wyborów, zwanego *exit poll*. Podstawą wnioskowania w tym badaniu jest próba losowa, której wielkość nie przekracza najczęściej jednego procenta liczebności populacji. W ostatnich latach w Polsce błąd badania *exit poll* nie przekraczał 1,6 pkt. proc. w odniesieniu do żadnego uczestniczącego w wyborach ugrupowania lub osoby – kandydata na urząd. Natomiast średni błąd był jeszcze mniejszy i

---

<sup>3</sup> Szerzej na ten temat por. [Szreder, 2019].

na przykład w wyborach prezydenckich w 2020 r. wyniósł w pierwszej i drugiej turze odpowiednio: 0,7 i 0,6 pkt. proc.<sup>4</sup> A w przedterminowych wyborach prezydenta Rzeszowa w czerwcu 2021 r. nie przekroczył 1,1 pkt. proc.<sup>5</sup>

Podobną precyzję osiąga się w badaniach exit poll w innych krajach. Nie ma więc żadnych podstaw, aby kwestionować samą metodę badawczą, gdy zestawia się ze sobą słabą jakość przedwyborczych sondaży z wysokiej jakości badaniami *exit poll*. Realizatorzy sondaży w okresach między wyborami powinni dużo bardziej rygorystycznie dbać o respektowanie naukowych podstaw statystycznych badań reprezentacyjnych.

Inny charakter mają te źródła problemów związanych ze stosowaniem wnioskowania statystycznego, które nie tyle wynikają z niespełnienia założeń teorii statystycznej, co z naturalnego w pewnym sensie dążenia badaczy do eliminowania niepewności przypisanej każdemu rezultatowi statystycznego wnioskowania. Amrhein, Trafimow i Greenland [2019, s. 264] stwierdzają słusznie, że: „nadużywa się statystyki jako maszyny do automatycznego podejmowania naukowych decyzji, zarówno w odniesieniu do weryfikowanych hipotez, jak i w procesie selekcji artykułów kierowanych do publikacji”. Konsekwencje takiego działania są dla statystyki oraz dla całej nauki coraz poważniejsze. Dostrzegany jest spadek zaufania do prac naukowych opartych na statystycznych badaniach próbkowych. Widoczne jest to przede wszystkim w redakcjach czasopism o uznanym międzynarodowym prestiżu. Interesujące wyniki badania (*nota bene* próbkowego) na ten temat w czasopiśmiennictwie z zakresu finansów prezentują w swojej pracy Baker, H.K. i T.K. Mukherjee [2007]. Autorzy podkreślają, że żadna z ankietowanych przez nich redakcji spośród czołowych światowych pism naukowych nie zgodziła się ze stwierdzeniem, że naukowe dociekania oparte na wykonanym badaniu ankietowym powinny być traktowane na równi z innymi oryginalnymi badaniami (*survey-based research should be considered equal to other types of original research*). Rolę badań ankietowych widzą jedynie jako uzupełniającą (*complementary*). Najdalej jednak w swojej rezerwie do ocen i rozstrzygnięć wnioskowania statystycznego poszła redakcja czasopisma „Basic and Applied Social Psychology” (BASP), która już w 2015 roku oznajmiła, że nie będzie publikowała prac zawierających elementy statystycznej procedury weryfikacji

---

<sup>4</sup> Pięć lat wcześniej w wyborach prezydenckich w Polsce maksymalny błąd w exit poll wyniósł w obu turach odpowiednio: 1,6 oraz 1,5 pkt. proc. W wyborach parlamentarnych tego samego roku błąd dla żadnego komitetu wyborczego nie był wyższy niż 1,6 proc.

<sup>5</sup> Por. [Szreder, 2021].

hipotez statystycznych (NHSTP)<sup>6</sup>, a dokładniej, że przed publikacją autor artykułu będzie musiał z niego usunąć wszelkie ślady zastosowań tej procedury: „*prior to publication, authors will have to remove all vestiges of the NHSTP (p-values, t-values, F-values, statements about “significant” differences or lack thereof, and so on)*”<sup>7</sup>. Klasyczną procedurę weryfikacji hipotez statystycznych redakcja uznała za nieskuteczną (*invalid*). W rzeczywistości za niepoprawną należałoby uznać nie samą procedurę testowania hipotez, a raczej praktykę postępowania w tym zakresie, ze wszystkimi uproszczeniami i błędami, o których wspomniano w tej części opracowania.

W następnym punkcie przedstawione są z kolei najważniejsze wyzwania dotyczące metod uczenia statystycznego, a więc odnoszące się głównie do metod eksploracyjnych.

### 3. Wyzwania stojące przed badaniami za pomocą metod uczenia statystycznego

#### 3.1. Istota uczenia statystycznego

Formalną definicję uczenia statystycznego wprowadził Vladimir Vapnik [1998]. Ogólny model uczenia statystycznego opisuje on przez podanie trzech jego komponentów:

- generatora (G) wektorów losowych  $x \in R^n$ , wylosowanych niezależnie z ustalonego, lecz nieznanego rozkładu prawdopodobieństwa, określonego funkcją  $F(x)$ ,
- nauczyciela (S) który przypisuje wartość wyjściową  $y$  do każdego wektora wejściowego  $x$ , w zależności od warunkowego rozkładu  $F(y/x)$ , także ustalonego lecz nie znanego,
- maszyny uczącej (LM) zdolnej do wdrożenia zbioru funkcji  $f(x, \alpha)$ ,  $\alpha \in A$ , gdzie  $A$  jest zbiorem parametrów.

Problem uczenia polega na znalezieniu z danego zbioru funkcji  $f(x, \alpha)$ ,  $\alpha \in A$ , która najlepiej aproksymuje reakcję nauczyciela.

W przytoczonej definicji kluczową rolę odgrywa maszyna ucząca (LM), która jest narzędziem uczenia maszynowego. Ze względu na to iż generator G ma z reguły charakter losowy, to funkcja  $f(x, \alpha)$ , łącząca dane wejściowe z danymi wyjściowymi, ma z reguły charakter indeterministyczny, stochastyczny, więc uczenie maszynowe najczęściej jest rozumiane jako jest rozumiane jako uczenie statystyczne, oparte na regułach probabilistycznych [Hastie, Tibshirani, Friedman, 2009].

---

<sup>6</sup> NHSTP jest popularnym w jęz. angielskim skrótem wyrażenia: the Null Hypothesis Significance Testing Procedure, czyli Neymanowsko-Pearsonowskiej procedury weryfikacji hipotez statystycznych.

<sup>7</sup> Trafimow i Marks (2015).

W procesie uczenia maszyna obserwuje pary  $(x,y)$  ze zbioru uczącego, zadanego przez nauczyciela, tak aby się nauczyć przypisać, korzystając z pewnej funkcji, dowolnemu  $x$  wartości  $\bar{y}$ , możliwie bliskie wartości  $y$  podanej przez nauczyciela. Wybór takiej pożądanej funkcji opiera się na zbiorze uczącym  $l$  obserwacji niezależnych i o identycznym rozkładzie, wylosowanych zgodnie z:

$$F(x,y) = F(x) F(y/x); (x_1,y_1), \dots, (x_l,y_l).$$

Aby wybrać najlepszą z dostępnych aproksymacji reakcji nauczyciela należy podać pewną miarę straty lub rozbieżności  $L(y, f(x,\alpha))$  pomiędzy odpowiedzią nauczyciela na dany sygnał wejściowy i odpowiedzią  $f(x,\alpha)$  daną przez maszynę uczącą. Oczekiwana wartość straty dana jest funkcjonałem ryzyka:

$$R(\alpha) = \int L(y, f(x, \alpha)) dF(x, y).$$

Zadaniem matematycznym jest znalezienie funkcji  $f(x,\alpha_0)$ , która minimalizuje funkcjonał  $R(\alpha)$  w klasach funkcji  $f(x,\alpha)$ ,  $\alpha \in A$  w sytuacji, gdy łączne prawdopodobieństwo  $F(x,y)$  jest nieznane i gdy jedynie posiadana informacja jest zawarta w zbiorze uczącym. Z powyższej, skrótowo zarysowanej istoty uczenia statystycznego wynika, że probabilistyczny charakter procesów społeczno-gospodarczych stanowi najbardziej ogólną platformę łączenia metod analizy danych, opartych na modelach sztucznej inteligencji, ze statystyczno-ekonometrycznymi metodami modelowania oraz prognozowania tych procesów.

### 3.2. Zasady uczenia statystycznego

W ogólnym ujęciu uczenia statystycznego mamy zmienną  $Y$  rozumianą w statystyce jako zmienna objaśniana a w uczeniu statystycznym jako zmienna wynikowa oraz  $k$  zmiennych objaśniających (predyktorów)  $X_1, X_2, \dots, X_k$ . Zakładamy że pomiędzy  $Y$  oraz  $X = (X_1, X_2, \dots, X_k)$  istnieje pewna relacja, którą ogólnie możemy zapisać jako:

$$Y = f(X) + \xi,$$

gdzie  $f$  jest nieznaną funkcją wiążącą  $Y$  z  $X$  a  $\xi$  - jest składnikiem losowym. Istotą uczenia statystycznego jest odgadywanie rzeczywistej funkcji  $f$  za pomocą funkcji  $h$ , będącej jedną z hipotez dotyczących tej nieznannej funkcji  $f$ , należącej do przestrzeni hipotez  $H$ , w której mieści się  $f$  [Hastie, Tibshirani, Friedman, 2009].

Istnieją dwa zasadnicze powody dla których staramy się odgadnąć funkcję  $f$ . Pierwszy z nich, o charakterze praktycznym, to predykcja  $Y$  na podstawie znajomości  $X$ . Drugi o charakterze poznawczym, to wnioskowanie o  $Y$  na podstawie  $X$ . Jeśli chcemy wykorzystać metody uczenia statystycznego dla celów wnioskowania o tym jaki jest sposób powiązania



między zmienną objaśnianą a zmiennymi objaśniającymi, to nie możemy traktować  $\hat{f}$ , będącej statystycznym szacowaniem rzeczywistej funkcji  $f$ , jako czarną skrzynkę, lecz musimy przyjąć konkretną postać funkcji  $\hat{f}$ . Natomiast jeśli chcemy dokonać możliwie dokładnej predykcji zmiennej wynikowej, to funkcję tą traktujemy jako czarną skrzynkę przy pomocy której dokonujemy możliwie najdokładniejszego szacunku wartości zmiennej wynikowej w przyszłości, w świetle przyjętych miar dokładności predykcji.

Kluczowym problemem uczenia statystycznego jest więc poznanie rzeczywistej funkcji  $f$ . Szacujemy ją na podstawie zbioru danych, który nazywamy zbiorem uczącym, zawierającym informacje o danych wejściowych i wyjściowych  $(x_{ij}, y_i)$ . Innymi słowy, poszukujemy takiej funkcji  $\hat{f}$ , dla której  $Y \approx \hat{f}(X)$ , dla dowolnej pary obserwacji ze zbioru  $(X, Y)$ . Możemy tutaj stosować podejście parametryczne lub nieparametryczne.

Parametryczna metoda uczenia statystycznego polega na sprecyzowaniu postaci analitycznej funkcji  $\hat{f}$ . Nieparametryczne metody uczenia statystycznego nie przyjmują wyraźnych założeń co do postaci analitycznej funkcji dotyczących  $f$ . Zamiast tego poszukują takiej postaci funkcji  $f$ , która dopasowuje się możliwie najlepiej do danych ze zbioru uczącego. Podejście nieparametryczne może mieć dużą przewagę nad podejściem parametrycznym, gdyż unikając przyjęcia określonej postaci analitycznej funkcji  $f$ , może ona dokładniej dopasować się do danych empirycznych. Podejście parametryczne niesie niebezpieczeństwo tego, że postać analityczna funkcji  $\hat{f}$  bardzo może odbiegać od rzeczywistej funkcji  $f$ , wiążącej ze sobą predyktory ze zmienną wynikową. Niemniej jednak, podejście nieparametryczne posiada tę wadę, że nie zmniejsza liczby szacowanych parametrów do tylko istotnych, a przez to wymaga o wiele liczniejszego zbioru uczącego [James i in., 2013].

Zaletą metod uczenia statystycznego jest to, że obejmują one zarówno podejście parametryczne jak i nieparametryczne. Wykraczają więc one daleko poza standardowe zastosowania klasycznych metod regresji, szczególnie w postaci regresji liniowej. Stąd bierze się stosowanie różnych metod uczenia statystycznego dla jednego zbioru uczącego, a następnie, na podstawie kryteriów empirycznych, wybór najlepszej metody. W uczeniu statystycznym żadna metoda nie dominuje nad wszystkimi innymi, we wszystkich możliwych zbiorach danych. W przypadku konkretnego zestawu danych określona metoda może działać najlepiej, ale inna metoda może działać lepiej na podobnym, ale nieco innym zestawie danych. W związku z tym ważnym zadaniem jest zdecydowanie, przy określonym zestawie danych, która metoda daje najlepsze wyniki. Jednak wybór

najlepszego podejścia jest jedną z najtrudniejszych decyzji, przy zastosowaniach praktycznych metod uczenia statystycznego.

### **3.3. Panorama metod uczenia statystycznego**

Zaprezentowany poniżej przegląd metod uczenia statystycznego nie jest wyczerpujący, ale ma za zadanie pokazać, że paleta metod uczenia statystycznego jest o wiele szersza niż zestaw klasycznych metod statystyki matematycznej. Liniowe modele uczenia statystycznego, takie jak liniowa funkcja regresji, model logitowy, liniowa funkcja dyskryminacyjna, będące narzędziami klasycznej statystyki matematycznej, mogą być szacowane na podstawie danych z próby jak również uczone na podstawie zbioru danych uczących, w kategoriach uczenia maszynowego. Nieliniową, nieparametryczną alternatywę w stosunku do liniowych, parametrycznych modeli stanowi metoda k-najbliższych sąsiadów. Stąd widać, że w badaniach empirycznych mamy alternatywę pomiędzy podejściem parametrycznym a nieparametrycznym. Znane są rozszerzenia i uogólnienia tych metod idące w kierunku łączenia podejścia parametrycznego z nieparametrycznym. Tego wyrazem są uogólnione modele liniowe oraz uogólnione modele addytywne, które są modelami typu semiparametrycznego.

Innym kierunkiem zastępowania ujęcia parametrycznego przez nieparametryczne są estymatory i klasyfikatory jądrowe. Estymacja jądrowa pozwala na wyznaczanie gęstości rozkładu zmiennej losowej, na podstawie uzyskanej próby lub zbioru uczącego. Szacowanie gęstości jądra metodą uczenia nienadzorowanego, prowadzi do rodziny procedur klasyfikacji nieparametrycznej, nazywanej klasyfikatorami jądrowymi. Zastosowanie klasyfikatora jądrowego pozwala na uzgadnianie nieliniowych zależności w analizowanych zbiorach danych z liniowym charakterem otrzymywanych klasyfikatorów. Alternatywą jest stosowanie modeli mieszanek rozkładów.

Ważną grupą metod uczenia statystycznego są metody drzewkowe, opierające się na zasadzie podziału rekurencyjnego, wśród nich metoda klasyfikacji i regresji oparta na drzewach decyzyjnych, realizowana w postaci różnych algorytmów. Rozwinięciem modeli indywidualnych, o stosunkowo prostej architekturze, jest podejście wielomodelowe, polegające na łączeniu modeli indywidualnych w model zagregowany, który z reguły jest bardziej dokładny niż którykolwiek z indywidualnych modeli wchodzących w jego skład, co pozwala na poprawę zdolności predykcyjnej. Spośród wielu znanych algorytmów podejścia wielomodelowego najważniejsze znaczenie mają: metoda agregacji

bootstrapowej (*bagging*), metoda wzmacniania (*boosting*) oraz lasy losowe (*random forests*).

Oryginalną i szeroko stosowaną metodą, stosowaną zarówno dla celów klasyfikacyjnych jak i regresyjnych, jest metoda wektorów nośnych (*support vectors machine*). Wykorzystuje ona ideę klasyfikacji za pomocą hiperpłaszczyzn dyskryminacyjnych. Jest to metoda bardzo elastyczna. Umożliwia ona również znalezienie rozwiązania, gdy klas w próbie uczącej nie da się rozdzielić w sposób liniowy, czyli że można ją rozpatrywać w wariacie liniowym i nieliniowym. Hiperpłaszczyzna rozdzielająca klasy uzupełniona jest dwiema równoległymi prostymi wyznaczającymi pewien pas separujący klasy, nazywany wektorem nośnym, co pozwala na ocenę stabilności dokonywanych klasyfikacji.

Sztuczne sieci neuronowe jako systemy maszynowego przetwarzania informacji są modelami zaliczanymi do sztucznej inteligencji. Prawidłowo skonstruowana sztuczna sieć neuronowa może służyć do rozwiązywania wielu typów zagadnień regresyjnych, klasyfikacyjnych oraz prognostycznych. Istotną sprawą jest wybór architektury sieci. Wyróżnia się tutaj trzy główne typy architektur: sieci jednokierunkowe, rekurencyjne oraz komórkowe. Dla każdego typu architektury mamy odmienne procedury ich budowania. Jednak proces uczenia sieci jest zawsze procesem indeterministycznym, czyli wynik uczenia nie jest nigdy całkowicie jednoznaczny. Sztuczne sieci neuronowe są najstarszymi i najczęściej stosowanymi metodami uczenia statystycznego, opartymi na idei sztucznej inteligencji.

### **3.4. Paradygmat uczenia statystycznego**

Scharakteryzowane uprzednio przesłanki i zasady uczenia się z danych pozwalają na sformułowanie paradygmatu uczenia statystycznego [Pociecha, 2021]. Już w połowie lat dziewięćdziesiątych ubiegłego wieku Vladimir Vapnik zwrócił uwagę na rewolucję zachodzącą w metodologii badań statystycznych w postaci tworzenia się nowego paradygmatu tych badań – paradygmatu uczenia statystycznego [Vapnik, 1998]. Zauważył on, że klasyczny (Fisherowski), paradygmat statystyki matematycznej, sformułowany w latach 20-tych i 30-tych ubiegłego wieku jest zastępowany przez nowy paradygmat. Według Vapnika problem uczenia się z danych jest tak ogólny, że prawie każdy problem statystyczny może być formułowany równolegle w teorii uczenia maszynowego, która w sytuacji uwzględniania losowości danych, staje się uczeniem statystycznym.

Główne cechy paradygmatu uczenia statystycznego przedstawione zostaną w kontekście różnic pomiędzy nim a klasycznym paradygmatem statystyki matematycznej. Punktem

wyjścia paradygmatu statystyki matematycznej jest teoria rachunku prawdopodobieństwa i statystyki matematycznej oraz sprawdzanie na ile dane empiryczne można wpasować w teoretyczne ramy statystyki matematycznej. W paradygmacie uczenia statystycznego punkt wyjścia jest odwrotny: „niech dane mówią same za siebie”, „uczmy się z danych”, punktem wyjścia jest dostępny zbiór danych.

Kluczowym pojęciem paradygmatu statystyki matematycznej jest pojęcie populacji generalnej oraz próby. Paradygmat uczenia statystycznego ignoruje pojęcie populacji i próby. W to miejsce przyjmuje, że dysponujemy odpowiednim zbiorem danych empirycznych na podstawie którego efektywnie możemy dokonywać predykcji a także wnioskować o rzeczywistości z której pochodzą te dane. Zbiorem tym może być próba, nawet mała, wylosowana zgodnie z regułami metody reprezentacyjnej, może to być zbiór danych oczyszczonych, zgodnie z regułami czyszczenia danych a mogą to być także zbiory Big Data, w tym dane strumieniowe.

Istotą paradygmatu uczenia statystycznego jest tworzenie systemów samouczących się, czyli poprawiających się automatycznie poprzez doświadczenie.

W paradygmacie klasycznego wnioskowania statystycznego kluczową rolę odgrywa pojęcie prawdopodobieństwa. W paradygmacie uczenia statystycznego, ze względu na wątpliwości co do uznania w pewnych wypadkach zbioru uczącego za zbiór danych losowych, odgrywa ono rolę drugorzędną. W tym sensie paradygmat uczenia statystycznego zbliża się do paradygmatu statystyki opisowej.

Paradygmat uczenia statystycznego jest platformą badawczą bardziej uniwersalną, gdyż wchłonął on w istocie paradygmat wnioskowania statystycznego, kosztem osłabienia jego pierwotnych założeń. Paradygmat ten daje szeroką możliwość spożytkowania możliwości obliczeniowych współczesnych komputerów i wykorzystania dużych zbiorów danych, jakie dostarcza współczesne życie społeczno-gospodarcze. W badaniach statystycznych jest miejsce zarówno na klasyczny paradygmat statystyki matematycznej jak i nowy paradygmat uczenia statystycznego, oparty na modelach sztucznej inteligencji.

## **Literatura**

Amrhein, V., Trafimow, D., Greenland, S. [2019], *Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication*, The American Statistician, Vol. 73, Issue sup 1, pp. 262-270.

Baker, H.K., Mukherjee, T.K. [2007], *Survey Research in Finance: Views from Journal Editors*, International Journal of Managerial Finance, Vol. 3, Issue 1, pp.11-25.

- Goodman, S.N. [1999], *Toward evidence-based medical statistics. The P value fallacy*, Annals of Internal Medicine, Vol. 130, Issue 12, pp. 995-1004.
- Hastie T., Tibshirani R., Friedman J. (2009), *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, New York.
- Hirschauer, N., Grüner, S., Mußhoff, O., Becker, C., Jantsch, A. [2021], *Inference using non-random samples? Stop right there!*, Significance, Vol. 18, Issue 5, pp. 20-24.
- Jajuga K. [2019], *Nauki ekonomiczne – dylematy klasyfikacji dyscyplin. Tendencje zmian*, w: *Ewolucja nauk ekonomicznych*, PAN, s. 140-150, Warszawa.
- James G., Witten D., Hastie T., Robert Tibshirani R. [2013], *An Introduction to Statistical Learning*, Springer, New York.
- McCulloch W., Pitts W. [1943], *A Logical Calculus of Ideas Immanent in Nervous Activity*, Bulletin of Mathematical Biophysics, 5, s. 115–133.
- Merton R.C. [1994], *Influence of Mathematical Models in Finance on Practice: Past, Present and Future*, Philosophical Transactions, 1684, pp. 451-463, Royal Society of London.
- Pociecha J. [2021], *The paradigm of statistical inference and the paradigm of statistical learning*, Przegląd Statystyczny, vol. 68, 1, p. 1-16.
- Szreder, M. [2022], *Szanse i iluzje korzystania z dużych prób we wnioskowaniu statystycznym*, Wiadomości Statystyczne nr 8, s. 1-16.
- Szreder M. [2015], *Zmiany w strukturze całkowitego błędu badania próbkowego*, Wiadomości Statystyczne, nr 1, s. 4-12.
- Szreder, M. [2019], *Istotność statystyczna w czasach big data*, Wiadomości Statystyczne, nr 11, s. 42-57.
- Szreder, M. [2021], *Exit poll - optymistyczna lekcja z wyborów w Rzeszowie*. <https://wyborcza.pl/7,162657,27208314,exit-poll-optymistyczna-lekcja-z-wyborow-w-rzeszowie.html>, (25.10.2022).
- Shalev-Shwartz S., Ben-David S. [2014], *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press, Cambridge.
- Trafimow, D., Marks, M. [2015], *Editorial*, Basic and Applied Social Psychology, Vol. 37, Issue 1, pp. 1-2.
- Tukey J.W. [1977], *Exploratory Data Analysis*, Pearson, New York.
- Vapnik V. [1998], *Statistical Learning Theory*, Wiley, New York.

Profesor nauk ekonomicznych, wykładowca Uniwersytetu Ekonomicznego we Wrocławiu, Politechniki Warszawskiej, Uniwersytetu Ekonomicznego w Krakowie, Akademii WSB w Dąbrowie Górniczej oraz uczelni za granicą. Doktor honorowy Uniwersytetu Ekonomicznego w Krakowie, Akademii WSB w Dąbrowie Górniczej i profesor honorowy Politechniki Warszawskiej. W swojej pracy naukowo-badawczej koncentruje się wokół statystyki i ekonometrii, finansów, zarządzania ryzykiem, nieruchomości i nowych technologii. Kierownik Katedry Inwestycji Finansowych i Zarządzania Ryzykiem na Uniwersytecie Ekonomicznym we Wrocławiu. Jest przewodniczącym Komitetu Statystyki i Ekonometrii Polskiej Akademii Nauk oraz Zespołu Nauk Społecznych Rady Doskonałości Naukowej.

**Prof. dr hab. Józef Pociecha**

Emerytowany profesor Uniwersytetu Ekonomicznego w Krakowie i kierownik Katedry Statystyki. Autor i współautor prac z zakresu metod klasyfikacji i analizy danych, zastosowań metod ilościowych w analizach procesów demograficznych metod probabilistycznych w auditingu oraz metod prognozowania bankructwa firm. Przewodniczący Komisji Nauk Ekonomicznych Polskiej Akademii Umiejętności oraz przewodniczący Komisji Nauk Ekonomicznych i Statystyki PAN, Oddział w Krakowie.

**Prof. dr hab. Mirosław Szreder**

Profesor nauk ekonomicznych specjalizujący się w rozwoju oraz zastosowaniach metod wnioskowania statystycznego w problematyce ekonomicznej i społecznej. Autor m.in. dwóch wydań książki pt. „Metody i techniki sondażowych badań opinii”. Popularyzator nauki w dzienniku „Rzeczpospolita” oraz w tygodnikach „Polityka” i „Tygodnik Powszechny”. Pełni funkcję dziekana Wydziału Zarządzania Uniwersytetu Gdańskiego. Jest członkiem Komitetu Statystyki i Ekonometrii PAN oraz Naukowej Rady Statystycznej.