Statystyka jako gramatyka nowoczesnej racjonalności

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje rolę statystyki jako fundamentu współczesnego myślenia racjonalnego, wykraczając poza proste operacje matematyczne. Centralnym punktem rozważań jest cykl PPDAC (Problem, Plan, Dane, Analiza, Konkluzja), który autor określa mianem 'miniaturowej konstytucji' dla działań opartych na danych. Tekst zgłębia filozoficzne napięcia między klasycznym podejściem częstotliwościowym a subiektywną epistemiką bayesowską, rzucając światło na ograniczenia tradycyjnych metod, takich jak p-wartość czy błędy statystyczne. Czytelnik dowie się, jak nowoczesne techniki, w tym bootstraping i przedziały wiarygodności, pomagają zarządzać niepewnością epistemiczną w naukach społecznych. To głębokie spojrzenie na to, jak struktury statystyczne kształtują nasze rozumienie rzeczywistości i procesy podejmowania decyzji w skomplikowanym świecie.

This article examines the role of statistics as a foundation for contemporary rational thinking, extending beyond simple mathematical operations. Central to the discussion is the PPDAC (Problem, Plan, Data, Analysis, Conclusion) cycle, which the author describes as a 'miniature constitution' for data-driven action. The text explores the philosophical tensions between the classical frequentist approach and subjective Bayesian epistemics, illuminating the limitations of traditional methods such as p-values and statistical errors. The reader will learn how modern techniques, including bootstrapping and credibility intervals, help manage epistemic uncertainty in the social sciences. This is a profound look at how statistical structures shape our understanding of reality and decision-making in a complex world.

Artykuł dekonstruuje pozornie neutralny cykl PPDAC (Problem, Plan, Dane, Analiza, Konkluzja) używany w statystyce, ukazując go jako potężny instrument kształtowania społecznego postrzegania. Autor argumentuje, że każdy etap tego procesu, od formułowania pytania badawczego po wyciąganie wniosków, jest głęboko osadzony w normatywnych założeniach i politycznych decyzjach. Analiza danych, interpretacja prawdopodobieństwa i

regresji, a także rozumienie populacji, stają się areną walki o uprzywilejowaną perspektywę i definicję rzeczywistości. Kluczowa teza mówi, że statystyka nie jest obiektywnym odzwierciedleniem świata, lecz narzędziem, które może zarówno demistyfikować, jak i utrwalać nierówności. W dobie algorytmów i uczenia maszynowego, krytyczne zrozumienie cyklu PPDAC staje się niezbędne do świadomego kształtowania przyszłości, w której liczby służą publicznemu rozumowi, a nie technokratycznej kontroli.

SŁOWA KLUCZOWE

PPDAC | statystyka empiryczna | prawdopodobieństwo subiektywne | p-wartość | współczynnik Bayesa | bootstraping | przedział ufności | niepewność epistemiczna | regresja | hipoteza zerowa | racjonalność | populacja | wnioskowanie statystyczne | szanse początkowe | błąd drugiego rodzaju

Cykl PPDAC: fundament nowoczesnej epistemiki

W podręcznikach statystyki cykl PPDAC przedstawiany jest zazwyczaj jako pięć niewinnych kroków: Problem, Plan, Dane, Analiza, Konkluzja. W rzeczywistości mamy tu do czynienia z czymś znacznie potężniejszym – miniaturową konstytucją dla wszelkiego działania opartego na danych. Problem zakreśla horyzont pytania, Plan koduje decyzję o sposobie wejścia w świat empirii, Dane materializują selektywny wycinek tego świata, Analiza przekształca surowy zbiór obserwacji w struktury niosące znaczenie, a Konkluzja to efektowne, lecz ryzykowne przejście od liczb z próby do roszczeń do prawdy i słuszności w odniesieniu do całej populacji i praktyki społecznej.

Z tej perspektywy PPDAC nie jest jedynie techniczną procedurą, lecz protokołem, w którym roszczenia do zasadności ujawniają się na kolejnych etapach. W fazie Problemu badacz formułuje pytanie w języku, który siłą rzeczy jest zanurzony w świecie naszych wspólnych, nie zawsze uświadomionych założeń. Pytanie o efektywność programu antydyskryminacyjnego, nierówności płacowe czy skuteczność terapii onkologicznej nigdy nie jest czysto techniczne – od samego początku jest nasycone normatywnie. W fazie Planu trzeba podjąć decyzję, kto i co będzie się liczyć: czyje głosy włączymy do badanej populacji, które zmienne uwzględnimy, a które milcząco pominiemy. Dane nie spadają z nieba – są śladem, jaki zostawiają po sobie polityczne i organizacyjne decyzje.

Dopiero w fazie Danych rozpoczyna się to, co klasyczna statystyka lubi uważać za swój punkt zerowy: uruchomienie maszyn losujących, ankiet czy rejestrów administracyjnych. Jednak nawet tutaj jedynie odtwarzamy wcześniejsze wybory dotyczące definicji jednostki obserwacji, konstrukcji kwestionariusza czy

sposobu radzenia sobie z odmową udziału w badaniu. Faza Analizy, w której na scenę wkraczają p-wartości, współczynniki regresji czy czynniki Bayesa, jest próbą informacyjnego zagęszczenia danych i oddzielenia sygnału od szumu. Jednocześnie generuje ona nowe artefakty, które ulegają reifikacji, czyli uprzedmiotowieniu: współczynnik staje się w debacie publicznej „wielkością efektu", p-wartość zamienia się w magiczną pieczęć „istotności", a przedział ufności traktowany jest jak wygodny zamiennik dla całej złożoności niepewności.

Etap Konkluzji jest zarazem najbardziej niebezpieczny i społecznie twórczy. To tutaj statystyk, epidemiolog czy inżynier danych zaczyna mówić do świata językiem potocznym: „a zatem program należy wzmocnić", „leku nie warto refundować", „algorytm można wdrożyć w praktyce kredytowej". W tym momencie roszczenia do prawdy, normatywnej słuszności i wiarygodności splatają się w jedną figurę. Musi ona zostać poddana krytyce nie tylko z perspektywy wewnętrznej elegancji modelu, ale przede wszystkim z punktu widzenia realnych konsekwencji dla społecznego działania.

W tym sensie PPDAC stanowi socjologiczny schemat, w którym społeczeństwo późnej nowoczesności uczy się postrzegać samo siebie jako populację mierzalnych obiektów. Jeśli przyjąć za Davidem Spiegelhalterem, że niepewność nie jest obiektywną właściwością rzeczy, lecz relacją między kimś a światem, to cykl PPDAC jest właśnie procedurą kształtowania tej relacji, a nie tylko mechaniczną operacją liczenia.

Częstotliwościowa vs subiektywna wizja prawdopodobieństwa

Jeżeli cykl PPDAC jest konstytucją, to liczby są językiem, w którym ta konstytucja ożywa. Statystyka empiryczna zaczyna się bowiem w momencie, gdy surowe obserwacje zostają przekształcone w rozkład: histogram, wykres punktowy czy pudełkowy. Może to być rozkład Poissona opisujący liczbę zabójstw w ciągu dnia, rozkład normalny wagi urodzeniowej, asymetryczny wykres liczby partnerów seksualnych czy rozkład zgadywanej liczby żelków w słoju. Lecz same liczby są czymś więcej niż tylko opisem świata; stanowią zarazem efekt kompromisu między różnymi filozofiami prawdopodobieństwa.

Klasyczny ideał częstotliwościowy głosi, że prawdopodobieństwo to graniczna proporcja zdarzeń w nieskończonej serii identycznych eksperymentów. W tym ujęciu szansa na wyrzucenie dwóch orłów w dwóch rzutach monetą wynosi jedną czwartą, ponieważ w milionach powtórzeń układy orzeł–orzeł, orzeł–reszka, reszka–orzeł i reszka–reszka wystąpią z niemal identyczną częstością. Taka konstrukcja, choć ma ogromną siłę pedagogiczną i porządkuje intuicję, okazuje się krucha w zderzeniu z rzeczywistością. W praktyce życia społecznego nikt nie dysponuje nieskończonym ciągiem identycznych prób; nikt nie zamierza przeżyć tej samej kampanii wyborczej czy tego samego kryzysu finansowego milion razy.

Tutaj na scenę wkracza interpretacja subiektywna, w której prawdopodobieństwo nie jest obiektywną cechą świata, lecz miarą postawy konkretnego aktora wobec niego. Bruno de Finetti nieprzypadkowo zatytułował swój esej „Prawdopodobieństwo nie istnieje", aby wymusić przesunięcie z metafizyki obiektywnych szans do pragmatyki rozsądnego zakładu. David Spiegelhalter idzie tym tropem, proponując, by konsekwentnie mówić „moja szansa na zdarzenie X", co podkreśla, że jest to zawsze relacja między określoną osobą, jej wiedzą a pewnym stanem rzeczy.

Z tej perspektywy rzut monetą przestaje być archetypem wszelkiej niepewności, a staje się zaledwie jednym z wielu modeli sytuacji, w których ktoś, dysponując określonymi informacjami, przypisuje liczby szansom i jest gotów zawierać zakłady. To przejście od obiektywistycznej metafizyki do relacyjnej epistemiki ma znaczenie nie tylko filozoficzne, ale i głęboko polityczne. Oznacza bowiem, że spór o prawdopodobieństwa jest zawsze także sporem o to, czyją perspektywę uznajemy za uprzywilejowaną: lekarza, pacjenta, analityka ryzyka w korporacji, unijnego regulatora danych czy właściciela platformy handlującej prognozami naszych zachowań.

Kiedy Spiegelhalter pisze, że jego książki są próbą uczenia się „powolnego myślenia o naszej niewiedzy", ustanawia tym samym etos stojący w ostrej opozycji do dominującej dziś logiki „natychmiastowego predykcyjnego przesądzania" przyszłości przez algorytmy. Zamiast traktować liczby jako nieodwołalne orzeczenia losu, proponuje widzieć w nich tymczasowe kondensacje niewiedzy, które muszą być nieustannie rewidowane w świetle nowych danych i nowych pytań.

Uczenie bayesowskie rozbija epistemiczny dogmatyzm

Na tym tle bayesowskie uczenie się można zrekonstruować jako formalizację zdroworozsądkowej zasady, że nasze przekonania powinny zmieniać się w uporządkowany sposób pod wpływem dowodów. Szanse początkowe są tu numerycznym zapisem naszej dotychczasowej wiedzy i doświadczenia, wiarygodność danych opisuje, jak dobrze poszczególne hipotezy „przewidują" obserwacje, a szanse końcowe stanowią nową równowagę przekonań, osiągniętą po zderzeniu tych dwóch elementów. To nic innego jak zdyscyplinowana aktualizacja poglądów w świetle faktów.

W przykładzie z monetami mamy do czynienia z czystą niepewnością epistemiczną: świat jest ustalony, w kieszeni leży konkretna moneta, lecz my po prostu nie wiemy, która. Intuicja podpowiadająca „pół na pół" po wyrzuceniu orła zdradza resztki myślenia w kategoriach symetrii, które ignoruje kluczowy fakt – moneta z dwoma orłami ma znacznie większą zdolność generowania obserwowanego wyniku. Rachunek bayesowski mówi wprost: jeśli w puli mamy równą liczbę monet uczciwych i skrajnych, to zaobserwowany orzeł jest dalece bardziej kompatybilny z hipotezą „skrajnie orłową". Pedagogiczna metafora częstości, którą posługuje się źródłowy tekst, jest próbą przełożenia tej logiki na język potoczny: liczymy po prostu, z ilu

teoretycznych orłów, jakie mogłyby wytworzyć wszystkie monety, pochodzi ten jeden, który trzymamy w dłoni.

Jeszcze dobitniej widać to w przykładzie testu antydopingowego, gdzie hipotezy nie dotyczą już struktury świata fizycznego, lecz moralno-prawnego statusu sportowca. Domyślne szanse początkowe wynikają z wiedzy o częstości dopingu w całej populacji, a zatem są już produktem wcześniejszych cykli PPDAC, poprzednich pomiarów i sporów o samą definicję niedozwolonych substancji. Test, który w warunkach laboratoryjnych wydaje się „dobry", z czułością i swoistością na poziomie dziewięćdziesięciu kilku procent, w populacji o niskim wskaźniku dopingu generuje jedynie umiarkowane powody do zmiany przekonań. Szanse końcowe rzędu jednej trzeciej oznaczają, że pozytywny wynik nie powinien automatycznie uruchamiać życiowej sankcji dyskwalifikacji, lecz raczej cały aparat dodatkowej weryfikacji i namysłu.

W świetle tego przykładu widać, że podejście bayesowskie jest nie tyle narzędziem do polowania na winnych, ile tarczą przeciwko epistemicznemu fanatyzmowi. Jeśli prawo – czy to antydopingowe, karne, czy dotyczące bezpieczeństwa – przekształca wynik testu statystycznego w automatyczny wyrok, nie jest to wina rachunku prawdopodobieństwa. Jest to wyraz kapitulacji rozumu wobec żądania prostych, bezrefleksyjnych procedur. Perspektywa bayesowska przypomina, że niepewność pozostaje nieusuwalnym elementem sytuacji i nie da się jej magicznie zlikwidować deklaracją, że „test ma dziewięćdziesiąt pięć procent dokładności".

Fetyszyzacja p-value: pułapki testowania hipotez

Tradycja testowania hipotez, ukształtowana przez Neymana i Pearsona, organizuje badanie wokół hipotezy zerowej. Traktuje się ją jako domyślne założenie o braku efektu, związku czy różnicy – przez analogię prawną jest to odpowiednik domniemania niewinności. Nie chodzi bynajmniej o to, że owa bierność świata jest ontologicznie bardziej prawdopodobna, lecz o to, że przyjmujemy ją jako punkt wyjścia, który można obalić jedynie za pomocą mocnych dowodów. To proceduralna konieczność, nie filozoficzne wyznanie wiary.

Źródłowy tekst słusznie piętnuje fetyszyzację wartości p, która zdominowała literaturę empiryczną, nie odpowiadając przy tym na pytanie, które naprawdę nas interesuje. Wartość p nie mówi bowiem, jakie jest prawdopodobieństwo hipotezy zerowej w świetle zebranych danych. Mówi coś zupełnie innego: jakie jest prawdopodobieństwo uzyskania danych tak skrajnych (lub bardziej) jak nasze, przy założeniu, że hipoteza zerowa jest prawdziwa. Ta subtelna inwersja logiczna, będąca w swej istocie zdaniem warunkowym, nie pozwala na bezpośredni wniosek o winie bądź niewinności oskarżonego.

Perspektywa bayesowska przywraca tu fundamentalną symetrię, pozwalając wprost mówić o stopniu poparcia dla hipotezy zerowej. Współczynnik Bayesa, czyli narzędzie służące do porównywania siły

dowodowej na rzecz konkurencyjnych hipotez, może przyjmować wartości mniejsze od jedności, co interpretujemy jako dowód przeciwko H_0. Może jednak również przyjmować wartości większe, stając się pozytywnym argumentem za brakiem efektu. Gdy uświadomimy sobie, że tradycyjny próg istotności p równej 0,05 odpowiada współczynnikowi Bayesa rzędu dwóch lub trzech – a zatem dowodom raczej anegdotycznym niż rozstrzygającym – napięcie między kulturą odkrywania „istotnych statystycznie wyników" a kulturą odpowiedzialnego wnioskowania staje się niemal namacalne.

Problem w tym, że instytucjonalna logika publikowania naukowego wzmocniła tę asymetrię na poziomie błędów. Błąd pierwszego rodzaju, czyli fałszywe odrzucenie prawdziwej hipotezy zerowej, traktowany jest jako dopuszczalne ryzyko zawodowe, byle utrzymane na poziomie pięciu procent. Tymczasem błąd drugiego rodzaju, a więc nieodrzucenie fałszywej hipotezy zerowej, pozostaje w cieniu, mimo że jego skutki społeczne bywają katastrofalne – od utrwalania nieskutecznych terapii po utrwalanie niewidocznych nierówności. Wszystkie te zabiegi, jak szacowanie mocy testu, projektowanie badań o odpowiedniej wielkości próby czy kontrola nad testowaniem wielokrotnym, są w istocie próbą przywrócenia zachwianej równowagi.

W kontekście społecznym ta techniczna z pozoru równowaga nabiera jednak wymiaru politycznego. Ustalenie progu istotności staje się de facto decyzją o tym, jaką część ryzyka błędnych interwencji lub zaniechań przerzucamy na obywateli, pacjentów czy pracowników. W tym sensie głośne postulaty zaostrzenia progu istotności można interpretować jako żądanie większej ostrożności wobec naukowych „konstatywów", które w niezwykle krótkim czasie potrafią zamienić się w polityczne „dyrektywy".

Przedziały niepewności precyzują komunikację ryzyka

Dla umysłu nawykłego do poszukiwania „najlepszej liczby" idea, że wokół niej rozpościera się cały wszechświat możliwych wartości, które równie dobrze mogliśmy zaobserwować, jest rewolucyjna. Aby ten wszechświat zwizualizować, statystyka sięga po metodę bootstrapingu, czyli technikę polegającą na wielokrotnym losowaniu z posiadanej próbki w celu symulacji niepewności. W tym kontekście przedział ufności staje się czymś więcej niż technicznym artefaktem; jest pedagogicznym narzędziem, które oddaje sprawiedliwość niepewności, pokazując, jak szeroki jest horyzont prawdopodobieństwa otaczający naszą pojedynczą estymację.

W dyskursie publicznym ta subtelność ginie, zastępowana przez poręczne metafory: margines błędu w sondażu, widełki dla prognozowanego bezrobocia, zakres skuteczności terapii. Kuszą one, by myśleć o przedziale jak o fizycznym pudełku, w którym ukryto prawdziwą wartość. Tymczasem jego natura jest proceduralna – przedział ufności jest elementem metody, która w nieskończonej serii powtórzeń

eksperymentu obejmowałaby prawdziwy parametr w określonym odsetku przypadków. Jest to więc relacja między naszym narzędziem a światem, a nie obiektywna rama narzucona na rzeczywistość.

Na scenę wkracza wówczas perspektywa bayesowska, oferując przedziały wiarygodności, które wprost wyrażają stopień naszego przekonania co do położenia parametru po uwzględnieniu danych. Jej język jest bliższy ludzkiej intuicji, bo odpowiada na pytanie, które faktycznie nas nurtuje: na ile powinniśmy wierzyć, że efekt leku jest bliski temu, co sugeruje punktowy wynik badania? Dla pacjenta i lekarza informacja, że redukcja ryzyka o dwadzieścia kilka procent jest otoczona wąskim przedziałem, który nie zahacza o zero, ma znacznie większą siłę perswazji niż abstrakcyjna wartość p rzędu jeden do miliona.

Ponownie jednak ujawnia się fundamentalne napięcie między modelem a żywym doświadczeniem. Przedział, czy to ufności, czy wiarygodności, konstruowany jest dla formalnie zdefiniowanej populacji, podczas gdy jednostka przeżywa jedno, konkretne życie, z unikalnym profilem ryzyka i w określonym systemie opieki. Jeśli społeczeństwo bezrefleksyjnie przyjmie język przedziałów jako ostateczny werdykt, dojdzie do cichego wywłaszczenia jednostki z prawa do własnej oceny sytuacji. Zadaniem krytycznego namysłu jest przywrócenie tej podmiotowości, nie poprzez odrzucenie statystyki, lecz przez wpisanie jej w szerszy dialog, w którym przedział jest ważnym argumentem, ale nigdy ostatecznym wyrokiem.

Modele regresji: statystyczny fundament polityki

Regresja, w swojej elementarnej postaci dopasowania prostej linii do chmury punktów, wydaje się niewinną techniczną operacją. Bierzemy zbiór par danych – wzrost matki i córki, dochód i konsumpcję, natężenie ruchu i liczbę wypadków – po czym sprowadzamy je do jednej, deterministycznej funkcji oraz tego, co nazywamy błędem. Ten pozornie neutralny krok ma jednak głębokie konsekwencje dla sposobu, w jaki społeczeństwo postrzega samo siebie: jako strukturę, w której każda jednostka jest tylko obserwacją wzdłuż linii trendu. Wszystko, co od tej linii odbiega, staje się „resztą”, statystycznym szumem, a w pewnym sensie – marnotrawstwem informacji.

Współczynnik regresji, teoretycznie będący tylko parametrem opisującym nachylenie owej linii, w praktyce szybko staje się normą. Gdy stwierdzamy, że z każdym dodatkowym centymetrem wzrostu rodzica oczekujemy jednej trzeciej centymetra wzrostu u dziecka, poruszamy się jeszcze w obrębie biologicznej regularności. Kiedy jednak mówimy, że każda nadgodzina podnosi ryzyko zawału, każda złotówka długu studenckiego zwiększa prawdopodobieństwo niewypłacalności, a każda kolejna operacja w szpitalu buduje doświadczenie zespołu i obniża śmiertelność, wkraczamy na grząski grunt. Relacja statystyczna jest tu natychmiast interpretowana jako ocena praktyk, instytucji i modeli biznesowych.

Z perspektywy krytycznej teorii społeczeństwa regresja okazuje się narzędziem o podwójnym ostrzu: może służyć zarówno demistyfikacji, jak i urzeczowieniu. Demistyfikuje, gdy obnaża, że gloryfikowane „talenty" uczniów są w dużej mierze produktem struktury klasowej, a rzekoma „naturalna skłonność do przedsiębiorczości" koreluje głównie z dostępem do kapitału. Prowadzi jednak do urzeczowienia, gdy tę samą zależność traktuje się jako niepodważalne prawo natury, usprawiedliwiające bierność. Skoro dzieci z biedniejszych domów i tak osiągną statystycznie gorsze wyniki, to po co nadmiernie inwestować w ich szkoły? Nachylenie linii staje się przecież „twardą daną".

David Spiegelhalter, przywołując badania nad kardiochirurgią dziecięcą, ilustruje to napięcie z niezwykłą klarownością. W latach dziewięćdziesiątych korelacja była wyraźna: szpitale wykonujące więcej zabiegów miały lepsze wyniki. Przełożyło się to na chwytliwy polityczny slogan – „koncentracja usług ratuje życie" – i posłużyło do delegitymizacji mniejszych ośrodków. Jednak po dwóch dekadach, w trakcie których wprowadzono rygorystyczne regulacje i standardy, korelacja niemal całkowicie zniknęła. Infrastruktura systemowa wyrównała szanse, a regresja przestała być prostym argumentem w politycznej walce. To moment, w którym statystyka przestaje być bronią w rękach interesu, a staje się narzędziem autokorekty całego systemu.

Jednocześnie opowieść o regresji musi być nieustannie chroniona przed pokusą utożsamiania korelacji z przyczynowością. Słynny przykład równoległego wzrostu spożycia mozzarelli i liczby doktoratów z inżynierii lądowej w USA to nie tylko zabawna anegdota. To fundamentalne ostrzeżenie przed strukturalnym błędem, jakim jest urzeczowienie upływającego czasu jako wspólnej przyczyny. Fakt, że dwie serie danych rosną równolegle, nie dowodzi, iż pizza generuje inżynierów. Pokazuje natomiast, że bez krytycznego namysłu nad czynnikami zakłócającymi jesteśmy skazani na bałwochwalstwo liczb.

Populacja docelowa vs próba: granice wnioskowania

Aby domknąć cykl PPDAC, musimy zmierzyć się z samym pojęciem populacji. Statystycy, na których powołuje się autor źródłowego tekstu, rozróżniają z myślową elegancją populacje literalne, wirtualne i metaforyczne. Populacja literalna to przypadek klasyczny: zbiór wszystkich jednostek, które formalnie mogą zostać włączone do badania, precyzyjnie określony w czasie i przestrzeni. Populacja wirtualna jest zbiorem wszystkich możliwych pomiarów, jakie moglibyśmy wykonać na tej samej jednostce, gdybyśmy powtarzali eksperyment w identycznych warunkach. Populacja metaforyczna to już konstrukcja czysto kontrfaktyczna: zbiór wszystkich możliwych historii świata, spośród których ta nasza, faktyczna, jest tylko jednym z wielu urzeczywistnień.

Ten trójpodział jest czymś więcej niż tylko zgrabnym narzędziem dydaktycznym; to klucz do zrozumienia, jak głęboko statystyka zakorzeniona jest w wyobraźni społecznej. Kiedy analizujemy wyniki sondaży

wyborczych, odwołujemy się do populacji literalnej – wszystkich uprawnionych do głosowania. Kiedy badamy rozkład błędów pomiaru ciśnienia krwi, korzystamy z populacji wirtualnej – kwantyfikowalnego marginesu wahań wokół „prawdziwej" wartości. A kiedy analizujemy liczbę zabójstw w danym roku, sięgając po rozkład Poissona, mimo że znamy wszystkie przypadki, zakładamy istnienie populacji metaforycznej: tych wszystkich innych lat, które mogły się zdarzyć, gdyby los potoczył się inaczej.

David Spiegelhalter sięga po tę metaforyczną kategorię szczególnie chętnie, gdy mówi o zdarzeniach rzadkich. Jeśli w Anglii i Walii dochodzi średnio do półtora zabójstwa dziennie, a rozkład tej liczby dobrze opisuje rozkład Poissona, można oszacować prawdopodobieństwo dnia z siedmioma lub więcej morderstwami. Okazuje się, że jest to zdarzenie skrajnie nieprawdopodobne, ale nie niemożliwe. Pojedynczy dzień z tak tragicznym bilansem nie musi więc być dowodem na załamanie się ładu społecznego, lecz może stanowić po prostu realizację jednego z nielicznych, lecz dopuszczalnych przez model scenariuszy.

Na poziomie polityki publicznej ta perspektywa ma fundamentalne znaczenie. Czy nagły wzrost liczby zabójstw w jednym roku traktujemy jako sygnał strukturalnego kryzysu, czy jako fluktuację w ramach stabilnego rozkładu? Czy anomalny wzrost liczby zamachów terrorystycznych to dowód na „nową erę", czy raczej wynik kilku zdarzeń z „długiego ogona" rozkładu rzadkich zjawisk? Czy wyjątkowo łagodny sezon grypowy interpretujemy jako trwały sukces służby zdrowia, czy jako statystyczny kaprys? Odpowiedzi na te pytania zależą od tego, czy potrafimy świadomie operować kategorią populacji metaforycznej, zamiast ulegać panice moralnej lub triumfalizmowi na podstawie pojedynczych obserwacji.

W epoce sztucznej inteligencji pojawia się tu jednak zupełnie nowy wymiar. Modele generatywne, algorytmy predykcyjne i systemy rekomendacyjne wpisują wyobraźnię populacji metaforycznej bezpośrednio w infrastrukturę naszych decyzji. Populacje, które kiedyś istniały tylko w umyśle statystyka, stają się wirtualnymi poligonami, po których algorytmy błądzą, optymalizując swoje parametry na milionach symulowanych scenariuszy. Otwiera to drogę do radykalnie nowej formy zarządzania: zamiast reagować na faktyczne dane, systemy mogą działać w oparciu o rozkłady potencjalnych zdarzeń, profilując polisy ubezpieczeniowe, oferty kredytowe czy poziom nadzoru policyjnego na podstawie setek tysięcy symulowanych historii, których nikt nigdy nie przeżył.

Od korelacji do przyczynowości: wyzwanie nauk społecznych

Przejście od regresji do przyczynowości jest krokiem, który przekracza pewną epistemologiczną przepaść – jest to ruch tyleż techniczny, co filozoficzny. Techniczny, ponieważ wymaga sięgnięcia po zaawansowane instrumentarium: randomizowane badania kontrolowane, analizę zmiennych instrumentalnych, metody

dopasowania czy modelowanie kontrfaktyczne. Filozoficzny, albowiem zmusza do porzucenia logiki biernego obserwatora świata na rzecz logiki aktywnej interwencji, która stawia fundamentalne pytanie: „jeśli zmienimy X, to jak dokładnie zmieni się rozkład Y?”.

Źródłowy tekst słusznie zauważa, że statystyczne ujęcie przyczynowości ma charakter probabilistyczny, a nie deterministyczny. Stwierdzenie „palenie powoduje raka” nie oznacza przecież, że każdy palacz nieuchronnie zachoruje. Znaczy ono coś znacznie subtelniejszego: w populacji, w której eksperymentalnie wymusilibyśmy palenie, odsetek zachorowań byłby istotnie wyższy niż w bliźniaczej populacji, której byśmy go zabronili. Randomizowane badania kliniczne w medycynie, infrastruktura prób losowych w ekonomii rozwoju czy eksperymenty naturalne w polityce społecznej są jedynie różnymi wcieleniami tej samej idei. Przyczynowość jest tym, co pozostaje jako sygnał, gdy wyeliminujemy tyle czynników zakłócających, ile tylko jest to metodycznie możliwe.

Spiegelhalter przywołuje w tym kontekście wstrząsający przykład Harolda Shipmana, aby pokazać, że przyczynowość można rekonstruować również ex post, o ile dysponujemy odpowiednio zaprojektowanymi procedurami alarmowymi. Gdyby w latach osiemdziesiątych zgony pacjentów lekarzy rodzinnych monitorowano przy użyciu testów sekwencyjnych, czyli procedury statystycznej zaprojektowanej do ciągłego monitorowania danych i wczesnego wykrywania anomalii, można byłoby zidentyfikować patologiczny wzorzec w praktyce Shipmana znacznie wcześniej, zapobiegając setkom zgonów. Przyczynowość jest tu rozumiana jako dramatyczna różnica między światem, w którym interweniujemy szybko, a światem, w którym pozostajemy bierni. Test sekwencyjny staje się narzędziem redukującym czas naszej ignorancji, przy zachowaniu akceptowalnie niskiego ryzyka fałszywego alarmu wobec tysięcy niewinnych lekarzy.

W szerszym ujęciu współczesna debata o przyczynowości toczy się między kilkoma intelektualnymi obozami. Obóz „randomizacyjny”, wyrastający z tradycji prób losowych i szkoły Neymana, utrzymuje, że autentyczną przyczynowość można ustalić wyłącznie tam, gdzie możliwa jest kontrolowana interwencja. Obóz „struktur przyczynowych”, którego czołowym przedstawicielem jest Judea Pearl, buduje formalne modele graficzne, w których przyczynowość jest własnością samej architektury zależności, a nie tylko planu eksperymentu. Wreszcie obóz „epidemiologiczny”, odwołujący się do kryteriów Bradforda Hilla, kładzie nacisk na syntetyzowanie dowodów z różnych poziomów – biologicznych, klinicznych i populacyjnych – aby uzasadnić związki przyczynowe tam, gdzie randomizacja jest niemożliwa lub nieetyczna.

Spiegelhalter, jako pragmatyczny, bayesowsko zorientowany empirysta, proponuje podejście, które można by nazwać proceduralnym minimalizmem. Przyczynowość jest dla niego tym, co pozostaje po tym, jak zrobiliśmy wszystko, co rozsądne, aby wyeliminować alternatywne wyjaśnienia. Nie potrzebujemy do tego wielkiej, metafizycznej teorii przyczyn. Potrzebujemy natomiast praktycznych kryteriów, które pozwolą

lekarzowi, sędziemu czy decydentowi gospodarczemu uzasadnić swoją interwencję w sposób zrozumiały i możliwy do obrony.

Czarne skrzynki: etyka algorytmów predykcyjnych

W kolejnym etapie cyklu PPDAC świat klasycznej statystyki zderza się z potęgą uczenia maszynowego. Na terytorium, które Spiegelhalter określa mianem „algorytmów, analityki i predykcji", wkracza cały arsenał nowych narzędzi: algorytmy klasyfikacyjne, krzywe ROC, miary AUC, wskaźniki Briera, drzewa decyzyjne, lasy losowe i coraz bardziej egzotyczne warianty sieci neuronowych. To moment, w którym statystyczna refleksja musi zmierzyć się z inżynierską skutecznością.

W sensie formalnym rewolucji tu nie ma. Wciąż chodzi o to samo: o przekształcanie danych w przewidywania, które następnie oceniamy pod kątem kalibracji, zdolności dyskryminacyjnej i stabilności. Wskaźnik Briera, który traktuje prognozy probabilistyczne jak obarczone błędem kwadratowym zakłady o wynik zero-jedynkowy, pozostaje w istocie miarą jakości w tej samej „bayesowskiej grze w zgadywanie". Dobrze skalibrowany algorytm to taki, który – jeśli tysiąc razy przewiduje 70% szans na deszcz – odnotowuje opady w około siedmiuset przypadkach. Testy kalibracji Spiegelhaltera są więc próbą utrzymania algorytmów w ryzach epistemicznej odpowiedzialności.

Prawdziwy problem rodzi się jednak, gdy logika optymalizacji predykcyjnej zostaje sprzężona z logiką korporacji i państwa. Klasyczne modele, takie jak regresja logistyczna, zachowują względną przejrzystość; wciąż można wyjaśnić obywatelowi, że odmówiono mu kredytu, ponieważ jego dochód i historia zatrudnienia w określony sposób podnoszą prawdopodobieństwo niewypłacalności. Tymczasem złożone sieci neuronowe czy potężne lasy losowe, choć często skuteczniejsze, stają się czarnymi skrzynkami, których wewnętrzna logika pozostaje nieprzenikniona nawet dla ich twórców. Skuteczność kupiono za cenę przejrzystości.

W odpowiedzi na to wyzwanie narasta świadomość, że w sytuacjach o wysokiej stawce społecznej nie można oddzielić precyzji od obowiązku wyjaśnienia. Coraz większą rolę odgrywają więc metody z nurtu „explainable AI", które próbują niczym detektywi rekonstruować lokalny wkład poszczególnych cech w ostateczną decyzję modelu. Czasem prowadzi to do selekcji cech, a czasem do świadomego powrotu do prostszych, bardziej interpretowalnych modeli, o ile nie wiąże się to z drastyczną utratą jakości predykcji.

USA, Europa i świat arabski: geopolityka danych i AI

W skali globalnej widać wyraźne pęknięcia cywilizacyjne w podejściu do algorytmicznych systemów decyzyjnych. W Stanach Zjednoczonych dominuje kultura zorientowana na efektywność, gdzie przewaga nawet o kilka punktów procentowych w precyzji predykcji niewypłacalności, recydywy czy ryzyka zdrowotnego staje się wystarczającym argumentem za wdrożeniem czarnej skrzynki. Dzieje się tak, o ile tylko jej działanie nie narusza w sposób zbyt jawny formalnych zakazów dyskryminacji. Europa, zakorzeniona w tradycji praw człowieka i regulacjach takich jak RODO, kładzie znacznie większy nacisk na przejrzystość, prawo do wyjaśnienia oraz ocenę wpływu technologii na prawa podstawowe. Z kolei w świecie arabskim obserwujemy fascynujące napięcie między ambicjami stania się liderem w dziedzinie AI, widocznymi w strategiach narodowych Arabii Saudyjskiej czy ZEA, a świadomością barier, które czynią te wdrożenia selektywnymi. Algorytmy trafiają tam często w ręce państwa, wspierając predykcję zachowań zbiorowych i zarządzanie opinią publiczną.

Ta regionalna dywersyfikacja okazuje się jednak powierzchowna, gdy zajrzymy pod maskę. W każdym z tych kontekstów rdzeniem maszyn decyzyjnych, działających w tle naszej codzienności – od rekomendacji treści, przez scoring kredytowy, po nadzór epidemiologiczny – stają się statystyka i uczenie bayesowskie. Paradoks polega na tym, że im więcej wysiłku wkładamy w to, by maszyny uczyły się z danych, tym bardziej rośnie ryzyko, że my, jako społeczeństwo, z tego trudu zrezygnujemy. Wezwanie Davida Spiegelhaltera, byśmy nauczyli się „powolnego myślenia o naszej niewiedzy”, stoi w fundamentalnym konflikcie z kulturą, która wynosi predykcję na piedestał, traktując ją jako wygodny substytut dla żmudnego procesu rozumienia.

Statystyka w sądzie: matematyczna weryfikacja dowodów

Prawnicy, którzy spoglądają na statystykę z nieufnością, mają po swojej stronie długą i ponurą historię sądowych pomyłek. Procesy, w których źle zrozumiane prawdopodobieństwo prowadziło do niesprawiedliwych wyroków, nie są anomalią. Ten sam błąd, który myli prawdopodobieństwo obciążającego wyniku testu z prawdopodobieństwem samej winy – zilustrowany wcześniej na przykładzie mammografii i testów antydopingowych – powraca niczym fatum w kolejnych odsłonach jako błąd prokuratora. Problem nie leży jednak w abstrakcyjności matematyki, lecz w tym, że praktyka dowodowa zbyt często odmawia sobie wysiłku, by włączyć formalne narzędzia do swoich procedur. Woli schronić się za mitem intuicyjnej mądrości ławy przysięgłych.

Gdybyśmy jednak potraktowali poważnie analogię między hipotezą zerową a domniemaniem niewinności, wówczas współczynnik wiarygodności i współczynnik Bayesa stałyby się naturalnymi kandydatami na miarę siły dowodu. Współczynnik wiarygodności, czyli narzędzie porównujące prawdopodobieństwo zaobserwowania danego śladu przy założeniu winy oraz niewinności, jest w istocie precyzyjną wersją pytania, które prawo powinno zadawać od zawsze: jak bardzo ten konkretny dowód pasuje do jednej, a jak do drugiej narracji? W połączeniu z jawnie określonymi szansami a priori pozwala to na spójną aktualizację przekonań, a nie na chaotyczne akumulowanie subiektywnych wrażeń.

David Spiegelhalter, opisując głośną sprawę identyfikacji szczątków Ryszarda III, pokazuje, jak wiele zyskała nauka sądowa na przyjęciu tej perspektywy. Zamiast mistycznego zaklęcia, że „prawdopodobieństwo pomyłki jest znikome", skrupulatnie zrekonstruowano i przemnożono współczynniki wiarygodności dla każdej z niezależnych przesłanek: miejsca znalezienia szkieletu, jego wieku, płci, deformacji kręgosłupa, charakteru obrażeń i wreszcie wyników badań genetycznych. Szanse a priori, że to król, wynosiły jeden do czterystu; szanse a posteriori sięgnęły rzędu kilkunastu tysięcy do jednego. Ta zmiana o charakterze geometrycznym nie jest tanią sztuczką retoryczną, lecz owocem konsekwentnej arytmetyki, która uczciwie włącza każdy element dowodu, zamiast opierać się na ulotnej impresji.

Z tego punktu widzenia prawo, które w imię ochrony zdrowego rozsądku odrzuca bayesowską interpretację dowodów, rezygnuje z potężnego narzędzia do porządkowania niewiedzy. Oczywiście, aparatu tego nie da się stosować mechanicznie w każdej sytuacji. W wielu sprawach cywilnych i karnych dane są zbyt ubogie, a hipotezy zbyt niedookreślone, aby przypisywanie im dokładnych liczb miało sens. Jednak nawet na poziomie czysto jakościowym logika Bayesa zmusza do zadawania właściwych pytań: co właściwie uznajemy za szanse a priori, które przesłanki wzmacniają którą opowieść i jak bardzo jeden dowód zmienia cały bilans.

W świecie arabskim, gdzie systemy prawne często splatają tradycję religijną z nowoczesnymi kodeksami, integracja statystyki bywa jeszcze trudniejsza, choć w pewnych obszarach, jak finanse czy arbitraż międzynarodowy, staje się nieunikniona. W Stanach Zjednoczonych debata koncentruje się na standardach dopuszczalności dowodów naukowych. W Europie zaś, gdzie dominuje model sędziego aktywnie prowadzącego postępowanie, potencjał bayesowskiej kultury wydaje się szczególnie obiecujący, gdyż mógłby zredukować arbitralność oceny, nie naruszając przy tym sędziowskiej swobody.

Warunkiem brzegowym jest jednak świadome kształcenie statystyczne prawników, sędziów i prokuratorów. Bez tego każda próba wprowadzenia liczb na salę sądową nieuchronnie skończy się ich fetyszyzacją albo całkowitym odrzuceniem.

Instytucje budują kulturę uczciwości statystycznej

Jeżeli potraktować poważnie ambicję, by cykl PPDAC stał się normatywną matrycą nowoczesnego rozumu publicznego, a nie tylko technicznym narzędziem, natychmiast rodzi się pytanie o instytucjonalne warunki jego uczciwego stosowania. W tradycji myślenia, do której należy Spiegelhalter, statystyka nigdy nie jest bytem abstrakcyjnym; jest zawsze osadzona w gęstej sieci bodźców, nagród i sankcji, która ostatecznie decyduje, jakie pytania w ogóle wolno zadać, jakie dane warto zebrać i które wyniki zasługują na nagłośnienie.

W obszarze nauk empirycznych ten ekosystem tworzą redakcje czasopism, agencje grantowe, komisje etyczne i koncerny farmaceutyczne, kształtując wspólną preferencję dla wyników istotnych statystycznie, spektakularnych i potencjalnie wdrażalnych. W ekonomii i finansach podobną rolę odgrywa logika rynków, ratingów analitycznych i kwartalnych raportów. W polityce publicznej analogicznym mechanizmem stają się sondaże, presja opinii publicznej i nieprzekraczalny horyzont wyborczy. W każdy z tych systemów wpisana jest strukturalna premia za szybkie odpowiedzi przedkładane nad powolne uczenie się, za proste wskaźniki przedkładane nad złożony opis niepewności oraz za gest decyzyjny przedkładany nad metodyczną powściągliwość.

Aby PPDAC mógł dotrzymać swojej obietnicy, każda z tych sfer wymagałaby przebudowy w kierunku instytucji, które nagradzają raczej przejrzyste komunikowanie niepewności niż teatralne ogłaszanie odkryć. Spiegelhalter powtarza, że uczciwa statystyka zaczyna się tam, gdzie badacz otwarcie przyznaje się do niewiedzy i traktuje ją jako poznawczy zasób, a nie wstydliwą porażkę. W tym sensie jego myśl nawiązuje do słynnej maksymy George’a Boxa, że wszystkie modele są błędne, ale niektóre bywają użyteczne. Ich użyteczność nie polega jednak na zastąpieniu świata, lecz na umożliwieniu nam bardziej rozumnego działania w jego obrębie.

Regresja do średniej: pułapka interpretacyjna w polityce

Pozwolę sobie na krótką, z pozoru absurdalną analogię, która jednak doskonale testuje, czy przyswoiliśmy już aparat pojęciowy statystyki. Wyobraźmy sobie ministra tłumaczącego wyborczą klęskę rządu: „Nasze wyniki są w regresji do średniej, a zatem za cztery lata, mocą praw statystyki, w sposób naturalny powrócimy do władzy”. Gdyby jakikolwiek obywatel uznał tę wypowiedź za sensowną, popełniłby podręcznikowy błąd. Regresja do średniej nie jest bowiem magiczną siłą dziejów, która kompensuje nieudolność, lecz wyłącznie statystycznym opisem faktu, że ekstremalny wynik jest po części dziełem losowych fluktuacji.

Absurd w wypowiedzi ministra polega na nieuprawnionym przeniesieniu prawa statystycznego z poziomu opisu rozkładu danych na poziom celowości historii. Jednak analogiczne błędy, choć w mniej spektakularnej formie, popełniane są nieustannie. Kiedy firma tłumaczy gwałtowny spadek sprzedaży tym, że „za rok statystyka to wyrówna", gdy program społeczny ogłasza sukces na podstawie kilku lat sprzyjających trendów, gdy uczelnia wyciąga wnioski o geniuszu studentów na podstawie jednego rocznika – wszędzie tam regresja do średniej staje się wygodną zasłoną, która maskuje potrzebę rzetelnej analizy przyczynowej i strukturalnej.

Humor tej figury retorycznej ma charakter dydaktyczny: pokazuje, że statystyka nigdy nie zastąpi odpowiedzialności politycznej i moralnej. Regresja do średniej nie jest przeznaczeniem, twierdzenie Bayesa nie jest wyrocznią, a przedział ufności nie stanowi tarczy przed krytyką. Jeśli w praktyce społecznej statystyka staje się językiem wymówek, a nie językiem rozliczeń, oznacza to, że porzuciliśmy jej radykalny potencjał jako narzędzia samokrytyki całego systemu.

Edukacja i media: oswajanie publicznej niepewności

Aby w pełni zrozumieć stawkę tej gry, warto naszkicować trzy profetyczne, lecz mocno zakotwiczone w realnych trendach scenariusze przyszłości. Każdy z nich opisuje odmienną relację między statystyką, sztuczną inteligencją a życiem społecznym.

W pierwszym scenariuszu – technokratycznej konsolidacji – algorytmy predykcyjne stają się głównym medium, przez które instytucje władzy, rynki i platformy cyfrowe postrzegają całe populacje. Cykl PPDAC ulega tu radykalnej redukcji do zaledwie trzech kroków: Problem jest narzucany przez interesy dominujących graczy, Dane zbiera się bezrefleksyjnie z każdego cyfrowego śladu, a Analizę deleguje się do czarnych skrzynek optymalizujących krótkoterminowe cele. Fazę Wniosków zastępuje automatyczna decyzja systemu. Bayesowskie uczenie się zachodzi w maszynach, podczas gdy ludzie pozostają w stanie poznawczej zależności, przyjmując predykcje niczym wyroki losu. W tym świecie niepewność zostaje oficjalnie wyparta, lecz powraca pod postacią lęku i teorii spiskowych, ponieważ nikt już nie rozumie, jak i dlaczego podejmowane są kluczowe decyzje.

Drugi scenariusz to reakcyjne kontroświecenie, w którym rodzi się ruch odrzucenia liczb jako narzędzia rzekomo odczłowieczającego. Statystykę i AI utożsamia się z opresją, a w ich miejsce proponuje się czyste doświadczenie, tradycję, narrację i intuicję. W skrajnej wersji tego scenariusza władze porzucają oparte na dowodach polityki zdrowotne, klimatyczne czy gospodarcze. Społeczeństwa wchodzą w epokę nowego irracjonalizmu technologicznego, gdzie urządzenia są powszechnie używane, ale ich działanie traktuje się jak magię, a nie przedmiot krytycznej refleksji. Brak kultury statystycznej nie usuwa niepewności – czyni ją jedynie nieprzejrzystą i podatną na manipulacje charyzmatycznych liderów.

Trzeci scenariusz, który można nazwać deliberatywnym oświeceniem, zakłada, że statystyka i AI zostają wbudowane w instytucje publicznego rozumu. Cykl PPDAC staje się nie tylko metodą naukową, lecz również schematem projektowania polityk publicznych. Problem jest tu współtworzony z obywatelami, Plan musi spełniać kryteria sprawiedliwości, Dane są traktowane jako zasób wspólny, Analiza jest prowadzona transparentnie, a Wnioski stają się punktem wyjścia do publicznej debaty, a nie jej końcem. Uczenie bayesowskie zostaje rozszerzone z poziomu algorytmów na poziom instytucjonalny, gdzie priorytety społeczne są świadomie aktualizowane w świetle nowych dowodów, a nie tylko wyników wyborów.

W tym ujęciu AI staje się narzędziem do symulowania możliwych przyszłości, lecz ostateczna decyzja zawsze należy do uczestników debaty. Zachowują oni świadomość, że każda predykcja jest tylko projekcją na tle naszej niewiedzy. Odróżnianie korelacji od przyczynowości, rozumienie błędów pierwszego i drugiego rodzaju, świadome posługiwanie się przedziałami niepewności czy zdyscyplinowane testowanie hipotez – wszystko to staje się obywatelskim elementarzem, a nie tajemną wiedzą wąskiej elity.

> Liczby, choć potężne, są jedynie echem naszych własnych pytań i wyborów. Czy w świecie zdominowanym przez algorytmy zdołamy zachować zdolność do „powolnego myślenia o naszej niewiedzy”? A może staniemy się ofiarami własnych, uprzedmiotowionych projekcji, skazani na życie w świecie, w którym statystyka przestała być narzędziem poznania, a stała się językiem wymówek?

Inspiracje

[1] **"The Art. od Statistics. Learning from data"** David Spiegelhalter

2020 | Pelican Books

Sir **David Spiegelhalter** jest emerytowanym profesorem statystyki na Uniwersytecie Cambridge. Znany jest z prac nad wnioskowaniem bayesowskim, komunikacją ryzyka i społecznym rozumieniem statystyki. W latach 2017-2018 był prezesem Królewskiego Towarzystwa Statystycznego.

*Sir **David Spiegelhalter** is Emeritus Professor of Statistics at the University of Cambridge. He is known for his work on Bayesian inference, risk communication, and public understanding of statistics. He was President of the Royal Statistical Society (2017-2018).*

University of Cambridge

Literatura uzupełniająca

Książki

Autorzy przywoływani w artykule:

[1] **"The Art of Uncertainty: How to Navigate Chance, Ignorance, Risk and Luck"**

David Spiegelhalter

2024 | Pelican Books

Google Scholar

[2] **"Covid by Numbers: Making Sense of the Pandemic with Data"**

David Spiegelhalter, Anthony Masters

2021 | Pelican

[3] **"Sex by Numbers: What Statistics Can Tell Us About Sexual Behaviour"**

David Spiegelhalter

2015 | Wellcome Collection

Literatura pokrewna (Google Scholar):

[4] **"Lectures and conferences on mathematical statistics and probability"**

Jerzy Neyman

1938 | Washington, D.C.

[5] **"Karl Pearson : an appreciation of some aspects of his life and work"**

Egon Pearson

1938 | Cambridge University Press

[6] **"Studies in the history of statistics and probability"**

Egon Pearson, Maurice George Kendall

1969

[7] **"Causality: Models, Reasoning, and Inference"**

Judea Pearl

2000 | Cambridge University Press

Google Scholar

[8] **"The Book of Why: The New Science of Cause and Effect"**

Judea Pearl, Dana Mackenzie

2018 | Basic Books

Google Scholar

[9] **"Causal Inference in Statistics: A Primer"**

Judea Pearl, Madelyn Glymour, Nicholas P. Jewell

2016

[10] **"Causal Inference: The Mixtape"**

Scott Cunningham

2021

[11] **"Causal Inference: What If"**

Miguel A. Hernán, James M. Robins

2020 | Chapman and Hall/CRC

Google Scholar

Artykuły naukowe

Autorzy przywoływani:

[1] **"Causal inference in statistics: An overview"**

Judea Pearl

2009 | Statistics Surveys

Google Scholar

[2] **"Robustness in the strategy of scientific model building"**

George E.P. Box

1979 | Robustness in Statistics

[3] **"Bayesian Methods in Machine Learning Applications and Challenges"**

Chengyue Li

2025

Google Scholar

[4] **"An Essay Towards Solving a Problem in the Doctrine of Chances"**

Thomas Bayes

1763 | Philosophical Transactions of the Royal Society

Google Scholar

[5] **"A Review of Bayesian Machine Learning Principles, Methods, and Applications"**

Kajal Kiran Shende, Dr. Siddu P Algur

2023 | International Journal of Innovative Research in Science Engineering and Technology

Google Scholar

[6] **"On the Problem of the Most Efficient Tests of Statistical Hypotheses"**

Jerzy Neyman, Egon S. Pearson

1933 | Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character

[7] **"Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability"**

Jerzy Neyman

1937 | Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences

[8] **"Causal Machine Learning: A Survey and Open Problems"**

Jean Kaddour, Aengus Lynch, Qi Liu, Matt J. Kusner, Ricardo Silva

2022 | arXiv

[9] **"On the Use and Interpretation of certain Test Criteria for the Purposes of Statistical Inference"**

Jerzy Neyman, Egon Pearson

1928 | Biometrika

[10] **"P Value and the Theory of Hypothesis Testing: An Explanation for New Researchers"**

Mohammad Zamanzadeh, Ali Ghaneei, Tahereh Rezaei, Davood Izadi, Faroogh Maroufi, Reza Rahmani, Leili Tapak

2020 | Nephro-Urology Monthly

Google Scholar

[11] **"Causal diagrams for empirical research"**

Judea Pearl

1995 | Biometrika

Google Scholar

[12] **"Science and Statistics"**

George E. P. Box

1976 | Journal of the American Statistical Association

[13] **"Causal Inference in Data Analysis with Applications to Fairness and Explanations"**

Authors not specified in the source

2025

Artykuły pokrewne (Google Scholar):

[14] "Why probability probably doesn't exist (but it is useful to act like it does)"

D Spiegelhalter

2024 | Nature

Google Scholar

[15] "How to maintain trustworthiness when doctors act as policy advocates"

Sebastian Walsh, David Taylor-Robinson, David Spiegelhalter, Carol Brayne

2024 | The BMJ

[16] "The effects of communicating scientific uncertainty on trust and decision making in a public health context"

Claudia R Schneider, Alexandra L. J. Freeman, David Spiegelhalter, Sander van der Linden

2023 | Judgment and Decision Making

APA ONE Publishing System

Fundacja Dobre Państwo © 2026

Article ID: 24f93199-8a9e-42b6-bfac-575747a5498d