



**Zalán
Gyenis**

Adiunkt w Zakładzie Logiki w Instytucie Filozofii Uniwersytetu Jagiellońskiego. Jego zainteresowania naukowe dotyczą logiki i epistemologii formalnej, w szczególności zagadnień dotyczących interpretacji prawdopodobieństwa i przyczynowości probabilistycznej.

Czy piraci powodują globalne ocieplenie?

Budzący Niepokój
Biznesmen radzi się
swojego statystyka.

BNB: Często muszę
latać, a bardzo
obawiam się
bomb. Jakie jest
prawdopodobieństwo,
że pasażer wniesie
bombę na pokład?

S: Jak jeden do
pięciu tysięcy.

BNB: Och, to zbyt
wiele. Może mógłbym
coś z tym zrobić?

S: Cóż, zawsze miej
ze sobą w podróży
bombę, ponieważ
prawdopodobieństwo
znajdowania się dwóch
bomb na pokładzie
jest o wiele mniejsze.

Słowa kluczowe: korelacja,
przyczynowość, współczynnik
Pearsona, globalne
ocieplenie, statystyka,
prawdopodobieństwo

Nadużywanie pojęć

Ten żart bazuje na pewnym szczególnym nadużyciu pojęcia prawdopodobieństwa i statystyki. W moim tekście zajmę się powiązanymi, często źle rozumianymi pojęciami: korelacji i przyczynowości. Każdy z nas słyszał kiedyś tę mantrę: „korelacja nie implikuje przyczynowości”, ale właściwie, co to w ogóle znaczy? Posługując się garścią przykładów, pokażę podstawowe różnice między korelacją i przyczynowością oraz kiedy te terminy są nie tylko źle używane, ale również traktowane zamiennie.

Tytuł tego tekstu odnosi się do zabawnej obserwacji, że od kiedy liczba piratów zaczęła spadać (od roku 1800), globalna temperatura zaczęła rosnąć. Istnieje zatem istotna statystycznie relacja odwrotności między liczbą piratów a globalną temperaturą. Czy więc to spadająca liczba piratów powoduje globalne ocieplenie? Jeśli tak, to dlaczego nie zostać jednym z nich? Spróbujmy jednak przybliżyć, czym jest korelacja i przyczynowość, zostawiając żarty na boku.

W tym najprostszym przykładzie mamy dwie wielkości (zmienne) P i T (P – piraci, T – temperatura), które zmieniają się w czasie t . $P(t)$ odnosi się do liczby piratów, zaś $T(t)$ oznacza średnią temperaturę w czasie t . Na podstawie obserwacji możemy stwierdzić, że wraz z upływem czasu P się zmniejsza, a T rośnie. Stwierdzamy zatem, że P i T są ze sobą negatywnie skorelowane. Pozytywna korelacja dwóch zmiennych X i Y może być rozumiana w ten sposób, że zarówno X i Y zwiększają się lub zmniejszają wraz z upływem czasu, zaś w przypadku korelacji negatywnej kierunki spadku lub wzrostu obu zmiennych są przeciwnie.

Korelacja jest miarą tego, do jakiego stopnia pary zmiennych są liniowo ze sobą powiązane. Ci, których interesuje matematyczny punkt widzenia, powinni poznać współczynnik korelacji Pearsona. Współczynnik korelacji wynosi $+1$, gdy zachodzi doskonale pozytywna liniowa korelacja, zaś -1 , gdy zachodzi doskonale negatywna liniowa korelacja (nazywana często antykorelacją), w innych przypadkach osiąga jakąś wartość pomiędzy $+1$ i -1 . Im bardziej korelacja zbliża się do zera, tym w mniejszym stopniu zachodzi relacja pomiędzy zmiennymi; im bliżej zaś znajduje się $+1$ lub -1 , to tym mocniej są one ze sobą powiązane. Mówimy, że korelacja jest znacząca, jeśli jest powyżej pewnego (ustalonego wcześniej) progu, na przykład jeżeli jest większa niż 0.95 (lub mniejsza niż -0.95). Gdy korelacja wynosi 0 , to mówimy, że dwie zmienne są niezależne. Korelacja jest po prostu liczbą, która mierzy liniową zależność, ale nie mówi nam niczego o przyczynach stojących za relacją tych zmiennych, mówi jedynie, że relacja istnieje.

Intuicje a przyczynowość

Definiowanie przyczynowości stanowiło dla filozofii (metafizyki) i nauki (fizyki, psychologii, prawa etc.) twardy orzech do zgryzienia już w czasach Arystotelesa. To zagadnienie zaprowadziłoby nas jednak zbyt daleko, dlatego skupimy się na zdroworozsądkowych intuicjach. Za dobrą in-

tuicję należy uznać tę, która stwierdza, że przyczynowość wskazuje, czy jedno zdarzenie jest skutkiem zaistnienia innego zdarzenia. Trzymając się formalizmu w kwestii przyczynowości, można by stwierdzić, że zmiana wartości jednej zmiennej pociągnie za sobą zmianę wartości innej zmiennej.

Typowy przykład byłby następujący: naciśnięcie przełącznika jest przyczyną tego, że żarówki zaczynają się świecić. Rozpatrując tę sytuację dokładnie, możemy zauważyć, że za każdym razem, gdy uruchamiamy przełącznik, światło się zapala. Innymi słowy, dwie zmienne, które wskazują na pozycję przełącznika (0 = włączenie lub 1 = wyłączenie) i to, że żarówki się

świecą (0 = światło 1 = brak światła), są te same, a więc stanowią przykład doskonałej pozytywnej korelacji. A jeśli to nie korelacja? Jeśli to coś innego? Z pewnością to jest korelacja, ale jednocześnie coś więcej.

Posiadamy przecież teorię, która wyjaśnia, w jaki sposób działa mechanizm przełączników, kabli ukrytych w ścianie, żarówek, elektryczności, propagacji prądu etc. Nie dokonujemy więc jedynie zestawienia dwóch skorelowanych ze sobą zbiorów danych, na podstawie których stwierdzamy relację przyczynową. Znalezienie przyczynowości wymaga odwołania się do danych eksperymentalnych i teorii, a nie tylko do danych obserwacyjnych.

Czy korelacja implikuje przyczynowość? Zauważmy, że korelacja jest relacją symetryczną, zaś przyczynowość asymetryczną. Stwierdzić, że brak piratów jest powiązany ze wzrostem globalnej temperatury to to samo, co twierdzić, że wzrost temperatury jest powiązany z brakiem piratów. Jednak to, że brak piratów powoduje globalne ocieplenie, jest zupełnie innym zdaniem niż stwierdzenie, że globalne ocieplenie jest przyczyną wyginięcia piratów.

Dlatego pytając o to, czy korelacja implikuje przyczynowość, musimy przynajmniej zaznaczyć, o jaki kierunek przyczynowości nam chodzi. Oczywiście może być tak, że piraci nie powodują globalnego ocieplenia, ►

Paweł i Gaweł są w relacji symetrycznej, jeśli Paweł jest bratem Gawła, a Gaweł bratem Pawła.

Platon i Arystoteles są w relacji asymetrycznej, jeśli Platon jest nauczycielem Arystotelesa, ale Arystoteles nie jest nauczycielem Platona.



Ilustracja: Małgorzata Ugiłk

Warto doczytać
 ■ P. Kawalec, *Przyczyna i wyjaśnianie. Studium z filozofii i metodologii nauk*, Lublin 2006.
 ■ A. Grobler, *Metodologia nauk*, Kraków 2006.

ale globalne ocieplenie jest przyczyną śmierci piratów. Zostawmy jednak ten zabawny przykład i zajmijmy się takim, który brzmi bardziej wiarygodnie.

Im więcej strażaków walczy z pożarem, tym większe są zniszczenia.

A więc liczba strażaków jest skorelowana ze skalą zniszczeń. Choć byłoby ciekawe przyjąć, że to strażacy rozmyślnie dokonali demolki, to założmy jednak, że strażacy wysyłani do gaszenia pożaru sami nie powodują zniszczeń. Co z kierunkiem odwrotnym? Wydaje się, że można się zgodzić na to, że im więcej jest zniszczeń wywołanych przez ogień, tym więcej strażaków musi przybyć na miejsce pożaru. To przykład sytuacji, w której odwrotna relacja przyczynowa wyjaśnia korelację.

Okulary i telewizja

Podobne wyjaśnienie może występować w przypadku dzieci, które są nauczone i które otrzymują złe oceny. Z pewnością to nie fakt uczenia dzieci sprawia, że otrzymują złe oceny, ale coś zupełnie innego. Mówi się również, że dzieci, które oglądają za dużo telewizji, są najbardziej agresywne. Być może to telewizja sprawia, że dzieci są bardziej agresywne, ale równie dobrze może być tak, że po prostu to agresywne dzieci oglądają dużo telewizji.

Istnieje powiązanie między liczbą sprzedanych lodów i okularów przeciwsłonecznych: jeśli sprzedaż lodów się zwiększa, to samo dzieje się ze sprzedażą okularów.

Oczywiście spożywanie lodów nie wymusza na ludziach kupowania okularów przeciwsłonecznych i jest równie mało prawdopodobne, że kupno takich okularów sprawi, że będziemy jeść lody. Nie ma tu bezpośredniego ani (zwróć uwagę) przyczynowania. Co jednak wyjaśnia tę korelację? Być może jest to coś, co nazywane jest wspólną przyczyną (*common cause*). Jest to ukryty

czynnik, który stanowi przyczynę obu zdarzeń. W omawianej sytuacji być może wspólną przyczyną jest światło słoneczne: im jest go więcej, tym jest goręcej i sprzedaje się więcej lodów. Podobnie im więcej światła słonecznego, tym niebo jest jaśniejsze, a ludzie zaczynają nosić okulary. Ważne, żeby pamiętać, że nie twierdzimy tu, że to światło słoneczne jest przyczyną obu zdarzeń. Jest to jedynie możliwe i dopuszczalne wyjaśnienie.

Jeśli pomiędzy dwoma wydarzeniami nie ma bezpośredniej lub odwrotnej relacji przyczynowej, to wtedy często korelacja może być wyjaśniona za pomocą jednego lub większej liczby czynników przyczynowych. Dobry obserwator już pewnie zauważył, że przykład ze strażakami i zniszczeniami może być wyjaśniony w bardzo podobny sposób. Im większy jest ogień, tym większe są zniszczenia, a zatem więcej strażaków jest potrzebnych do walki z ogniem. W wyjaśnieniu tym jest zawarty trzeci czynnik, a więc zaognienie się pożaru, które zwiększa nie tylko zniszczenia, ale także liczbę strażaków. Które z tych wyjaśnień jest właściwe? Główna myśl jest taka, że to, iż dwie zmienne są linearnie ze sobą powiązane, samo w sobie nie daje nam żadnych wskazówek.

Populacja osób bezdomnych jest skorelowana ze współczynnikiem przestępczości w ten sposób, że w tym samym czasie oba mogą być wysokie lub niskie.

Wielu ludzi z pewnością wyciągnęłoby wniosek, że obecność bezdomnych osób będzie skutkować działaniami o charakterze kryminalnym. Jednak, jeśli odnieść się do danych statystycznych, to teza, że populacja bezdomnych jest przyczyną przestępstw, jest nieuzasadniona: to, że populacja bezdomnych jest skorelowana ze współczynnikiem przestępstw, to to samo, co stwierdzenie, że współczynnik przestępstw jest skorelowany z populacją bezdomnych, jednak zdania, które mówią, że bezdomność jest

przyczyną przestępstw lub przestępstwa są przyczyną bezdomności bardzo się od siebie różnią. Które z nich jest prawdziwe? Oba? Żadne z nich? Może rolę odgrywają tu wspólne czynniki przyczynowe? Takimi czynnikami mogłyby być: uzależnienie od narkotyków, przemoc domowa, rozwody albo bezrobocie. Same zbiory skorelowanych ze sobą danych nie wystarczą do tego, żeby zidentyfikować wspólne przyczyny, dlatego należy uwzględnić tu dane z zakresu socjologii, psychologii, urbanistyki, kryminologii itd.

Wspólne przyczyny Reichenbacha

Czy zawsze istnieje wspólna przyczyna (lub wiele takich przyczyn), która stanowi podstawę każdej korelacji? Teza Reichenbacha głosi, że korelacja pomiędzy dwoma zdarzeniami A i B wskazuje na to, że albo A jest przyczyną B, albo B jest przyczyną A, albo A i B mają wspólną przyczynę. Można dyskutować o tym, czy tego typu zasadę można utrzymać. Zaś na tej stronie www.tylervigen.com/spurious-correlations.

Na zakończenie zostawiam Cię z pewnymi pytaniami, które możesz zabrać ze sobą do domu, żeby o nich pomyśleć. W porządku, korelacja nie ma nic wspólnego z przyczynowością, ale czy przyczynowość implikuje korelację? A tak w ogóle, to czym jest przyczynowość? Czy naprawdę mamy pewność, że piraci nie powodują globalnego ocieplenia?

Tłumaczenie Błażej Gębura

Pytania do tekstu

1. Czy z tego, że dwa zdarzenia często ze sobą współwystępują, można wyprowadzić wniosek, że posiadają tę samą przyczynę?
2. Czy teza Reichenbacha zakłada zasadę racji dostatecznej, która głosi, że każde zdarzenie lub stan rzeczy posiada swoją rację?
3. Podczas rozgrywki w snookera bila uderza drugą, wprowadzając ją tym samym w ruch. Czy jest to przykład korelacji?