

UPIORY STATYSTYKI

Bywa i tak, że przekopywanie się przez kolejne warstwy danych w poszukiwaniu znaczących liczb jest prawdziwym wyzwaniem. W 1973 roku Uniwersytet Kalifornijski w Berkeley oskarżono o dyskryminację kobiet w procesie rekrutacyjnym dla kandydatów na wyższy stopień studiów. Dowody wydawały się przekonujące. Spośród osób, które dotarły do ostatniej tury selekcji, przyjęto czterdzieści cztery procent mężczyzn i zaledwie trzydzieści pięć procent kobiet. Taka dysproporcja wydawała się niekorzystna dla uczelni; w najlepszym wypadku mogła uzasadnić konieczność dalszego śledztwa.

Gdy rok później statystyk Peter Bickel wraz ze współpracownikami przeanalizował zgromadzone dane, odkrył niezwykle zaskakujące zjawisko. W pierwszym etapie rozbił dane na poszczególne wydziały uczelni, szukając źródeł dyskryminacji. Dysproporcja tajemniczo zniknęła. Badając opinie komisji rekrutacyjnych wydziałów, zespół Bickela zauważył „kilka jednostek decyzyjnych wykazujących statystycznie istotne odchylenia od oczekiwanych wartości przyjęć kobiet oraz co najmniej tyle samo jednostek zdających się sprzyjać raczej kobietom niż mężczyznom”. Ogólny obraz okazał się zupełnie odwrotny od spodziewanego – wykazano „niewielką, lecz statystycznie istotną dysproporcję na korzyść kobiet”. Uczelnię oczyszczono z zarzutów, zagadka pozostała jednak nierozwiązana, ponieważ dane źródłowe były poprawne.

Odpowiedź częściowo wiąże się ze wspomnianą zasadą, że korelacja nie wskazuje przyczyny. W tym wypadku mamy też do czynienia z obrazowym przykładem zjawiska noszącego nazwę paradoksu Yule’a–Simpsona na cześć dwóch statystyków brytyjskich: Udny Yule dostrzegł paradoks w roku 1903, zaś Edward Simpson opisał go w roku 1951. Podobnie jak we wcześniejszych przykładach i tutaj nieporozumienie wiąże się z nieuwzględnieniem ważnego czynnika.

Bickel posłużył się metaforą łowienia ryb. Załóżmy, że w rzece zanurzono dwie sieci: oka jednej są duże, drugiej zaś – mniejsze. Nic dziwnego, że w gęstszą sieć złapie się więcej ryb. Niech będzie to 60% wszystkich ryb dla sieci o mniejszych okach i 25% dla tej o większych okach. Gdy założymy, że w rzece żyją ryby obu płci z określonych gatunków oraz że osobniki różnych płci mają tę samą wielkość, żadna z płci nie będzie faworyzowana. Mimo to może dojść do powstania nierównowagi, której źródłem jest pewien dodatkowy czynnik.

Wyobraźmy sobie, że samice wykazują tendencję do pływania środkiem nurtu, natomiast samce poruszają się bliżej brzegów. Gdyby rzadką sieć zanurzyć na godzinę na środku rzeki, przepłynie przez nią więcej samic niż samców: powiedzmy, że odpowiednio 80 i 16 osobników. Gęstą sieć zarzucono w pobliżu brzegu, gdzie w ciągu godziny przepłynęło przez nią 20 samic i 100 samców. W ostatecznym rozrachunku schwytano więcej samców, mimo że żadna z sieci nie była przystosowana do połowu wyłącznie osobników jednej płci. Wynik mógłby kształtować się następująco:

	Schwytane samice	Schwytane samce
Rzadka sieć w nurcie	$\frac{20}{80}$ (25%)	$\frac{4}{16}$ (25%)
Gęsta sieć przy brzegu	$\frac{12}{20}$ (60%)	$\frac{60}{100}$ (60%)
Razem	$\frac{32}{100}$ (32%)	$\frac{64}{116}$ ($\approx 55\%$)

Innymi słowy, używając dwóch całkowicie symetrycznych sieci, otrzymaliśmy połów o znacznej przewadze osobników męskich. Nawet gdyby zmodyfikować dane, symulując, że w każdą z sieci bardziej skłonne są wpadać osobniki żeńskie (na przykład dlatego, że są nieco większe od męskich), czynnik lokalizacji sieci i tak z łatwością przeważa:

	Schwytane samice	Schwytane samce
Rzadka sieć w nurcie	$\frac{30}{80}$ (37,5%)	$\frac{4}{16}$ (25%)
Gęsta sieć przy brzegu	$\frac{15}{20}$ (75%)	$\frac{60}{100}$ (60%)
Razem	$\frac{45}{100}$ (45%)	$\frac{64}{116}$ ($\approx 55\%$)

Mamy tu wyraźnie do czynienia z paradoksem Yule’a–Simpsona: sieci przystosowano do chwytania przede wszystkim samic, trend ulega jednak odwróceniu po zsumowaniu wyników połowów z obu sieci. W przypadku uczelni z Berkeley dodatkowym czynnikiem analogicznym do lokalizacji sieci były różnice w preferencjach wydziałów. Kobiety chętniej aplikowały na wydziały, na których liczba chętnych na jedno miejsce była wyższa, w związku z czym odsiewano większy odsetek kandydatów. Mężczyźni częściej próbowali się dostać na wydziały mniej oblegane. (Co ciekawe, autorzy uznali w tym miejscu matematykę za czynnik decydujący: „Wydziały, na które łatwiej się dostać, są też na ogół tymi, które od kandydata wymagają większej wiedzy w zakresie matematyki”).

Regresja do średniej także może skutecznie rozwiewać statystyczne miraży, obalając nieweryfikowalne twierdzenia. Przypomina nam również o niezwykle istotnej, lecz często niedocenianej wartości w życiu, jaką jest szczęście.

Łut szczęścia to wieczne marzenie wszystkich, którym przyszło zdawać jakiś egzamin. Wyobraźmy sobie, że Annabel i Betty podchodzą do dwóch testów. Wiadomo, że średnia krajowa zdawalność tych testów kształtuje się na poziomie pięćdziesięciu procent punktów. W pierwszej turze Annabel świetnie sobie radzi, zdobywając osiemdziesiąt osiem procent punktów; w drugiej ma niestety nieco mniej szczęścia, chociaż uzyskany przez nią wynik – siedemdziesiąt procent punktów – nadal jest ponadprzeciętny. Tymczasem Betty wypada fatalnie, zbierając z pierwszego testu zaledwie dwadzieścia pięć procent punktów. Drugi test, mimo że wciąż poniżej przeciętnej, zalicza już znacznie lepiej, otrzymując czterdzieści trzy procent punktów.

Nietrudno przewidzieć, jak mogą zinterpretować przedstawione wyniki rodzice i przyjaciele dziewcząt: sukces wywołał u Annabel zbytnią pewność siebie i samozadowolenie, dlatego nie starała się wystarczająco podczas drugiego sprawdzianu. Betty przeżyła potężny zawód po pierwszym egzaminie, dlatego znacznie poważniej podeszła do drugiego. Być może w tych spekulacjach jest ziarno prawdy, jednak podobnych zjawisk należy się spodziewać z przyczyn czysto statystycznych. To, jak ktoś poradzi sobie na klasówce, podczas meczu piłki nożnej, w biznesie bądź w miłosnym związku – w zasadzie w dowolnej sytuacji – zależy od kombinacji jego umiejętności oraz wysiłku i przygotowania, lecz także od szczęścia.

Wyobraźmy sobie tym razem, że testy, które rozwiązywały dziewczęta, są całkowicie losowe: składa się na nie sto pytań wielokrotnego wyboru i w każdym przypadku Annabel i Betty mają pięćdziesiąt procent szans na udzielenie poprawnej odpowiedzi spośród dwóch dostępnych. Gdybyśmy mieli spróbować przewidzieć wyniki, najrozsądniej byłoby obstawić pięćdziesiąt procent. Annabel ma jednak szczęście i w pierwszej turze zdobywa osiemdziesiąt osiem procent punktów. Co by się stało, gdybyśmy pokusili się o prognozę wyniku

drugiego testu? Ano nic. Z punktu widzenia statystyki jej szanse nadal wynosiłyby pięćdziesiąt procent. Z tej perspektywy sam fakt, że Annabel osiągnęła wysoki pierwszy wynik, sugeruje gorszą drugą notę. Analogicznie, istnieje duże prawdopodobieństwo, że niski wynik Betty zostanie poprawiony podczas drugiego sprawdzianu. W obu przypadkach są spore szanse na to, że dziewczęta przesuną się z zajmowanych skrajnych pozycji w kierunku średniej.

Identyczna sytuacja ma miejsce, gdy punktacja nie zależy wyłącznie od szczęścia, ale – jak zazwyczaj się dzieje – od kombinacji szczęścia, umiejętności i innych czynników. Załóżmy, że na sto punktów możliwych do zdobycia siedemdziesiąt to efekt realnej wiedzy, a trzydzieści to sprawa szczęścia. Załóżmy też, że poziom wiedzy obu dziewcząt jest stały i wynosi odpowiednio sześćdziesiąt u Annabel i dwadzieścia u Betty.

Annabel ma niesamowite szczęście i w pierwszym teście zdobywa dwadzieścia osiem punktów, natomiast Betty ma wielkiego pecha i dostaje tylko pięć z trzydziestu, które można otrzymać za sprawą szczęścia. Należy się spodziewać, że w drugim teście wyniki obu dziewcząt będą podlegać regresji do średniej.

	Annabel			Betty		
	Wiedza/70	Szczęście/30	Suma/100	Wiedza/70	Szczęście/30	Suma/100
Wynik 1. testu	60	28	88	20	5	25
Wynik 2. testu	60	10	70	20	23	43

Regresja do średniej to powszechne, choć wyjątkowo niedoceniane zjawisko. Pojawia się pod wieloma postaciami – szczególnie tam, gdzie występuje korelacja dwóch zdarzeń, ale nie tylko. Wyjątkowo częsta jest w sporcie, komentatorzy rzadko zwracają jednak na nią uwagę. I tak skoczek w dal, który w pierwszym skoku osiągnie doskonały wynik, z tego tylko powodu prawdopodobnie poradzi sobie gorzej za drugim podejściem. W 1989 roku magazyn „Sports Illustrated” odnotował, że dziewięćdziesiąt procent baseballistów, którzy w pierwszej połowie sezonu zdobyli punkt ponad dwadzieścia razy, rzadko powtarzało ten wynik w drugiej połowie sezonu. Z punktu widzenia statystyki nie jest to nic nadzwyczajnego.

Ten sam tygodnik dostarcza również innego słynnego przykładu regresji do średniej – często przywoływanego w przypadku szczęściarzy odnoszących niespodziewany sukces. Sportowiec, który ma za sobą wyjątkowo udany tydzień, może znaleźć się na okładce „Sports Illustrated”. Zauważono jednak, że w następnym tygodniu zazwyczaj doświadczą spadku formy. Czyżby nad okładką pisma ciążyło jakieś fatum? Nie. Po ponadprzeciętnym sukcesie wyniki rzekomych ofiar klątwy wracają po prostu do normy. Wszak korelacja nie wskazuje przyczyny.