

DYDAKTYKA I NAUKA

MIROSŁAW SZREDER

O WERYFIKACJI I FALSZYFIKACJI HIPOTEZ

Opracowanie to dedykuję Doktorantom

WPROWADZENIE

Prawdopodobnie artykuł ten nie powstałby, gdyby nie coraz częściej pojawiające się w pracach młodych naukowców sformułowania typu: „testowana hipoteza okazała się prawdziwa”, „hipoteza zerowa została pozytywnie zweryfikowana”, „można więc przyjąć, że weryfikowana hipoteza H_0 jest prawdziwa”. Tego rodzaju sformułowania, będące często konkluzją przeprowadzonych badań empirycznych, są dla statystyka trudne do zaakceptowania. Odnosi się to w szczególności do hipotez statystycznych, ale nie wyłącznie. Strategie weryfikacji hipotez w badaniach naukowych i w życiu codziennym, stanowiące przedmiot zainteresowania filozofów i psychologów, rzadko zawierają możliwość przesądzenia o tym, że weryfikowana hipoteza jest prawdziwa¹.

Celem tego opracowania jest przypomnienie zasad statystycznej teorii weryfikacji hipotez, wzbogacenie jej interpretacji, a także pokazanie logicznej zgodności, jaka istnieje między decyzjami podejmowanymi w procesie testowania hipotez naukowych i statystycznych. Przez hipotezy naukowe będziemy rozumieli przypuszczenia o charakterze uniwersalnym (globalnym), wywodzące się z odpowiedniej teorii. Hipoteza statystyczna natomiast, rozumiana będzie jako przypuszczenie dotyczące szczegółowych kwestii odnoszących się do ściśle zdefiniowanej populacji, które zweryfikować można w oparciu o próbę losową reprezentującą tę populację.

W zagadnieniach ekonomicznych, podobnie jak w socjologii, psychologii, medycynie, poprawna interpretacja wyników weryfikacji hipotez jest ważna, gdyż dostarcza przesłanek do podejmowania decyzji odnoszących się do całych, nieraz bardzo

¹ Interesujące omówienie poglądów filozofów i psychologów na weryfikację hipotez, jako na element poznania naukowego można znaleźć m.in. w pracach D.J. Glass i N. Hall [2008], J. Klayman i Y. Ha [1987], K. Popper [1995].

dużych populacji, o których badacz wypowiada się na podstawie informacji zawartych w próbie. Biorąc pod uwagę to, że próba stanowi zwykle niedużą frakcję populacji (ułamek procenta lub promila), ważne jest zarówno zastosowanie właściwej procedury weryfikacji hipotez, jak i poprawne zinterpretowanie jej wyników.

1. DOWODZENIE PRZEZ ZAPRZECZENIE (DOWODZENIE NIEPRAWDZIWOŚCI HIPOTEZY)

W statystycznej teorii weryfikacji hipotez, której twórcami są E.S. Pearson i J. Neyman, sprawdzaną hipotezę nazywa się *hipotezą zerową* i oznacza się przez H_0 . Równolegle formułuje się konkurencyjną wobec niej tzw. *hipotezę alternatywną* (H_1), która w przeciwieństwie do hipotezy zerowej ma najczęściej charakter złożony. Oznacza to, że hipoteza alternatywna nie określa jednej wartości parametru populacji, którego dotyczy wnioskowanie, lecz dopuszcza różne jego wartości, inne od tej, którą zawiera hipoteza zerowa. Na przykład: hipoteza zerowa stwierdza, że współczynnik korelacji między dwiema cechami w populacji jest równy zero, a hipoteza alternatywna głosi, że współczynnik ten jest większy od zera (czyli przyjąć może jedną z nieskończonej liczby wartości większych od zera, lecz nie równą zero). Mimo że hipotezą sprawdzaną (testowaną) jest hipoteza zerowa, to formułowana jest ona w taki sposób, aby dowieść jej **nieprawdziwości**. Jest to jedna z kluczowych kwestii w dobrym rozumieniu teorii wnioskowania statystycznego. Aby dowieść istnienia statystycznie istotnej korelacji między dwiema cechami, stawia się hipotezę, że współczynnik korelacji między tymi cechami jest równy zero. Odrzucenie tej hipotezy (**stwierdzenie jej nieprawdziwości**) stanowi statystyczny dowód na to, że między badanymi cechami istnieje istotna zależność korelacyjna. Podobnie jest w przypadku hipotez nieparametrycznych, odnoszących się nie do parametrów, lecz pewnych atrybutów populacji. Aby dowieść, że między dwiema cechami w populacji istnieje współzależność, sprawdzaną hipotezę formułuje się na zasadzie zaprzeczenia, tj. że analizowane cechy są niezależne. Także tutaj odrzucenie sprawdzanej hipotezy w oparciu o wynik próby losowej będzie statystycznym dowodem na to, że cechy te są zależne. Zanim omówimy powody takiego postępowania (pkt 2), wyjaśnijmy najpierw, co właściwie oznacza odrzucenie hipotezy i dlaczego decyzję tę traktujemy jako dowód statystyczny nieprawdziwości hipotezy zerowej.

Procedura weryfikacji hipotez statystycznych zakłada, że hipoteza zerowa jest dotąd prawdziwa, dopóki nie znajdą się silne przesłanki każące wątpić w jej prawdziwość². Jedynym źródłem możliwych przesłanek mogą być – w klasycznym ujęciu – informacje pochodzące z próby losowej. Odrzucenie testowanej hipotezy następuje wówczas, gdy bardzo małe jest prawdopodobieństwo wystąpienia (realizacji) zaobserwowanej próby, przy założeniu że prawdziwa jest hipoteza zerowa. Na przykład, jeżeli hipoteza H_0 głosi, że w populacji średni wzrost osób wynosi 177 cm, a z próby losowej uzyskano średni wzrost 155 cm, to możemy albo przyjąć, że zrealizowało się zdarzenie o bardzo małym prawdopodobieństwie (wylosowano nietypową próbę zdominowaną przez naj-

² *A null hypothesis is regarded as valid unless the evidence suggests that it is not true*, J. Kmenta [1990], s. 112. Por. także Cz. Domański [1990].

nizsze osoby z tej populacji), albo że hipoteza zerowa jest nieprawdziwa³. Jeżeli mielibyśmy uznać, że realizacja próby była zdarzeniem o prawdopodobieństwie mniejszym od pewnego minimalnego α (np. 0,01), to skłonni jesteśmy raczej odrzucić hipotezę zerową, czyli stwierdzić, że jest nieprawdziwa. Podejmując decyzję o odrzuceniu H_0 możemy jednak popełnić błąd, gdy hipoteza ta będzie w rzeczywistości prawdziwa. Na szczęście znamy prawdopodobieństwo popełnienia takiego błędu. Odrzucenie sprawdzanej hipotezy, gdy jest ona prawdziwa, oznacza, że zrealizowała się próba losowa o prawdopodobieństwie mniejszym od α . Jeżeli $\alpha = 0,01$, to próba taka zdarza się będzie nie częściej niż średnio jeden raz na sto prób. Podejmując decyzję o odrzuceniu (stwierdzeniu nieprawdziwości) hipotezy zerowej wiemy, z jakim prawdopodobieństwem może to być decyzja błędna. Nie tylko znamy to prawdopodobieństwo, ale sami jego wielkość ustalamy. Posługujemy się więc procedurą, która może być zawodna z prawdopodobieństwem nie większym niż α . Właśnie z tego powodu, iż w pełni kontrolujemy prawdopodobieństwo podjęcia błędnej decyzji, mamy prawo twierdzić, że odrzucenie hipotezy zerowej jest statystycznym dowodem jej nieprawdziwości, a tym samym prawdziwości hipotezy konkurencyjnej H_1 (najczęściej złożonej).

2. PRÓBA STATYSTYCZNA A PRAWDZIWOŚĆ TESTOWANEJ HIPOTEZY

Jeżeli zastosowanie testu statystycznego dla danego zbioru obserwacji w próbie nie prowadzi do odrzucenia hipotezy zerowej, to właściwą w tych okolicznościach decyzją jest stwierdzenie, że brak jest podstaw do odrzucenia H_0 . Innymi słowy – próba nie dostarczyła wystarczających przesłanek do odrzucenia sprawdzanej hipotezy. Z kolei **brak przesłanek do stwierdzenia nieprawdziwości hipotezy H_0 nie oznacza dowodu na jej prawdziwość**. W teorii wnioskowania statystycznego nie istnieje konkluzja o stwierdzeniu prawdziwości hipotezy zerowej, a jedynie o braku podstaw do uznania jej za nieprawdziwą. Nie są to kategorie tożsame⁴. Wyjaśnienie, dlaczego empiryczne bada-

³ Zauważmy, że im większa liczebnie będzie próba, a także im mniejsze rozproszenie (dyspersja) wzrostu osób w tej populacji, tym trudniej będzie zaakceptować tak duże (22 cm) odchylenie średniej w próbie od średniej w populacji. Szczególnie warto tu zwrócić uwagę na znaczenie liczebności próby. Im większa próba, tym mniej prawdopodobne jest wystąpienie dużej różnicy między danym parametrem populacji a jego odpowiednikiem (estymatorem) w próbie. W codziennym myśleniu nie zawsze relację tę bierzemy pod uwagę, stając się „niewrażliwi na liczebność próby” (ang. *insensitivity to sample size*). A. Tversky i D. Kahneman [1974] oraz M.H. Bazerman i D. Moore [2009] ilustrują to następującym przykładem: W pewny mieście funkcjonują dwa szpitale. W większym z nich rodzi się dziennie około 45 noworodków, a w mniejszym – około 15 noworodków. Jak wiadomo, przeciętnie 50% noworodków jest płci męskiej. Jednak rzeczywisty udział chłopców wśród noworodków jest różny każdego dnia, bywa mniejszy lub większy od 50%. Przez jeden rok każdy z tych szpitali zanotował dni, w których udział względny urodzonych chłopców przekraczał 60%. Na pytanie, który ze szpitali zanotował więcej takich dni w roku, pada często odpowiedź, że szpitale te zanotowały w przybliżeniu tę samą liczbę dni. W rzeczywistości jednak, przy dużej liczbie urodzin (dużej liczebności próby) zdarzenie polegające na znacznym odchyleniu proporcji w próbie (o 10 pkt proc.) w stosunku do proporcji w populacji jest mniej prawdopodobne, niż w przypadku małej liczby urodzin. Mniejszy szpital będzie więc rejestrował więcej takich dni w roku.

⁴ W jęz. angielskim do wyrażenia tej myśli używa się powszechnie stwierdzenia *absence of evidence is not evidence of absence*. Brak dowodów (np. współzależności cech) nie jest dowodem braku (współzależności tych cech).

nie **nie może** stanowić podstawy do stwierdzenia prawdziwości testowanej hipotezy, ma podstawy logiczne, statystyczne i filozoficzne. Te ostatnie, jako że dotyczą szerszej rozumianych hipotez (naukowych), zostaną zaprezentowane w odrębnym punkcie (pkt 3).

Logiczne uzasadnienie podanej wyżej tezy wymaga przypomnienia, że w procedurze weryfikacji hipotez statystycznych jednym z ważnych etapów jest przyjęcie założenia o prawdziwości hipotezy H_0 . Przy założeniu, że sprawdzana hipoteza jest prawdziwa, następuje obliczenie zarówno wartości statystyki testowej, jak i wyznaczenie prawdopodobieństwa krytycznego (ang. *p-value*)⁵. Dopuszczenie możliwości stwierdzenia prawdziwości hipotezy zerowej w wyniku zastosowania procedury testowania prowadziłoby do następującej sprzeczności logicznej. Najpierw zakłada się, że hipoteza H_0 jest prawdziwa, stosuje się odpowiednią procedurę testową, aby w jej efekcie dojść do wniosku, że hipoteza H_0 jest prawdziwa. Ale przecież założono na wstępie, że jest prawdziwa! Warto więc powtórzyć, że:

a) logicznej sprzeczności unika się, gdy możliwymi decyzjami w procesie weryfikacji hipotez są: albo odrzucenie sprawdzanej hipotezy, albo stwierdzenie o braku wystarczających podstaw do jej odrzucenia;

b) rzeczywista użyteczność testowania statystycznego przejawia się przede wszystkim w możliwości stwierdzenia nieprawdziwości sprawdzanej hipotezy (odrzuconiu H_0 na rzecz konkurencyjnej hipotezy H_1).

Statystyczne uzasadnienie niemożności stwierdzenia prawdziwości hipotezy zerowej brzmi najbardziej przekonująco w ogólniejszym kontekście wnioskowania statystycznego. Odwołamy się tu w szczególności do estymacji przedziałowej. Zwróćmy uwagę, że w większości testów parametrycznych określona statystyka testowa służy także do konstrukcji przedziału ufności dla parametru populacji θ , którego dotyczy wnioskowanie. Informacje uzyskane z próby pozwalają zarówno obliczyć wartość statystyki testowej, jak i skonstruować odpowiedni przedział ufności dla θ . Jeżeli w wyniku testowania hipotezy zerowej otrzymano wartość statystyki leżącą poza obszarem krytycznym, a więc stwierdzono brak podstaw do odrzucenia H_0 na poziomie istotności α , to w kontekście estymacji oznacza to, że z dużym prawdopodobieństwem $(1 - \alpha)$ parametr θ mieści się w przedziale zawierającym wartość hipotetyczną θ_0 (tj. wartość parametru θ wyspecyfikowaną w hipotezie zerowej). Innymi słowy oznacza to, że **z dużym prawdopodobieństwem można twierdzić, że jedną z możliwych wartości nieznanego parametru populacji θ jest θ_0** . Nie jest ona jedyną wartością (co dawałoby podstawy do stwierdzenia, że udowodniono prawdziwość $H_0 : \theta = \theta_0$), a zaledwie jedną z nieskończonej liczby wartości mieszczących się w przedziale ufności dla θ . Zilustrujemy to niżej prostym przykładem.

Przykład 1.

Przyjmijmy, że w populacji o rozkładzie normalnym, w której odchylenie standardowe (σ) wynosi 4, nieznana jest wartość średnia μ . W oparciu o losową próbę $n = 16$ elementów, z której obliczono średnią ($\bar{X} = 11$), należy na poziomie istotności

⁵ Przez prawdopodobieństwo krytyczne rozumie się najmniejszą wartość poziomu istotności, na którym może być odrzucona hipoteza zerowa.

$\alpha = 0,10$ zweryfikować hipotezę, że średnia w populacji μ jest równa 10. Hipotezy zerową i alternatywną sformułowano następująco:

$$\begin{aligned} H_0 : \mu &= 10 \\ H_1 : \mu &\neq 10 \end{aligned}$$

Wartość odpowiedniej statystyki testowej Z , obliczona przy założeniu prawdziwości hipotezy zerowej, równa się:

$$Z = \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} = \frac{11 - 10}{4} \cdot 4 = 1.$$

Wartość ta nie mieści się w obszarze krytycznym $(-\infty, -1,64) \cup (1,64, +\infty)$, stąd właściwą decyzją jest stwierdzenie o braku podstaw do odrzucenia hipotezy H_0 . Dlaczego nie jest to tożsame z uznaniem hipotezy zerowej za prawdziwą? Prezentując uzyskany wynik z próby w kontekście estymacji przedziałowej:

$$\begin{aligned} P\left\{\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right\} &= 1 - \alpha \\ 11 - 1,64 \cdot \frac{4}{4} \leq \mu \leq 11 + 1,64 \cdot \frac{4}{4} \\ 9,36 \leq \mu \leq 12,64 \end{aligned}$$

oznacza on, że z prawdopodobieństwem 90% można twierdzić, iż nieznaną średnią w populacji μ zawiera się w przedziale (9,36; 12,64) i że prawdziwa wartość tej średniej może być dowolną liczbą z tego przedziału. Może to być także wartość 10 (ujęta w sprawdzanej hipotezie), ale nie jest ona jedyną możliwą wartością. A skoro nie jedyną, to nie możemy twierdzić, że wskazana w hipotezie wartość jest prawdziwa, gdyż wykluczałoby to prawdziwość jakiegokolwiek innej wartości z przedziału (9,36; 12,64).

Głównym powodem tego, że wynik z próby losowej nie rozstrzyga o prawdziwości hipotezy określającej konkretną wartość parametru populacji jest niedoskonałość mechanizmu losowania⁶. Każda technika losowania jest obciążona błędem losowania, sprawiającym że uzyskana próba prawie nigdy nie jest doskonałą reprezentacją populacji. Wynik z próby, który prowadzi do stwierdzenia braku podstaw do odrzucenia sprawdzanej hipotezy statystycznej, nie jest dowodem prawdziwości tej hipotezy. Analogiczne do tego rozumowanie jest szeroko stosowane także w odniesieniu do hipotez naukowych.

3. FALSYFIKACJA HIPOTEZ NAUKOWYCH

Przedstawiona wyżej w pkt 1 i 2 interpretacja procedury testowania hipotez statystycznych jest zgodna z zasadami stosowanymi w odniesieniu do hipotez naukowych.

⁶ Szerzej o naturze błędu losowania por. M. Szreder [2010], s. 170-173.

W obu przypadkach ważna jest możliwość stwierdzenia nieprawdziwości weryfikowanej hipotezy, możliwość jej obalenia. E. Babbie w książce pt. *Badania społeczne w praktyce* pisze: „Każda znacząca hipoteza musi być obalalna. Jeżeli nie jesteś w stanie określić, jakie odkrycia empiryczne stanowiłyby zaprzeczenie twojej hipotezy, nie jest to prawdziwa hipoteza w naukowym znaczeniu tego słowa” (E. Babbie [2004], s. 474). W obu też przypadkach – hipotez statystycznych i hipotez naukowych – podkreśla się niemożność udowodnienia prawdziwości hipotezy na podstawie cząstkowych (próbkowych) danych empirycznych. „Żadna liczba białych łabędzi nie ustala tego, że wszystkie łabędzie są białe” – pisze Karl Popper [1995], s. 121⁷. I dalej dodaje: „Gdy mówimy, że nasza wiedza nie jest pewna, mamy na myśli tylko to, że nigdy nie możemy być pewni, czy nasze przypuszczenia są prawdziwe. Gdy stwierdzamy, że hipoteza nie jest prawdziwa lub przynajmniej że nie wydaje się stanowić lepszego przybliżenia do prawdy niż jej konkurentki, możemy ją odrzucić. Hipotezy nigdy nie są weryfikowalne, można je jednak sfalsyfikować. Można je krytykować i sprawdzać” (K. Popper [1995], s. 226). Termin „sfalsyfikować” oznacza wykazać lub ustalić nieprawdziwość (fałszywość) hipotezy. Formułowanie pewnych hipotez, obalenie niektórych z nich i na tej podstawie stawianie nowych – to jedna z dróg poznania naukowego.

Oczywiście, obserwacje empiryczne (także statystyczne) mogą potwierdzać pewne hipotezy lub je wspierać, ale znaczenie takiego faktu jest dla nauki mniejsze niż wykazanie nieprawdziwości hipotezy. W codziennym życiu bywa odmiennie. Staramy się niekiedy poszukiwać faktów lub interpretować je w taki sposób, aby potwierdzały nasze pierwotne aprioryczne przekonania (nasze hipotezy). Zjawisko to nazywane jest „pozytywnym testowaniem hipotez” (ang. *positive hypothesis testing*) i ściśle łączone jest z tzw. błędem konfirmacji (ang. *confirmation bias*), por. M.H. Bazerman, D. Moore [2009], s. 8-9. Błąd ten polega na poszukiwaniu i przetwarzaniu tych informacji, które potwierdzają postawioną hipotezę, z jednoczesnym pomniejszaniem znaczenia obserwacji, które hipotezie tej przeczą. W działalności naukowej tego typu skłonność prowadzić może do nierzetelnych lub fałszywych wyników. Analogią z zakresu wnioskowania statystycznego byłoby postępowanie, którego celem jest zwiększanie liczbeności próby dla wzmocnienia siły argumentacji o prawdziwości hipotezy zerowej. Tymczasem, jak stwierdziliśmy wcześniej, nawet przy bardzo dużej próbie nie da się wykazać prawdziwości sprawdzanej hipotezy, a jedynie pokazać, iż próba wspiera tę hipotezę (nie daje przesłanek do jej odrzucenia). Statystyk, filozof, psycholog – wszyscy oni zainteresowani są takim formułowaniem hipotez, których fałszywość można udowodnić materiałem empirycznym i w ten sposób potwierdzić hipotezę konkurencyjną (alternatywną).

LITERATURA

- Babbie E., [2004], *Badania społeczne w praktyce*, PWN, Warszawa.
Bazerman M.H., Moore D., [2009], *Judgment in managerial decision making*, (7th edition), J. Wiley.
Calaprice, A., [2005], *The new quotable Einstein*, Princeton University Press.

⁷ Podobne stwierdzenie przypisuje się też Einsteinowi: „Żadna liczba eksperymentów nie przekona mnie, że mam rację; jeden eksperyment może udowodnić, że nie mam racji” (na podst. A. Calaprice [2005], s. 291).

Domański Cz., [1990], *Testy statystyczne*, PWE, Warszawa.

Glass D.J., Hall N., [2008], *A brief history of the hypothesis*, „Cell” Vol. 134, No. 3.

Klayman J., Ha Y.W., [1987], *Confirmation, disconfirmation, and information in hypothesis testing*, „Psychological Review”, Vol. 94, No. 2.

Kmenta J., [1990], *Elements of econometrics*, Macmillan Publishing Company, New York.

Popper K.R., [1995], *Mit schematu pojęciowego. W obronie nauki i racjonalności*, Książka i Wiedza, Warszawa.

Szreder M., [2010], *Metody i techniki sondażowych badań opinii*, PWE, Warszawa.

Tversky A., Kahneman D., [1974], *Judgment under uncertainty: Heuristics and biases*, „Science”, Vol. 185, pp. 1124-1131.