BADANIA PRÓBKOWE – PROJEKTOWANIE I WNIOSKOWANIE

Mirosław Szreder, Katedra Statystyki, Uniwersytet Gdański

Wśród wielu nowych trendów obserwowanych w opisie rzeczywistości na uwagę zasługuje upowszechnienie się pomiaru liczbowego, który w wielu dziedzinach życia społecznego i gospodarczego zdaje się wypierać opis jakościowy (werbalny). Dążenie do oceny nieraz bardzo skomplikowanych zjawisk za pomocą kilku wskaźników, indeksów, pozycji w rankingu staje się wyznacznikiem współczesnego trybu życia – szybkiego, oferującego dostęp do wielu źródeł informacji. Opis liczbowy jest zwykle oparty na wynikach badań próbkowych (niewyczerpujących), obejmujących jedynie małą reprezentację populacji.

Fundamentem wnioskowania o populacji pozostaje w tego typu badaniach teoria wnioskowania statystycznego Neymana-Pearsona. Z wielu jednak powodów próba losowa przestaje być jedynym i wystarczającym źródłem informacji dla statystyka. Głównymi powodami dokonującej się ewolucji w tym zakresie są: z jednej strony – wzrost znaczenia błędów nielosowych w całkowitym błędzie badania próbkowego (braki odpowiedzi, wady operatów losowania), a z drugiej – rosnące zasoby informacji o wielu populacjach poddawanych badaniom i coraz łatwiejszy dostęp do tych informacji. Te nowe zjawiska powinny być brane pod uwagę w projektowaniu badania już na etapie wyboru techniki próbkowania. Jednym z najważniejszych wyzwań dla statystyka jest obecnie umiejętne wykorzystanie całości zgromadzonej na wstępie wiedzy (*a priori*) w badaniu, w którym głównym źródłem informacji pozostaje wciąż próba statystyczna.

Przesłanki realizacji badań próbkowych

Na upowszechnienie się badań próbkowych wpływ wywarły w ostatnich dziesięcioleciach zarówno postępy w teorii statystyki i rozwoju odpowiedniego oprogramowania, jak i ogólne przyspieszenie tempa życia i oczekiwań szybkich wyników przez zlecających badania[1]. W stosunku do badań pełnych (wyczerpujących) atutem badań próbkowych są niższe koszty, a także – na co nie zawsze wystarczająco zwraca się uwagę – większa ich efektywność, jeżeli mierzyć ją stosunkiem jakości wyników (ich wiarygodności i precyzji) do poniesionych nakładów. Jest tak dlatego, że wśród kilku kategorii błędów, jakimi są

[1] „*Statystyki GUS są opóźnione zwykle o ponad miesiąc, a jeśli chodzi o dane szczegółowe – o kilka miesięcy. Dziś wiedza o tym, co się działo przed kilku miesiącami, jest prehistorią*" (W. Gadomski [2009], s. 22).

Copyright © StatSoft Polska 2011, info@DaneWiedzaSukceS.pl

obciążone wyniki prawie każdego badania statystycznego, w badaniach próbkowych pojawia się tylko jedna dodatkowa kategoria błędu – błąd losowania (ang. *sampling error*). Poza błędem losowania istnieją cztery inne kategorie błędów, którymi mogą być obciążone wyniki, zarówno badań wyczerpujących, jak i niewyczerpujących. Są to:

- błąd pokrycia jednostek badanej zbiorowości przez operat losowania (ang. *coverage error*), który może być wynikiem wykorzystania w badaniu nieaktualnego lub niekompletnego spisu jednostek populacji;
- błąd spowodowany brakiem odpowiedzi respondentów (ang. *nonresponse error*);
- błąd pomiaru (ang. *measurement error*), związany z zarejestrowaniem nieprawdziwych informacji o badanym respondencie;
- błąd przetwarzania zebranych danych statystycznych (ang. *postsurvey processing error*).

W stosunku do błędu losowania, który najczęściej jako jedyny w postaci skwantyfikowanej zostaje przypisany wynikom badania próbkowego, inne błędy mogą znacznie bardziej zniekształcać ostateczne rezultaty, bez względu na to, czy obserwacji poddano wszystkie jednostki populacji, czy tylko ich reprezentację.

Błąd pokrycia odgrywa tu szczególną rolę, gdyż od jakości odwzorowania populacji w operacie losowania zależy rzeczywisty dostęp badacza do populacji docelowej (ang. *target population*). W tym zakresie nawet organa statystyki publicznej, które – jak mogłoby się wydawać – powinny mieć dobrej jakości rejestry pełniące funkcje operatów, nie zawsze są w stanie sprostać wymogom jakościowym w tym zakresie.

„Wydaje się, że obecność błędów pokrycia w badaniach statystycznych GUS była dostrzegana dość dawno, ale stanowiły one problem wstydliwy, o którym oficjalnie nie pisano”.

J. Paradysz [2010], s. 52

Tym bardziej trudno jest o dostęp do kompletnego i aktualnego operatu w badaniach wykonywanych przez mniejsze podmioty na rynku, mające skromniejsze zasoby i możliwości. Dotyczy to m.in. badań opinii społecznej realizowanych drogą telefoniczną. W wielu przypadkach można mieć uzasadnione obawy co do kompletności operatu w takich sondażach. Ważne jest jednak to, że błąd pokrycia nie jest wyłączną specyfiką badań próbkowych.

Podobnie rzecz się ma z błędem braków odpowiedzi części respondentów odmawiających współpracy. Wystąpienie znacznej liczby odmów może prowadzić do systematycznego obciążenia wyników badania. I nawet w tak ważnych badaniach, jak spisy powszechne, trudno jest błędy te wyeliminować:

„Błędy systematyczne w narodowych spisach powszechnych są znacznie większe niż się nam dotychczas wydawało”.

J. Paradysz [2010], s. 63

W badaniach ankietowych i sondażowych jest to współcześnie ten rodzaj błędów, który sprawia najwięcej trudności statystykom. Odmów jest w wielu badaniach poważny

Copyright © StatSoft Polska 2011, info@DaneWiedzaSukceS.pl

odsetek[2], a ich skutków nie niweluje zwiększenie liczebności próby. Błąd, jaki generują braki odpowiedzi, może znacznie przewyższać typową wielkość błędu losowania.

Wskazując zatem na przesłanki badań próbkowych, warto zaznaczyć, że towarzyszący tym badaniom **błąd losowania bywa często najmniejszym składnikiem całkowitego błędu**, którym obciążone są wyniki badań statystycznych. W tym kontekście wyzwaniem dla statystyków jest z jednej strony lepsze poznanie natury i wielkości błędów nielosowych, a z drugiej doskonalenie sposobów minimalizacji skutków tych błędów. Już przed ponad pół wiekiem niektórzy znani statystycy zauważali zbyt małe zainteresowanie badaczy błędami innymi niż błąd losowania. Słynny hinduski statystyk P.C. Mahalanobis w 1951 r. pisał:

„Ogólna postawa jest taka, że błąd nielosowy postrzega się jako coś, co nie dotyczy statystyka, a w każdym razie uważany jest za niezbyt jasną kwestię, którą statystyk nie musi się przejmować”[3].

Z kolei amerykański znawca technik próbkowania L. Kish twierdził, że:

„Błąd losowania jest nadmiernie badany”[4].

Pomiar błędów nielosowych, a także korekty wyników obciążonych tymi błędami (m.in. z wykorzystaniem technik imputacji i kalibracji) – oba te zagadnienia wiążą się ściśle z szeroką problematyką wykorzystania informacji *a priori* w statystycznych badaniach próbkowych. Gromadzenie, ocena i umiejętne wykorzystanie informacji spoza próby staje się jednym z najważniejszych czynników wpływających na jakość badań próbkowych. Źródeł tych dodatkowych informacji o wielu poddawanych badaniom populacjach jest obecnie więcej, niż było kiedykolwiek w przeszłości. Ważne jest więc dostrzeżenie możliwości ich wykorzystania na wszystkich etapach badania próbkowego.

Próby losowe czy nielosowe?

Dla statystyka dylemat, czy próba powinna być losowa czy nielosowa, właściwie nie istnieje. Teoria klasycznego wnioskowania statystycznego oparta jest na modelu matematycznym, w którym zakłada się, że do próby dostają się jednostki wygenerowane przez mechanizm losujący, który każdej jednostce populacji daje szansę znalezienia się w próbie. Statystyka dostarcza znacznie więcej sformalizowanych narzędzi wnioskowania w sytuacji,

---

[2] W co najmniej kilku sondażach zrealizowanych przez największe w naszym kraju instytuty badawcze w czasie ubiegłorocznej kampanii prezydenckiej udziału w badaniu odmawiało ponad 80% badanych, a były też takie, w których odmów było 93% (por. *Ocena metodologii i rezultatów...* [2010]).

[3] “*In fact, the general attitude is to look upon the non-sampling error as something, which does not concern the statistician, or in any case is a kind of dirty job, which a highbrow statistician need not bother about*”, P.C. Mahalanobis [1951], s. 4.

[4] "*Sampling error is »over-researched«*". Sformułowanie to pojawiło się w interesującym artykule znanych statystyków Richarda Platka i Carla-Erika Särndala pt. "*Can a statistician deliver?*", opublikowanym w jęz. polskim wraz z dyskusją przez miesięcznik "Wiadomości Statystyczne" nr 4, 2001 r.

Copyright © StatSoft Polska 2011, info@DaneWiedzaSukceS.pl

gdy do wspomnianego modelu matematycznego może się odwołać, aniżeli wtedy, gdy próba zostaje pobrana z populacji w inny sposób. Główną zaletą prób probabilistycznych, w szczególności próby losowej prostej, jest możliwość stosowania w dalszej analizie (wnioskowaniu) zasad rachunku prawdopodobieństwa (stąd nazwa: próba probabilistyczna, ang. *probability* – prawdopodobieństwo). Tam, gdzie nie ma próby losowej, mechanizmu losującego, czy wreszcie zdarzeń losowych, tam nie ma zastosowania klasyczny rachunek prawdopodobieństwa.

Najważniejsze elementy wnioskowania statystycznego, w tym: określenie wielkości próby, interpretacja wyników estymacji i testowania hipotez, charakterystyka wielkości błędu – wszystkie one w sposób bezpośredni odwołują się do prawdopodobieństwa. Oznacza to, że u samych podstaw najważniejszych aspektów wnioskowania statystycznego leży założenie o możliwości stosowania rachunku prawdopodobieństwa (założenie losowości próby badawczej). Prawdopodobieństwo otrzymuje tutaj interpretację częstościową – jedną z czterech najpopularniejszych interpretacji, obok: klasycznej, logicznej i personalistycznej (subiektywnej)[5].

W estymacji przedziałowej prawdopodobieństwo występuje *explicite*, jako poziom ufności *(1-α)*, który interpretować należy jako częstość względną liczby przedziałów ufności pokrywających szacowany parametr populacji w wielokrotnie powtarzanych losowaniach prób, z których każda określa własne granice przedziału ufności. Gdyby tę samą procedurę wielokrotnie powtarzanych prób zastosować do technik nielosowych próbkowania, w których mechanizm generowania obserwacji w próbie nie jest losowy, to nie byłoby żadnych podstaw do zastosowania rachunku prawdopodobieństwa. W konsekwencji poziom ufności nie pozwalałby na intepretację mówiącą cokolwiek o prawdopodobieństwie popełnienia (lub niepopełnienia) błędu w przedziałowej estymacji parametru. Dlatego zastosowanie podanej wyżej interpretacji do estymacji opartej na próbach nieprobabilistycznych (nielosowych) uznać należałoby za nieupoważnione.

Podobnie rzecz się ma z testowaniem hipotez statystycznych. Pojęcie prawdopodobieństwa występuje tu wprost jako poziom istotności (α) – prawdopodobieństwo popełnienia błędu polegającego na odrzuceniu hipotezy, gdy w rzeczywistości jest ona prawdziwa. Prawdopodobieństwo popełnienia tego błędu jesteśmy w stanie ustalić, bo potrafimy obliczyć częstość względną nietypowych prób w długim ciągu losowań. Zwróćmy uwagę, że prawdopodobieństwo α odnosi się wyłącznie do niedoskonałości mechanizmu losującego, odpowiedzialnego za błąd losowania. Błąd losowania i przypisane mu prawdopodobieństwo nie uwzględniają niczego poza niedoskonałością samego aktu losowania. A jeżeli tak, to zrozumiałe jest, że błąd ten nie występuje tam, gdzie w ogóle losowania nie ma. W technikach nieprobabilistycznych prawdopodobieństwo to (α) nie miałoby żadnej interpretacji. **Zdarzające się zastosowania teorii weryfikacji hipotez do prób nielosowych powodują, że traci się w tych warunkach możliwość określenia prawdopodobieństwa podjęcia błędnej decyzji**, a także możliwość interpretacji przyjętego poziomu istotności (szerzej na ten temat w kolejnym punkcie niniejszego opracowania).

[5] Szerzej na ten temat - por. Szreder [1994] i Szreder [2010].

Copyright © StatSoft Polska 2011, info@DaneWiedzaSukceS.pl

Do prawdopodobieństwa odwołujemy się także w ważnym zagadnieniu określenia niezbędnej wielkości próby badawczej. Zwróćmy najpierw uwagę na to, że pytanie: "*Jak duża powinna być w danym badaniu próba?*" jest niepełne i trudno jest na nie w ogóle odpowiedzieć, jeżeli nie poda się jakiegoś kryterium precyzji lub dokładności wnioskowania. Dopiero w połączeniu z takim kryterium, pytaniu o liczebność próby można nadać odpowiedni sens logiczny i sformułować je przy pomocy właściwych kategorii statystycznych. Na przykład: "*Jaka powinna być liczebność próby, aby przeciętne odchylenie uzyskanych w badaniu ocen od prawdziwej wartości szacowanego parametru nie różniło się więcej niż o 2% lub o 5 jednostek miary, w której wyrażona jest badana cecha?*". W badaniach reprezentacyjnych przyjęło się najczęściej stosować łącznie dwa kryteria: średniego lub maksymalnego błędu oszacowania oraz ustalonego poziomu ufności. Oba te kryteria wiążą się bezpośrednio z losowaniem próby, a nie z jakimkolwiek innym sposobem jej wyboru.

Do niedawna dość powszechne było przekonanie, że techniki nielosowego doboru próby są tym gorszym wariantem, wymuszonym okolicznościami, które nie pozwalają na zastosowanie technik probabilistycznych. Sytuacja ta jednak zmienia się wraz ze wzrostem zasobów informacji, które obok obserwacji w próbie statystyk chciałby włączyć do wnioskowania. Kiedyś mógł ich nie mieć albo nie mieć do nich łatwego dostępu. Obecnie coraz rzadziej badacz znajduje się w sytuacji zupełnego braku wiedzy o badanej populacji. Najczęściej wiedzę taką, mimo że cząstkową i niedoskonałą, ma lub może mieć. I problemem nie jest to, czy ją wykorzystać, lecz jak ją wykorzystać. Jest to kluczowa sprawa dla zrozumienia zarówno coraz śmielszego odwoływania się statystyków do personalistycznej interpretacji prawdopodobieństwa, jak i coraz większej popularności prób nielosowych w badaniach niewyczerpujących. Ostatecznym bowiem **celem wnioskowania statystycznego nie jest osiągnięcie doskonałości w próbkowaniu, lecz jak najlepsze poznanie badanej populacji**. Gdyby mechanizm losowania próby był doskonały, to badacz nie miałby powodów, by w niego ingerować. A ingeruje coraz silniej, gdyż ma coraz bogatszą wiedzę o populacji, wiedzę użyteczną, którą pragnie włączyć do próbkowania w celu poprawy jakości wnioskowania.

Ingerencja badacza w klasyczny schemat próby losowej prostej jest widoczna w powszechnie znanych schematach próbkowania, takich jak losowanie systematyczne, czy losowanie warstwowe. W tym ostatnim szczególnie dobrze widać przekonanie o niedoskonałości modelu próby losowej prostej w sytuacji, w której badacz zna (*a priori*) zróżnicowanie wewnętrzne populacji ze względu na cechy istotne dla celu badania. Ingerowanie zaś w mechanizm losowania to nic innego jak stopniowe jego zastępowanie wyborem opartym na informacjach apriorycznych o badanej zbiorowości. Dlatego wraz ze zwiększaniem się zasobów informacji o badanych populacjach widoczna staje się tendencja do coraz częstszego korzystania w praktyce ze schematów losowania nieprostego[6], a także z technik nielosowego doboru próby, takich jak wybór kwotowy. Gdy o wielokrotnie badanej populacji badacz ma stosunkowo bogatą wiedzę, to kwotowy (nieprobabilistyczny) wybór

[6] O wnioskowaniu dla prób nieprostych w wielu swoich pracach pisze prof. Cz. Domański wraz z zespołem, por. np. Cz. Domański i K. Pruska [2000].

Copyright © StatSoft Polska 2011, info@DaneWiedzaSukceS.pl

próby może się okazać lepszy od schematów probabilistycznych. Ilustracją tego niech będzie stwierdzenie znanego brytyjskiego ośrodka badania opinii i rynku MORI o tym, że wybór kwotowy jest nie tylko najpowszechniej stosowany w Wielkiej Brytanii ("*most polls in Britain use quota sampling which pre-determines the demographic make-up of the sample to ensure it matches the profile of the whole adult population*"), ale także efektywność tej techniki wyboru okazywała się dotychczas większa niż efektywność technik probabilistycznych ("*Historically in Britain, the record of quota samples in predicting elections has been better than that of random samples*")[7].

Zaobserwowane prawidłowości w próbie nielosowej można odnieść do zbiorowości, której próba ta jest reprezentantem, jednakowoż bez przypisywania im wielkości błędu lub prawdopodobieństw ich prawdziwości. W praktyce ośrodki badawcze stosujące tę grupę technik próbkowania, jeżeli tylko doświadczenia ich są wystarczająco bogate, odwołują się do danych z przeszłości i do wiedzy ekspertów, aby w sposób skwantyfikowany określić błąd, jaki może wynikać z posłużenia się w opisie populacji informacjami z próby (błąd próbkowania).

Weryfikacja hipotez statystycznych

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.

[7] Fakt ten potwierdza także Worcester R.M. [1991], s. 154.

Copyright © StatSoft Polska 2011, info@DaneWiedzaSukceS.pl

Procedura weryfikacji hipotez statystycznych zakłada, że hipoteza zerowa jest dotąd prawdziwa, dopóki nie znajdą się silne przesłanki każące wątpić w jej prawdziwość[8]. 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 najniższe osoby z tej populacji), albo że hipoteza zerowa jest nieprawdziwa[9]. 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 α =0,01, to próba taka zdarzać 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.

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 tożsame[10]. Wyjaśnienie, dlaczego empiryczne badanie **nie może** stanowić podstawy do stwierdzenia prawdziwości testowanej hipotezy, ma podstawy logiczne, statystyczne i filozoficzne.

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

[8] „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].

[9] 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.

[10] 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).

Copyright © StatSoft Polska 2011, info@DaneWiedzaSukceS.pl

obliczenie zarówno wartości statystyki testowej, jak i wyznaczenie prawdopodobieństwa krytycznego (ang. *p-value*)[11]. 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:

- 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;
- rzeczywista użyteczność testowania statystycznego przejawia się przede wszystkim w możliwości stwierdzenia nieprawdziwości sprawdzanej hipotezy (odrzuceniu 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-α) 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

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 ($\overline{X} = 11$), należy na poziomie istotności α = 0,10 zweryfikować hipotezę, że średnia w populacji μ jest równa 10. Hipotezy zerową i alternatywną sformułowano następująco:

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

Copyright © StatSoft Polska 2011, info@DaneWiedzaSukceS.pl

$$H_0 : \mu = 10$$
$$H_1 : \mu \neq 10$$

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

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

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:

$$P\left\{ \overline{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \;\leq\; \mu \;\leq\; \overline{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right\} \;=\; 1-\alpha$$

$$P\left\{ 11 - 1{,}64 \cdot \frac{4}{4} \;\leq\; \mu \;\leq\; 11 + 1{,}64 \cdot \frac{4}{4} \right\} \;=\; 0{,}90$$

$$9{,}36 \;\leq\; \mu \;\leq\; 12{,}64$$

oznacza on, że z prawdopodobieństwem 90% można twierdzić, iż nieznana średnia 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 jakiejkolwiek 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 losowania. 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.

W obu 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. 121)[12]. Oczywiście, obserwacje empiryczne (także statystyczne) mogą potwierdzać pewne hipotezy lub je wspierać, ale znaczenie takiego

[12] 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).

Copyright © StatSoft Polska 2011, info@DaneWiedzaSukceS.pl

faktu jest dla nauki mniejsze niż wykazanie nieprawdziwości (falsyfikacja) hipotezy. W codziennym życiu bywa odmiennie. Staramy się niekiedy poszukiwać faktów lub interpretować je w taki sposób, aby potwierdzały nasze pierwotne 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]. Błąd ten polega na poszukiwaniu i przetwarzaniu tylko tych informacji, które potwierdzają postawioną hipotezę, z jednoczesnym pomniejszaniem znaczenia obserwacji, które hipotezie tej przeczą. W nauce tego typu skłonność prowadzić może do nierzetelnych lub fałszywych wyników. Analogią z zakresu wnioskowania statystycznego byłoby postępowanie, którego celem jest zwiększanie liczebności 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

1. Bazerman M.H., Moore D., [2009], *Judgment in managerial decision making*, (7th edition), J. Wiley.
2. Calaprice, A., [2005], *The new quotable Einstein*, Princeton University Press.
3. Domański Cz., [1990], *Testy statystyczne*, PWE, Warszawa.
4. Domański Cz., K. Pruska, [2000], *Nieklasyczne metody statystyczne*, PWE, Warszawa.
5. Gadomski W., [2009], *Polska w globalnej lawinie*, "Gazeta Wyborcza" z 16.02.2009.
6. Glass D.J., Hall N., [2008], *A brief history of the hypothesis*, "Cell" vol. 134, no. 3.
7. Kmenta J., [1990], *Elements of econometrics*, Macmillan Publishing Company, New York.
8. Mahalanobis P.C., [1951], *Professional Training in Statistics*, "Bulletin of the International Statistical Institute", vol. 33, no. 5.
9. *Ocena metodologii i rezultatów badań poprzedzających pierwszą i drugą turę wyborów prezydenckich 2010 roku*, [2010], Raport Komisji pod przew. prof. H. Domańskiego, Wyd. OFBOR, Warszawa.
10. Paradysz J., [2010], *Konieczność estymacji pośredniej w spisach powszechnych*, [w:], E. Gołata (red.), „Pomiar i informacja w gospodarce", Zeszyty Naukowe nr 149, Uniwersytet Ekonomiczny w Poznaniu, s. 43-66.
11. Popper K.R., [1995], *Mit schematu pojęciowego. W obronie nauki i racjonalności*, Książka i Wiedza, Warszawa.
12. Szreder M., [1994], *Informacje a priori w klasycznej i bayesowskiej estymacji modeli regresji*. Wydawnictwo Uniwersytetu Gdańskiego.

Copyright © StatSoft Polska 2011, info@DaneWiedzaSukceS.pl

13. Szreder M., [2010], *Metody i techniki sondażowych badań opinii*, (wyd. II zmienione), PWE, Warszawa.
14. Worcester R.M., [1991], *British Public Opinion. A Guide to the History and methodology of Political Opinion Polling*. Basil Blackwell, London.

Copyright © StatSoft Polska 2011, info@DaneWiedzaSukceS.pl