Iwona Mróz,

Instytut Fizyki Doświadczalnej,

Uniwersytet Wrocławski

**WNIOSKOWANIE STATYSTYCZNE – PODSTAWY TEORETYCZNE – materiały do wykładów**

Niniejsze materiały mają charakter roboczy. Bardzo proszę o zgłaszanie zauważonych błędów, braków, niedociągnięć i niejasności. Prośba dotyczy też przypisów. Z góry dziękuję za pomoc☺.

**Wnioskowanie statystyczne – wprowadzenie**

Jednym z podstawowych zadań, z którymi mierzy się statystyk jest wyciąganie wniosków dotyczących zbiorowości generalnej na podstawie wiedzy uzyskanej dla zbiorowości próbnej. Wnioskowanie takie zawsze jest obarczone błędem. Trzeba zatem uświadomić sobie, jaka może być natura popełnionych błędów oraz jakie jest prawdopodobieństwo popełnienia błędu. Poniżej przybliżam wstępnie tę problematykę dla metody wnioskowania statystycznego opartego o tzw. testowanie hipotez statystycznych.[[1]](#footnote-1)

Wyobraźmy sobie, że chcemy dowiedzieć się, jaki jest średni wzrost Polaków, mężczyzn między 25 a 60 rokiem życia. Nie jesteśmy w stanie zmierzyć wzrostu wszystkich mężczyzn, zatem badamy zbiorowość próbną, dobraną w odpowiedni sposób (o metodach doboru próby porozmawiamy w przyszłości). Mając dane jedynie dla próby będziemy wnioskować o populacji generalnej. Interesuje nas odpowiedź na pytanie: „Ile wynosi średni wzrost Polaków, mężczyzn między 25 a 60 rokiem życia?” Wnioskowanie statystyczne, będące procedurą matematyczną, wymaga przerobienia tego pytania na zdania twierdzące:

„Średni wzrost Polaków, mężczyzn między 25 a 60 rokiem życia wynosi 1,78 m” (\*)

oraz jedno ze zdań:

„Średni wzrost Polaków, mężczyzn między 25 a 60 rokiem życia nie wynosi 1,78 m” (\*\*)

lub:

„Średni wzrost Polaków, mężczyzn między 25 a 60 rokiem życia jest większy niż 1,78 m” (\*\*\*)

lub

„Średni wzrost Polaków, mężczyzn między 25 a 60 rokiem życia jest mniejszy niż 1,78 m”. (\*\*\*\*).

Postawiliśmy hipotezę statystyczną (\*), którą będziemy sprawdzać. Nazywamy ją hipotezą zerową (H0). Hipotezę H0 możemy w wyniku wnioskowania statystycznego przyjąć (wobec braku podstaw do jej odrzucenia) lub odrzucić. Jeżeli odrzucimy H0, musimy przyjąć inną hipotezę, tzw. hipotezę alternatywną (H1). Zdania (\*\*)-(\*\*\*\*) to hipotezy alternatywne, z których wybieramy tylko jedną i formułujemy ją, tak jak H0, na początku procesu wnioskowania. Sprawdzanie hipotezy H0 nie prowadzi do udowodnienia jej prawdziwości. Może prowadzić wyłącznie do przyjęcia lub odrzucenia H0.

Pamiętajmy, że hipotezy statystyczne to zdania twierdzące.

Podsumowując, „***Hipotezą statystyczną*** nazywamy każdy sąd o zbiorowości generalnej, wydany bez przeprowadzania badania całkowitego. Prawdziwość hipotezy statystycznej orzeka się na podstawie próby losowej.”[[2]](#footnote-2)

Zauważmy, że wnioskujemy o zbiorowości generalnej na podstawie zbiorowości próbnej. Może się zatem zdarzyć tak, że wniosek będzie błędny.

Przyjmijmy, że hipotezą alternatywną H1 jest zdanie (\*\*).

Jeżeli w rzeczywistości średni wzrost interesującej nas grupy Polaków (zbiorowości generalnej) jest równy 1,78 m to H0 jest prawdziwa, a H1 – fałszywa.

Może być też tak, że H0 jest fałszywa, a H1 – prawdziwa, czyli w rzeczywistości średni wzrost nie jest równy 1,78 m.

Widzimy, że w wnioskowanie może prowadzić do czterech sytuacji:

1. H0 jest w rzeczywistości prawdziwa i przyjęliśmy ją.
2. H0 jest w rzeczywistości prawdziwa, a odrzuciliśmy ją i przyjęliśmy hipotezę alternatywną H1 – popełniliśmy BŁĄD.
3. H0 jest w rzeczywistości fałszywa, a przyjęliśmy ją – popełniliśmy BŁĄD.
4. H0 jest w rzeczywistości fałszywa i odrzuciliśmy ją przyjmując H1.

Tylko sytuacje 1 i 4 opisują poprawne decyzje. Sytuacje 2 i 3 ukazują decyzje błędne, ale błędy są różnej natury. Aby to zrozumieć, przyjrzyjmy się przykładowi opisywanemu często w literaturze i przytoczonemu za A. Staniszem[[3]](#footnote-3). Wyobraźmy sobie, że jesteśmy sędziami, którzy mają skazać na karę śmierci człowieka oskarżonego o zabójstwo. W kulturze zachodniej powszechnie uważa się, że lepiej jest uniewinnić winnego niż skazać niewinnego.

Jeżeli H0 i H1 postawimy następująco:

H0: Oskarżony jest niewinny

H1: Oskarżony jest winny.

to „gorszy” błąd popełnimy jeżeli skażemy niewinnego (odrzucimy H0, która jest prawdziwa), niż gdy uniewinnimy winnego (przyjmiemy H0, która jest fałszywa).

Jeżeli w trakcie wnioskowania statystycznego odrzucimy hipotezę zerową, która jest prawdziwa, to popełnimy tzw. **błąd pierwszego rodzaju**. Jeżeli przyjmiemy hipotezę zerową, która w rzeczywistości jest fałszywa, popełnimy tzw. **błąd drugiego rodzaju**.

W rozważanym przykładzie błędem I-go rodzaju będzie skazanie niewinnego, za błędem II-go rodzaju – uniewinnienie winnego.

W praktyce tak formułujemy hipotezy H0 i H1, żeby błąd I-go rodzaju był „gorszym” błędem i żebyśmy, w celu uniknięcia jego popełnienia, musieli się namęczyć, żeby odrzucić hipotezę zerową. Innymi słowy, H0 powinna zostać odrzucona tylko przy mocnych przesłankach. Z tego względu H0 jest często formułowana jako „coś nie działa, nie ma efektu”, podczas gdy badacz życzyłby sobie, żeby efekt był. W dalszej części kursu zobaczymy wiele przykładów poprawnie zdefiniowanych H0. Warto też zapamiętać, że hipoteza zerowa H0 to ta, którą chcemy odrzucić.

Ponieważ błąd I-go rodzaju jest błędem „gorszym”, chcemy, żeby prawdopodobieństwo jego popełnienia było małe. Prawdopodobieństwo popełnienia błędu I-go rodzaju nazywamy **poziomem istotności** i oznaczamy symbolem α. Wnioskowanie prowadzimy na poziomie istotności, którego wartość musi być określona przed przeprowadzeniem wnioskowania i podana jawnie. Manipulowanie poziomem istotności w trakcie badania jest niedopuszczalną praktyką, która prowadzi do fałszerstw naukowych. Brak wartości α dyskwalifikuje wnioskowanie. Poziomy istotności przybierają wartości zależne od branży, której dotyczy wnioskowanie statystyczne. Najczęściej w zastosowaniach ogólnych przyjmuje się , ale w niektórych dziedzinach, np. związane z medycyną, wymaga się wnioskowania na niższym poziomie istotności: , a nawet . Zatem, planując testowanie hipotez statystycznych należy ustalić, jaka wartość poziomu istotności jest wymagana dla danych zastosowań. Przyjęcie zbyt wysokiej wartości stawia pod znakiem zapytania jakość wnioskowania i może sprawić, że wyciągnięte wnioski nie będą traktowane poważnie przez innych badaczy.

Prawdopodobieństwo popełnienia błędu II-go rodzaju oznaczamy zwykle jako β. Jego wartość od wielu czynników takich jak: procedura wnioskowania statystycznego (test statystyczny), wielkość próby oraz poziom istotności (α i β nie są niezależne, zmniejszanie α prowadzi do wzrostu β). Zbyt wysoka wartość β także nie świadczy dobrze o jakości prowadzonego wnioskowania. O prawdopodobieństwie popełnienia błędu II-go rodzaju powiem więcej w przyszłości.

Podsumowanie powyższych rozważań przedstawia Tabela 4.2

Tabela 4.2

|  |  |  |
| --- | --- | --- |
|  | Decyzja | |
| Przyjąć H0 | Odrzucić H0 |
| H0 prawdziwa | Decyzja prawidłowa | Błąd I-go rodzaju |
| H0 fałszywa | Błąd II-go rodzaju | Decyzja prawidłowa |

*Źródło:* Stanisz, A., Przystępny kurs statystyki z zastosowaniem STATISTICA.PL na przykładach z medycyny, StatSoft, Kraków 2006, t.1, s. 199.

**Wnioskowanie statystyczne – ciąg dalszy nr 1[[4]](#footnote-4)**

Przypomnijmy, że przez hipotezę statystyczną rozumieć będziemy każde przypuszczenie dotyczące zbiorowości generalnej wysunięte bez przeprowadzania badania całkowitego. Prawdziwość hipotezy będziemy weryfikować na podstawie wyników próby losowej.[[5]](#footnote-5)

Przywołując A. Stanisza, hipotezy statystyczne dzielimy na:

…”- parametryczne, gdy dotyczą wartości parametrów statystycznych populacji (np. średnia, wariancja),

- *nieparametryczne*, gdy dotyczą postaci rozkładu cech lub losowości próby.”[[6]](#footnote-6)

Hipotezę zerową parametryczną możemy przykładowo sformułować jako:

„Wartość przeciętna w populacji wynosi 50” lub „Wartości przeciętne w dwóch porównywanych populacjach są równe” lub „Wariancje w porównywanych dwóch populacjach są równe”.

Hipoteza zerowa nieparametryczna może mieć postać np.: „Rozkład badanej cechy statystycznej jest zgodny z rozkładem normalnym” lub „Dystrybuanty badanej cechy statystycznej w dwóch porównywanych populacjach są równe”.

Przypomnijmy, że hipotezy zerowe są, z powodów omówionych na wykładzie nr 10, tak formułowane, jakby nie było efektu, którego wystąpienie chcemy uprawdopodobnić.

Test statystyczny to pewien schemat postępowania pozwalający na odrzucenie hipotezy zerowej na określonym wcześniej (!!!) poziomie istotności lub pozwalający na jej przyjęcie na tym poziomie istotności, jeżeli nie ma podstaw do odrzucenia. Pamiętajmy, że poziomu istotności NIE MODYFIKUJEMY podczas wnioskowania!!! Wprowadzanie na tym etapie zmian przyjętej wartości poziomu istotności jest złą praktyką i może prowadzić do fałszerstwa naukowego.

W typowym teście statystycznym znajdziemy:

1. Założenia, które powinny być spełnione, aby wnioskowanie statystyczne przebiegło poprawnie, i żebyśmy mogli zaufać uzyskanym wynikom. W praktyce wielokrotnie zdarza się, że niewielkie odstępstwa od założeń jeszcze pozwalają na stosowanie testu, Sytuacje takie są znane statystykom i opisane w literaturze. Decyzja, czy zastosować test jeśli założenia nie są ściśle spełnione wymaga wiedzy i doświadczenia. Założenia mogą dotyczyć np. liczebności próby, skali pomiarowej danych, normalności rozkładu badanej cechy w populacji, z której została pobrana próba, równości wariancji badanej cechy w porównywanych populacjach i innych. Zasadniczo, więcej założeń znajdziemy w testach parametrycznych niż nieparametrycznych, ale testy parametryczne są uznawane za „lepsze” od nieparametrycznych. Dlaczego tak jest wyjaśnię Państwu później.

2. Schematy do formułowania hipotez statystycznych: zerowej i alternatywnej, pod które podłożymy treści merytoryczne podlegające badaniu.

3. Wyrażenie (wzór) definiujące zmienną losową nazywaną sprawdzianem hipotezy lub statystyką. Do wzoru podstawiamy dane z próby, którą dysponujemy. Obliczona wartość sprawdzianu hipotezy jest zatem wartością charakteryzującą konkretną próbę. Dla innej próby pobranej z tej samej populacji generalnej dostaniemy inną wartość.

4. Określenie wskazujące jaki powinien być rozkład statystyki opisanej w p. 3 jeżeli hipoteza zerowa jest prawdziwa. Może ono przybrać postać np.: „Jeżeli hipoteza zerowa jest prawdziwa to statystyka określona wzorem … (tu wzór z p. 3) ma rozkład … (tu podany rozkład, np. N(0,1)). Często spotykane rozkłady sprawdzianów hipotez to standardowy rozkład normalny, rozkład Studenta o liczbie stopni swobody zależnej od liczebności próby, rozkład chi-kwadrat i rozkład F Snedecora (o liczbach stopni swobody zależnych od liczebności prób). Istnieje wiele testów, dla których sprawdziany hipotez mają rozkłady inne od wymienionych.

Niejednokrotnie dane hipotezy możemy sprawdzać przy użyciu wielu różnych testów. Staramy się wtedy wybrać tzw. test najmocniejszy, czyli taki, który przy przyjętym poziomie istotności (czyli prawdopodobieństwie popełnienia błędu I-go rodzaju) daje najmniejsze prawdopodobieństwo popełnienia błędu drugiego rodzaju.[[7]](#footnote-7) Do pojęcia mocy testu wrócimy później.

Idea wnioskowania statystycznego przy pomocy testu statystycznego jest następująca[[8]](#footnote-8): Zakładamy, że hipoteza zerowa jest prawdziwa. Mamy zdefiniowana statystykę (zmienną losową) Z. Przy założeniu prawdziwości hipotezy zerowej Z podlega pewnemu, podanemu w opisie testu, rozkładowi, np. rozkładowi normalnemu N(0,1). Obliczamy wartość statystki Z dla naszej próby i patrzymy, czy dla założonego rozkładu Z uzyskaną wartość możemy uznać za „wystarczająco prawdopodobną”. W przypadku rozkładu N(0,1) prawdopodobieństwo, że statystyka Z przyjmie wartości dalekie od średniej jest małe. Jednak sytuacja taka jest możliwa i można ją wytłumaczyć na dwa sposoby. Po pierwsze, mamy wyjątkowego pecha lub źle dobraliśmy próbę. Po drugie: hipoteza zerowa NIE JEST PRAWDZIWA, zatem statystyka Z nie podlega założonemu rozkładowi. Przy wnioskowaniu statystycznym przyjmujemy, że jeżeli wartość statystyki Z wyliczonej z próby przyjmie wartość uznaną przez nas za mało prawdopodobną, to hipoteza zerowa nie jest prawdziwa i należy ją odrzucić.

To, czy obliczoną wartość statystyki uznamy za mało prawdopodobną zależy od przyjętej wartości poziomu istotności. Przypomnijmy, że poziom istotności to prawdopodobieństwo popełnienia błędu pierwszego rodzaju, czyli odrzucenia hipotezy zerowej w sytuacji, gdy jest ona prawdziwa (na wykładzie nr 4 omawialiśmy błąd pierwszego rodzaju polegający na skazaniu na śmierć człowieka, który w rzeczywistości jest niewinny). Zatem, jeżeli hipoteza zerowa jest prawdziwa, czyli statystyka Z podlega danemu rozkładowi, a my odrzucimy H0, to popełnimy błąd pierwszego rodzaju. Przyjmujemy teraz, że hipotezę zerową odrzucamy jeżeli wartość statystyki Z znajdzie się w przedziale (lub dwóch przedziałach, zob. dalej) ”wartości nietypowych” (np. dla rozkładu N(0,1) leżących daleko od średniej, na „ogonach” rozkładu), dla których całkowite pole pod funkcją gęstości rozkładu statystyki Z jest równe poziomowi istotności. Taka konstrukcja pozwala nam przyjąć, że jeżeli hipoteza zerowa jest prawdziwa, to prawdopodobieństwo jej odrzucenia jest równe poziomowi istotności. Podkreślmy, że wartość poziomu istotności musi być przyjęta wcześniej, jeszcze na etapie projektowania eksperymentu. Bez znajomości poziomu istotności nie jesteśmy w stanie określić, czy obliczoną wartość statystyki Z uznamy za „typową” czy „nietypową”, a zatem nie jesteśmy w stanie podjąć decyzji o odrzuceniu H0.

Przedziały wartości statystyki Z, dla których odrzucamy hipotezę zerową nazywamy zbiorem krytycznym, a wartości odcinające zbiory krytyczne to wartości krytyczne statystyki Z.

Definicja:

„Zbiorem krytycznym (…) nazywamy zbiór tych wartości sprawdzianu hipotezy, które przemawiają za odrzuceniem hipotezy H0.”[[9]](#footnote-9)

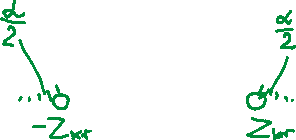
To, czy zbiór krytyczny będzie się składał z jednego, czy z dwóch przedziałów wartości sprawdzianu hipotezy zależy od postaci hipotezy alternatywnej H1 (czy jest to tzw. hipoteza jednostronna czy dwustronna). Do zagadnienia wrócimy później. W poniższym przykładzie przyjmijmy, że wartości statystyki Z przemawiające za odrzuceniem H0 możemy zaobserwować na obu „ogonach” rozkładu (H1 jest tzw. hipotezą dwustronną).

Przykład:

Niech sprawdzian hipotezy Z podlega standardowemu rozkładowi normalnemu przy założeniu, że H0 jest prawdziwa (Rys. 5.1).

Dla przyjętego z góry poziomu istotności α określamy przedziały na „ogonach” rozkładu, dla których sumaryczne pole pod krzywą jest równe poziomowi istotności. Zatem, na każdym „ogonie” odcinamy przedział, dla którego pole pod krzywą wynosi α/2. Przedziały te stanowią zbiór krytyczny, a wartości statystyki, które je odcinają to wartości krytyczne statystyki.





Rys. 5.1. Zbiór krytyczny dla hipotezy dwustronnej. Na osi odciętych mamy wartości statystyki testowej (sprawdzianu hipotezy), a na osi rzędnych – wartości funkcji gęstości prawdopodobieństwa rozkładu tej statystyki. Półproste na osi odciętych na prawo i na lewo od niebieskich, pionowych linii wyznaczają zbiór krytyczny dla hipotezy alternatywnej dwustronnej (jest podzielony w tym przypadku podzielony na dwa podzbiory). Pole pod krzywą obliczone dla CAŁEGO zbioru krytycznego jest równe pozimowi istotności α. Wartości sprawdzianu hipotezy, które odcinają zbiór krytyczny nazywamy wartościami krytycznymi statystyki. Dla analizowanego przykładu są to oznaczone na rysunku kółkami wartości -Zkr i Zkr (UWAGA! Rysunek 5.1 to poglądowa ilustracja, indeksy „kr” nie są typowe, później je zmienimy dostosowując do konkretnej sytuacji badawczej). W przypadku hipotez alternatywnych jednostronnych zbiór krytyczny budujemy „w jednym kawałku” i mamy jedną wartość krytyczną statystyki. *Źródło: opracowanie własne na podstawie J.E. Freund, B.M. Perles, Modern Elementary Statistics, Twelfth Edition,, s. 309.*



Podkreślmy, że na podstawie testowania hipotez nie jesteśmy w stanie udowodnić, że dana hipoteza H0 jest prawdziwa lub fałszywa. Możemy jedynie odrzucić ją lub nie mieć podstaw do jej odrzucenia. Ponadto, H0 odrzucamy lub nie mamy podstaw do jej odrzucenia dla przyjętego poziomu istotności (dla innego poziomu istotności wniosek końcowy może być odmienny). Zatem, formułując wniosek końcowy ZAWSZE podajemy poziom istotności i używamy słów „odrzucam” lub „nie mam podstaw do odrzucenia H0”, np.:

„Na poziomie istotności 0.05 nie mam podstaw do odrzucenia H0” (poprawna formuła) zamiast „H0 jest prawdziwa” (błędna formuła).

„Na poziomie istotności 0,05 odrzucam H0” (poprawna formuła) zamiast „H0 jest fałszywa” (błędna formuła).

Temat będzie kontynuowany☺.

1. Testowanie hipotez statystycznych omówiono we wszystkich podręcznikach polecanych do wykładu, dlatego szczegółowe przypisy podaję tylko w przypadku dosłownego cytowania książek lub przytaczania przykładów. [↑](#footnote-ref-1)
2. Ostasiewicz, S., …, op. cit. s. 235. [↑](#footnote-ref-2)
3. Stanisz, A., Przystępny kurs statystyki z zastosowaniem STSTISTICA PL na przykładach z medycyny, StatSoft, Kraków 2006, t.1, s. 198-199. [↑](#footnote-ref-3)
4. Informacje zawarte w niniejszym paragrafie można znaleźć w wielu książkach. Wykorzystałam głównie podręczniki: A. Stanisza, A. Aczela, S. Ostasiewicz i wsp. oraz Freunda i Perlesa. Szczegółowe przypisy zamieszczam jeśli przytaczam definicje lub cytuję autorów dosłownie lub prawie dosłownie. [↑](#footnote-ref-4)
5. Ostasiewicz, S, Rusnak, Z., Siedlecka U., Statystyka. Elementy teorii i zadania, Wydawnictwo Akademii Ekonomicznej im Oskara Langego we Wrocławiu, Wrocław 1999, s 235. [↑](#footnote-ref-5)
6. Stanisz, A., Przystępny kurs statystyki z zastosowaniem STATISTICA.PL na przykładach z medycyny, StatSoft Polska sp. z.o.o., Kraków 2006, Tom 1. Statystyki podstawowe, s. 197. [↑](#footnote-ref-6)
7. Ibidem, s. 201. [↑](#footnote-ref-7)
8. Por. np. Aczel A.D., Statystyka w zarządzaniu, Wydawnictwo Naukowe PWN, Warszawa 2000, s. 271-277 i źródła wyszczególnione w przypisie 1. [↑](#footnote-ref-8)
9. Ostasiewicz, S. i wsp., op. cit. s. 236. [↑](#footnote-ref-9)