

Interpretacja wyników analizy statystycznej danych

Interpretation of the statistical analysis of data

Łukasz J. Krzych

Zakład Kardioanestezji i Intensywnej Terapii Pooperacyjnej, SP Szpital Kliniczny ŚUM nr 7, Katowice

Kardiochirurgia i Torakochirurgia Polska 2007; 4 (3): 315–321



Streszczenie

Jakość publikowanych doniesień naukowych jest często poruszonym problemem ze względu na wiarygodność i użyteczność przedstawianych w nich danych. Metodologia badań naukowych pozostaje niedocenianym i często bagatelizowanym punktem w prowadzeniu medycznych prac naukowych, wpływającym na zasady dobrej praktyki publikacyjnej. W prezentowanym artykule zwrócono uwagę na często występujące problemy związane z interpretacją uzyskanych wyników oraz zilustrowano schemat podstawowego postępowania statystycznego, które powinno towarzyszyć analizie otrzymanych danych. Uwzględniono ponadto problem doboru uczestników do badania oraz metod kontroli czynników potencjalnie zakłócających wyniki.

Słowa kluczowe: medycyna oparta na dowodach naukowych, analiza statystyczna danych, naukowe badanie medyczne.

Abstract

The problem of the quality of published scientific researches is often criticized because of the questionable reliability and utility of the data they contain. Research methodology is still an underestimated and neglected issue and has an impact on principles of good publication practice. This paper describes frequent difficulties in data interpretation and illustrates the basic scheme of statistical analysis of data that should be taken under consideration by researchers. Proper choice of group of patients and statistical methods to control potential confounders are also discussed in this article.

Key words: evidence-based medicine, statistical analysis, medical research.

Wstęp

Medycyna oparta na dowodach *stricte* naukowych (EBM – *Evidence Based Medicine*) ma ponad 10-letnią historię [1]. Problem jakości publikowanych doniesień naukowych jest często poruszany ze względu na wiarygodność i użyteczność przedstawianych w nich danych. W 1994 r. Douglas Altman, redaktor *British Medical Journal*, zwrócił uwagę na błędy często występujące w pracach publikowanych na łamach najbardziej poczytnych czasopism [2]. Najliczniejsze zastrzeżenia dotyczyły metodologicznego opracowania zagadnienia. Do dziś metodologia badań naukowych pozostaje niedocenianym i często bagatelizowanym punktem w prowadzeniu prac naukowych, wpływającym na zasady dobrej praktyki publikacyjnej [3–5]. W 2000 r. na łamach *Science Editor* ukazał się artykuł, w którym podsumowano opinie redaktorów i recenzentów renomowanych, recenzowanych czasopism medycznych [6]. Wyniki tego przeglądu pokazały, że najczęstszą przyczyną odmowy publikacji artykułów była obecność błędów metodologicznych (36% negatywnych opinii), na-

stępnie nieadekwatne wyniki (25%), niewłaściwe metody statystyczne (11%) oraz niedostatki w edycji i prezentacji problemu (11%). Podsumowując, częścią artykułu, która decydowała o negatywnej opinii recenzentów, była sekcja *Materiał i metody* (52% przypadków), natomiast żadna praca nie została odrzucona ze względu na niepoprawny *Wstęp*.

W obliczu szybko postępującego rozwoju różnych dziedzin medycyny, chęci, a nierzadko potrzeby, publikacji własnych osiągnięć naukowych zagadnienie to staje się szczególnie istotne. Badacz powinien znaleźć optymalne rozwiązanie między zachowaniem pełnej metodologicznej poprawności a umiarkowaniem w epatowaniu czytelnika wszystkimi informacjami płynącymi z prowadzonych badań (sumarycznie często dziesiątki liczb, tabel czy rycin). Celem prezentowanej pracy jest zwrócenie uwagi na często występujące problemy związane z metodologiczną interpretacją uzyskanych wyników, a przede wszystkim zobrazowanie podstawowego postępowania statystycznego, które powinno towarzyszyć analizie otrzymanych danych.

Adres do korespondencji: dr n. med. Łukasz J. Krzych, Zakład Kardioanestezji i Intensywnej Terapii Pooperacyjnej ŚUM, ul. Ziołowa 45/47, 40-635 Katowice, tel. +48 32 359 88 99, faks +48 32 202 64 64, e-mail: l.krzych@wp.pl

Etapy statystycznej analizy danych

Celem prowadzonych badań – nie tylko w medycynie – jest stworzenie możliwości odnoszenia otrzymanych wyników do całej badanej populacji, czyli możliwości uogólniania rezultatów. Jednak przebadanie wszystkich elementów wybranego zbioru (całej populacji) rzadko jest możliwe. Z tego powodu często wybiera się jedynie tak zwaną próbę, stanowiącą materiał badawczy, nazywaną grupą badaną czy uczestnikami badania. Każda próba musi być reprezentatywna, co w zasadniczym stopniu uwarunkowane jest prawidłowym doбором uczestników pod względem ilościowym (tu: minimalna niezbędna liczebność próby) i jakościowym (tu: arbitralny lub losowy/randomizowany sposób doboru uczestników) [7]. Minimalna niezbędna liczebność próby uzależniona jest od liczebności populacji, do której mają być odnoszone wyniki, przewidywanej częstości występowania lub średniej wartości ocenianej zmiennej oraz zakładanego *a priori* poziomu istotności statystycznej (zwykle 0,05) i mocy zastosowanego testu (zwykle 0,8) [7, 8]. Dobór minimalnej liczebności próby zależy w znacznej mierze od epidemiologicznego typu prowadzonego badania (np. przekrojowe, kliniczno-referencyjne, kohortowe) [8]. Randomizacja, czyli losowy dobór uczestników, może uwzględniać schemat doboru prostego, systematycznego, warstwowego lub zespołowego, przy czym właściwy sposób doboru w znacznej mierze podyktowany jest typem badania (np. arbitralny dobór kolejnych przypadków w badaniu kliniczno-referencyjnym, zespołowy przy ocenie dużych zbiorowości w badaniach przekrojowych) [7, 9]. Reprezentatywność grupy badanej przekłada się na wiarygodność otrzymanych wyników, dlatego stadium projektowania badania warunkuje potencjalny sukces dalszych jego etapów [10].

Zadaniem statystyki jest badanie własności dużych zbiorów elementów, czyli wspomnianej populacji. Ze statystycznego punktu widzenia nie musi jej stanowić grupa osób czy grupa zwierząt, ale także zbiór dokumentacji medycznej (np. wypisy szpitalne, szpitalne bazy danych, historie chorób) czy dokumenty demograficzne (np. świadectwa urodzeń lub zgonów).

Każda analiza statystyczna powinna rozpoczynać się od zdefiniowania typów cech (własności elementów), które będą badane, w statystyce nazywanych zmiennymi. Wyróżnia się dwa typy zmiennych: jakościowe oraz ilościowe [11]. Zmienne jakościowe świadczą o „jakości” próby, nie posiadają jednostki pomiaru, są niemierzalne (np. płeć, obecność choroby, miejsce zamieszkania, stopień zaawansowania choroby). Cechy te są zazwyczaj wyrażane liczbą sprzyjających obserwacji (np. liczba mężczyzn i kobiet w grupie badanej) za pomocą wartości bezwzględnej (np. 20 mężczyzn i 30 kobiet), liczby względnej (np. 20 mężczyzn na 50 badanych, czyli 20/50 mężczyzn) oraz odsetka (np. 40% mężczyzn i 60% kobiet). Ważne jest ponadto zwrócenie uwagi, czy zmienna jest nominalną, czyli bez zachowanej gradacji (np. płeć, obecność choroby, grupa krwi) czy porządkową: z wyróżnialnymi stopniami, tak jak np. stopień zaawansowania nowotworu. Ponadto dla ułatwienia analizy statystycznej danych, „ze względów taktycznych”, bardzo często – choć zwykle

niepotrzebnie [12] – wykorzystywany jest przez badaczy tzw. podział dychotomiczny zmiennej jakościowej (na dwie kategorie). Podział taki jest oczywisty dla tych zmiennych, które posiadają tylko dwie kategorie (płeć: mężczyzna/kobieta, obecność choroby/brak choroby). W przypadku zmiennych o większej liczbie kategorii należy zadać konkretne pytanie badawcze, na które uda się udzielić dwóch odpowiedzi: tak lub nie (np. Czy analizowana grupa krwi to A?). Drugi typ zmiennych to zmienne ilościowe, czyli te, które posiadają swoje jednostki pomiarowe. Wśród zmiennych ilościowych wyróżnia się z kolei zmienne ciągłe, czyli dające się wyrazić z coraz większą dokładnością, z większą precyzją (używając coraz to mniejszych jednostek pomiarowych, np. masa ciała, stężenie cholesterolu) oraz zmienne skokowe, których „nie da się podzielić” (wyrażane w sztukach, np. ilość osób w sali, ilość komórek w polu widzenia). Aby wyrazić zmienną ilościową w sposób dychotomiczny, należy przekształcić ją w zmienną nominalną, używając tak zwanych punktów odcięcia [np. normoglikemia (glikemia <100 mg/dl) oraz podwyższone stężenie glukozy (glikemia ≥100 mg/dl)].

Z punktu widzenia medyka-badacza, analiza statystyczna danych, czyli popularnie nazywana „statystyka” ogranicza się do wprowadzenia do bazy danych komputera wielu informacji uzyskanych w toku badania i poszukiwania albo różnic, albo zależności między grupami (zmiennymi). Aby jednak nie pogubić się we własnych dociekaniach, na początku należy postawić sensowne pytanie badawcze, na które dane cyfry mają ułatwić odpowiedź. Pytanie to – zawarte w celach pracy – powinno być precyzyjne, czytelne oraz umożliwiające jednoznaczną odpowiedź. Pytanie to może dotyczyć opisu zaobserwowanych zjawisk (tu: epidemiologiczne badanie opisowe, szukające odpowiedzi na pytania: kto? gdzie? kiedy?), różnicy bądź zależności między grupami, a na samym końcu związku przyczynowo-skutkowego (tu: epidemiologiczne badanie analityczne). Rozwiązanie problemu badawczego często zależy od logicznej gradacji zadawanych kolejno pytań.

Pierwszym krokiem jakiegokolwiek statystycznej analizy danych w tak zwanej perspektywie statystycznej jest podanie zwięzłej charakterystyki badanych cech z wykorzystaniem właściwych technik (statystyka opisowa; patrz niżej). Etap drugi (statystyka analityczna) uwzględnia opis różnic lub zależności, odpowiednio do celu pracy. Krok drugi nie może jednak zostać osiągnięty bez realizacji kroku pierwszego! Gdy oba etapy są wypełnione, można wejść na kolejny stopień. Trzeci etap to odpowiedź na pytanie dotyczące związku przyczynowo-skutkowego, co jest możliwe jedynie wtedy, gdy uwzględniony zostanie wpływ czynników zakłócających na wcześniejsze rezultaty (zwłaszcza dotyczy to oceny różnic i zależności w drugim etapie analizy). Kontrolę czynników zakłócających można osiągnąć albo na etapie projektowania badania (poprzez randomizację, ograniczenie zasięgu badania czy parowanie), albo właśnie na etapie analizy statystycznej danych (z wykorzystaniem analizy stratyfikacyjnej i analizy wielu zmiennych) [8–10, 12]. Nie można zapominać, że to treść pytania badawczego inicjuje i warunkuje sposób wnioskowania statystycznego!

Statystyka opisowa

Jak wiadomo, na wnioskowanie statystyczne dotyczące natury zjawiska badanej populacji składają się statystyka opisowa oraz statystyka analityczna.

Statystyka opisowa podaje charakterystykę badanych zmiennych z wykorzystaniem właściwych sobie miar [7, 11, 14, 15]. Dla zmiennych jakościowych jest to, jak wspomniano, częstość sprzyjających obserwacji (liczba bezwzględna, względna, procent). W tym miejscu należy przytoczyć opinię, iż dla liczebności grup mniejszych niż 30 osób powinno się raczej używać liczby względnej lub bezwzględnej zamiast podawania wartości odsetkowej. Dla zmiennych ilościowych korzystne jest podanie miar skupienia (np. wartość największa i najmniejsza, średnia arytmetyczna, moda, centyle) oraz miar rozproszenia (tj. rozstęp, rozstęp kwartylowy, odchylenie standardowe, błąd standardowy, wariancja i współczynnik zmienności), czasem też miar asymetrii (skośność rozkładu) i koncentracji (kurtoza). Sposób prezentacji danych jest uzależniony od dwóch czynników. Po pierwsze, decyduje o tym autor pracy. Po drugie, przyjęta jest konwencja, że dla opisu bardziej liczebnych grup (oraz dla zmiennych spełniających kryteria rozkładu normalnego) zwykle się używać średniej arytmetycznej i odchylenia standardowego (lub błędu standardowego). W przypadku badanych grup o mniejszej liczebności i (co za tym często podąża) dla zmiennych odbiegających od rozkładu normalnego powinno używać się mediany lub mody oraz rozstępu kwartylowego, ponieważ one lepiej opisują badaną grupę [15, 16].

Statystyka analityczna – analizy proste

Kolejnym etapem, po uwzględnieniu w analizie danych statystyki opisowej, jest prosta analiza różnic bądź zależności, co tożsame jest z testowaniem stawianych hipotez badawczych (korespondujących z celami pracy). Testowanie hipotez służy ocenie, czy zaobserwowane różnice bądź zależności są raczej efektem błędu lub przypadku czy prawdziwą zależnością/różnicą między badanymi populacjami [7, 11]. Do celów weryfikacji zawsze stawiana jest hipoteza zerowa, która brana jest niejako „pod obrzuteł krytyki”. Weryfikowana statystycznie hipoteza zerowa mówi zazwy-

czaj, że obserwowana różnica/zależność jest dziełem przypadku, skutkiem błędu, czyli testowane próby pochodzą z tej samej populacji (między grupami nie ma statystycznej różnicy bądź zależności). Druga stawiana hipoteza, nazywana alternatywną, jest natomiast przeciwnością hipotezy zerowej (zależności/różnice nie są dziełem przypadku ze statystycznego punktu widzenia).

Po postawieniu hipotezy zerowej przed badaczem staje problem doboru odpowiedniego testu statystycznego (matematycznej formuły), który pozwoli odpowiedzieć na postawione pytanie. Interpretacji wyników obserwacji dokonuje się na podstawie wartości prawdopodobieństwa (p) dla zastosowanego testu, choć należy z całą stanowczością zaznaczyć, iż wartość p nie powinna bezkrytycznie determinować dalszego wnioskowania. Równie ważne jest dociekanie na temat wielkości samego efektu, który nie zawsze musi być statystycznie znamienny [17, 18].

Zwykło się przyjmować, że jeżeli p jest „wystarczająco małe”, to jest mało prawdopodobne, że obserwacja wynika z przypadku (należy odrzucić hipotezę zerową i przyjąć hipotezę alternatywną). Jeżeli p jest „wystarczająco duże”, to jest bardzo możliwe, że zaobserwowane zależności są skutkiem przypadku (brak wtedy podstaw do odrzucenia hipotezy zerowej). Próg między p małym a p dużym (prawdą statystyczną a fałszem) zakładany jest na początku badania (*a priori*) i oznaczany jako wspomniany wcześniej poziom istotności statystycznej (alfa) [7, 11]. Najczęściej przyjmuje się, że alfa = 0,05, natomiast bardziej wymagający badacze mogą przyjąć alfa = 0,01 lub alfa = 0,001 (rozumiejąc przez to, że mogą popełnić mniejszy błąd we wnioskowaniu). Co więcej, w testowaniu hipotez, w analizie wielu zmiennych (patrz niżej) bardzo często przyjmuje się alfa = 0,1 (a wartość pomiędzy alfa = 0,05 a alfa = 0,1 nazywana jest graniczną znamiennością statystyczną). Dobór wartości poziomu istotności statystycznej również jest konwencją, dlatego tak naprawdę jego wartość zależy od badacza. Interpretując wynik analizy, powinno się pamiętać, iż jest on odzwierciedleniem jedynie związków statystycznych (znamienność statystyczna różnic bądź zależności), co niekoniecznie przekłada się na prawdę biologiczną. Interpretację wyników wybranych testów statystycznych, opisanych poniżej, przedstawiono w tabeli I.

Tab. I. Interpretacja wyników wybranych testów statystycznych

Typ zmiennej	Nazwa testu	Wynik testu*	Interpretacja wyniku testu
jakościowa	chi ² (także chi ² z poprawką Yatesa oraz dokładny test Fishera)	p<0,05	znamienne statystycznie różnice między częstościami w porównywanych grupach
		p≥0,05	zaobserwowane różnice są nieznamienne statystycznie
ilościowa	Shapiro-Wilka (także Smirowa-Kolmogorowa)	p<0,05	rozkład badanej zmiennej odbiega od rozkładu normalnego (<i>zastosuj test nieparametryczny</i>)
		p≥0,05	rozkład zmiennej jest normalny (<i>zastosuj test parametryczny</i>)
	Levena	p<0,05	wariancje nie są jednorodne (<i>zastosuj test Cochrańa-Cox</i>)
		p≥0,05	wariancje są jednorodne (<i>zastosuj test t-Studenta lub analizę wariancji</i>)
	t-Studenta (także U Manna-Whitneya)	p<0,05	zmienna ma znamienne statystycznie wyższą wartość w jednej grupie w porównaniu z drugą grupą
		p≥0,05	różnice między wartościami w grupach są nieznamienne statystycznie
analiza wariancji (także Kruskala-Wallisa)	p<0,05	różnice między wartościami w grupach są znamienne statystycznie	
	p≥0,05	różnice między wartościami w grupach są nieznamienne statystycznie	

*gdy poziom istotności statystycznej alfa = 0,05

Analiza statystyczna różnic między grupami dla zmiennych ilościowych

Pierwszym krokiem tego etapu analizy jest ocena, czy zmienne są ze sobą powiązane (czy są od siebie zależne) [11]. W przypadku oceny różnic między zmiennymi niepowiązanymi [np. dystrybucji ciśnienia tętniczego (zmienna zależna) w zależności od płci badanych (zmienna niezależna/grupująca)], powinno się ocenić, czy rozkład analizowanej zmiennej jest normalny (Gaussowski) czy też odbiega od normalnego [11, 19]. Najpopularniejszymi i najbardziej rzetelnymi testami wykorzystywanymi do tego celu są testy Shapiro-Wilka oraz Smirnowa-Kołmogorowa [11, 19].

Jeżeli rozkład zmiennej ilościowej odbiega od normalnego ($p < 0,05$ dla każdego z wymienionych dwóch testów), do oceny różnic należy zastosować test nieparametryczny (test U Manna-Whitneya lub test Kruskala-Wallisa), przy czym do oceny różnic między dwiema grupami służy test U Manna-Whitneya [np. ocena różnic między wartością ciśnienia tętniczego (zmienna zależna) w zależności od płci (zmienna niezależna)], natomiast dla większej liczby grup [np. ocena różnic w wartości ciśnienia tętniczego (zmienna zależna) między mieszkańcami czterech miast (zmienna niezależna)] – test Kruskala-Wallisa [11, 19].

Jeżeli rozkład analizowanej zmiennej ilościowej jest normalny ($p > 0,05$ dla testu Shapiro-Wilka), należy zbadać, czy wariancje w badanych grupach są jednorodne, np. za pomocą testu Levena [19]. Jeżeli wariancje są niejednorodne (wynik testu Levena: $p < 0,05$), do oceny różnic między grupami służy test Cochran-Cox. W przeciwnym przypadku (wynik testu Levena: $p > 0,05$) należy rozważyć zastosowanie testu t -Studenta lub analizy wariancji (ANOVA) [11, 19]. Test t -Studenta jest rodzajem analizy wariancji dla dwóch grup [np. ocena różnic pomiędzy wartością ciśnienia tętniczego (zmienna zależna) w zależności od płci (zmienna niezależna)], natomiast analiza wariancji służy do oceny różnic występujących między co najmniej trzema grupami [np. ocena różnic w wartości ciśnienia tętniczego (zmienna zależna) między mieszkańcami czterech miast (zmienna niezależna)]. Co więcej, znamienny statystycznie wynik testu Kruskala-Wallisa czy ANOVA ($p < 0,05$) tak naprawdę nie definiuje, między którymi grupami występuje istotność statystyczna. Aby rozwikłać ten problem, należy wykorzystać tzw. analizę *post hoc* z wykorzystaniem testów Tukeya czy najmniejszych istotnych różnic (NIR), które – jak test t – testują zmienne parami [11, 19].

W ocenie różnic międzygrupowych dla zmiennych ilościowych powiązanych [np. wartość ciśnienia tętniczego (zmienna zależna) w grupie badanej przed leczeniem i po leczeniu (zmienna niezależna)] wykorzystywany jest test t dla grup zależnych (gdy rozkład zmiennej jest normalny) lub test Wilcoxon (gdy rozkład zmiennej odbiega od normalnego) [11, 19].

Analiza statystyczna różnic między grupami dla zmiennych jakościowych

W ocenie różnic międzygrupowych dla zmiennych jakościowych niepowiązanych wykorzystywany jest test χ^2 (z analizą ilorazu szans) [11, 19, 20]. Jednak dla mało liczeb-

nych grup zmiennych test ten może dawać mylne wyniki. W celu ich weryfikacji stosowane są modyfikacje testu χ^2 : test χ^2 z poprawką Yatesa oraz tzw. dokładny test Fischera. Gdy występuje mniej niż 40 obserwacji lub gdy w dowolnym polu w tabeli czteropolowej występuje mniej niż pięć obserwacji, wskazane jest zastosowanie poprawki Yatesa albo dokładnego testu Fischera. Natomiast gdy grupa badana liczy mniej niż 20 osób, z całą pewnością powinno się stosować dokładny test Fischera [11, 19].

Kiedy porównywane są zmienne jakościowe powiązane (np. procent sukcesu przeprowadzonej terapii hipotensyjnej), zastosowanie znajduje test McNemary, który również jest pewną modyfikacją testu χ^2 [11, 19].

Wynik testu χ^2 wskazujący na znamienność statystyczną różnic w częstości cech między grupami ($p < 0,05$) w tabeli wielopolowej również nie mówi, między którymi parami zmiennych występują różnice istotne statystycznie. Aby określić, między którymi parami zmiennych znajdują się wskazane różnice, należy dla każdej takiej pary zmiennych przeprowadzić osobny test χ^2 , czyli ułożyć oddzielną tabelę czteropolową. Alternatywą jest również zastosowanie tzw. poprawki Bonferroniego, która uwzględnia ilość testowanych hipotez zerowych (liczba tabeli czteropolowych, które można skonstruować na podstawie tabeli wielopolowej) [21]. W tym przypadku poziom istotności statystycznej musi zostać skorygowany o liczbę hipotez zerowych (jeżeli na podstawie tabeli można sformułować osiem hipotez zerowych, to „skorygowana” alfa wynosi $0,05/8 = 0,006$).

Analiza statystyczna zależności między grupami

Pierwszym krokiem na tym etapie analizy jest ocena rodzaju badanej zmiennej zależnej i niezależnej (ciągła, skokowa/nominalna, porządkowa). Gdy istnieje potrzeba oceny zależności między dwiema zmiennymi ciągłymi [np. zależność między stężeniem cholesterolu (zmienna zależna) a wskaźnikiem masy ciała (zmienna niezależna)], zastosowanie znajduje korelacja liniowa z wykorzystaniem współczynnika korelacji Pearsona (r) [11, 19]. W przypadku interpretacji graficznej korelacji liniowej w układzie współrzędnych na płaszczyźnie, na osi OY (oś rzędnych) umieszczana jest zmienna zależna, natomiast na osi OX (oś odciętych) – niezależna. Analiza korelacji powinna uwzględniać: kształt zależności (liniowa/nie liniowa), jej kierunek (dodatni/ujemny), siłę (słaba/umiarkowana/wysoka) i znamienność statystyczną r . Wartość współczynnika korelacji zawiera się w przedziale (0; 1], gdy korelacja jest dodatnia oraz w przedziale [-1; 0), gdy jest ujemna. Im bardziej wartość współczynnika korelacji zbliżona jest do zera, tym słabsza korelacja ($r = 0$ oznacza brak związku liniowego). Jeżeli wartość jest bliska jedności, korelacja jest prawie pełna. Siła związku oceniana jest niezależnie od jego kierunku. Warto zaznaczyć, iż niska wartość r Pearsona nie oznacza braku zależności między zmiennymi ilościowymi, ale jedynie brak zależności liniowej między nimi! Tymczasem może być obecny inny typ zależności, np. zależność logarytmicz-

na. Dlatego przed oszacowaniem r Pearsona (i jego znamienności statystycznej) zwykło się konstruować wykres rozrzutu, aby zweryfikować kształt zależności [11].

W przypadku gdy jedna ze zmiennych (zależna/niezależna) ma charakter zmiennej skokowej lub porządkowej, zależność ocenia się z wykorzystaniem współczynnika korelacji rang Spearmana [11, 19]. I wreszcie, dla zmiennych nominalnych wykorzystuje się całą gamę współczynników, obliczanych na bazie tabeli wielopolowej i testu χ^2 , takich jak np. współczynniki tau Kendalla, kontyngencji C , V Cramera, zbieżności Czuprowa czy ϕ Yule'a (ponadto już sam iloraz szans ze swoim 95-procentowym przedziałem ufności opisuje siłę zależności między zmiennymi) [11, 19]. Zwykle to od badacza zależy, który z nich zostanie zastosowany – ich interpretacja jest bowiem zbliżona. Opisane współczynniki przyjmują wartość od 0 (brak relacji między zmiennymi) do +1 (całkowita zależność); oprócz tau Kendalla i współczynnika korelacji rang Spearmana, dla których interpretacja jest zbliżona do r Pearsona.

Analiza stratyfikacyjna – warstwowanie

Analiza stratyfikacyjna (*stratum* = warstwa) stanowi kolejny etap testowania różnic lub zależności i z pewną precyzją pozwala kontrolować czynniki potencjalnie zakłócające wynik [7, 22]. Jeżeli celem pracy jest ocena wpływu stylu życia na występowanie nadciśnienia tętniczego, a w badanej grupie jest statystycznie więcej kobiet niż mężczyzn lub badani różnią się pod względem dystrybucji badanych czynników stylu życia, obserwacje te mogą znacznie wpłynąć na uzyskane dane i formułowane wnioski. Dlatego czasami badaną grupę warto jest podzielić na podgrupy, czyli warstwy (*strata*), w ramach których ponawia się analizy proste. Najczęściej warstwy takie definiowane są w oparciu o płeć, kategorie wieku, obecność cechy społeczno-ekonomicznej lub stanu zdrowia. I tak, analizując wpływ regularnego picia alkoholu na występowanie raka płuc, warto zastosować warstwowanie i ocenić ten związek u osób palących i niepalących, ponieważ informacja na temat natęgu palenia może mieć istotne znaczenie. Metodą służącą do tego celu jest statystyczna procedura Cochran-Mantel-Haenszel dla zmiennych jakościowych oraz analiza korelacji dla zmiennych ilościowych [23]. Interpretacja ich wyników jest tożsama z interpretacją surowego ilorazu szans [20] (w przypadku procedury Cochran-Mantel-Haenszel) oraz r Pearsona, obliczanych w analizach prostych [19].

Analiza wielu zmiennych

Ostatnim etapem statystycznej analizy danych jest tzw. analiza wielu zmiennych. Jej wyniki dostarczają najcenniejszych informacji, bowiem pozwalają odpowiedzieć na pytanie dotyczące związku przyczynowo-skutkowego oraz kontrolować wpływ czynników zakłócających, których często nie da się uniknąć [7, 11, 22]. Wybór typu analizy wielu zmiennych warunkowany jest również typem zmiennej zależnej. Gdy zmienną zależną stanowi zmienna ciągła (np. ocena czynników wpływających na stężenie glukozy w surowicy), zastosowanie znajduje regresja liniowa. Jeśli

jest ona zmienną nominalną (np. ocena czynników ryzyka nadciśnienia tętniczego) – regresja logistyczna. W pozostałych przypadkach (zmienna skokowa i porządkowa) można stosować regresję Poissona, jednak najlepiej – wykorzystując tzw. punkty odcięcia – wyrazić zmienną w sposób dychotomiczny i zastosować regresję logistyczną. Wybór zmiennych niezależnych do testowanego modelu opiera się zazwyczaj na wynikach analiz prostych, jednak nie można zapomnieć, iż to model biologiczny steruje modelem statystycznym, czyli typowaniem zmiennych, a nie odwrotnie. Zatem jeśli model biologiczny zdaje się potwierdzać udział zmiennej niezależnej we wnioskowaniu przyczynowo-skutkowym, ale wyniki analiz prostych tego nie potwierdzają, mimo wszystko warto uwzględnić taką zmienną w testowanym modelu wielu zmiennych. Pewnym ograniczeniem analizy wyników jest możliwość występowania współliniowości między zmiennymi niezależnymi, co pozwalają stwierdzić wyniki analiz prostych (np. wynik testu χ^2 $p < 0,05$ lub znamienne statystycznie r Pearsona o sile związku $> 0,3$). Aby zniwelować tę niedogodność, zmienna o mniejszym znaczeniu najczęściej usuwana jest z modelu (tak czyni się np. w regresji krokowej) lub „zakłócająca” zmienna wyrażona zostaje w postaci logarytmicznej [24].

Wynik równania regresji interpretowany jest w oparciu o istotność statystyczną współczynnika regresji (b) dla poszczególnych zmiennych niezależnych [11]. Jeśli wartość p współczynnika regresji wynosi $p < 0,05$ (czasem 0,1, gdy definiowana jest tzw. graniczna znamienność), to testowana zmienna niezależna ma znamienne statystycznie wpływy na wartość zmiennej zależnej. Wielkość tego wpływu determinuje wartość opisywanego współczynnika regresji (zwłaszcza liniowej). Określa on, o ile jednostek przeciętnie wzrośnie ($b > 0$) lub zmaleje ($b < 0$) wartość zmiennej zależnej, gdy wartość zmiennej niezależnej wzrośnie o jedną jednostkę. Miarą związku przyczynowo-skutkowego w modelu regresji logistycznej jest natomiast wartość logistycznego ilorazu szans ($\log OR$, $\log IS$) [11, 19], którego interpretacja nie odbiega od interpretacji surowego ilorazu szans (konstruowanego w oparciu o tabelę czteropolową w analizach prostych).

Podsumowanie

Podsumowując, analiza statystyczna danych stanowi pewnego rodzaju ciąg logicznych decyzji, które powinny być podjęte w dążeniu do realizacji zaplanowanego celu badawczego. Jedynie rzetelnie przeprowadzona analiza dostarcza cennych (wiarygodnych i użytecznych) informacji. Nie jest ona *panaceum* na osiągnięcie sukcesu w ocenie zależności i różnic między grupami oraz związku przyczynowo-skutkowego, choć w znaczącym stopniu decyduje o metodologicznej poprawności badania. Ocena wpływu wyniku procedury statystycznej na ostateczne wnioski zawsze powinna uwzględniać informacje płynące z literatury przedmiotu i powinna być uzupełniona zdrowym rozsądkiem badacza, gdyż tylko takie połączenie pozwoli uniknąć nieporozumień.

Praca została przedstawiona na Warsztatach Sekcji Kardiotorakoanestezjologii odbywających się w Szczawnicy w dniach 8–10 marca 2007 r.

Piśmiennictwo

1. Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. *BMJ* 1995; 310: 1122-1126.
2. Altman DG. The scandal of poor medical research. *BMJ* 1994; 308: 283-284.
3. Altman DG. Poor-quality medical research: what can journals do? *JAMA* 2002; 287: 2765-2767.
4. Zejda JE. Medyczny artykuł naukowy. Zasady dobrej praktyki publikacyjnej. *Ann Acad Med Siles* 2006; 60: 323-329.
5. Altman DG, Goodman SN, Shroter S. How statistical expertise is used in medical research. *JAMA* 2002; 287: 2817-2820.
6. Byrne DW. Common reasons for rejecting manuscripts at medical journals: a survey of editors and peer reviewers. *Science Editor* 2000; 23: 39-44.
7. Beaglehole R, Bonita R, Kjellstrom T. Podstawy epidemiologii. WHO, Geneva 1993. tłum. pod red. Neonili Szeszeni-Dąbrowskiej. IMP, Łódź 2002.
8. Lwanga SK, Lemeshow S. Sample size determination in health studies. Case-control studies. Hypothesis tests for an odds ratio. Geneva: World Health Organization, 1991.
9. Altman DG, Bland JM. How to randomize. *BMJ* 1999; 319: 703-704.
10. Altman DG. Randomization. *BMJ* 1991; 302: 1481-1482.
11. Swinscow TDV. *Statistics at Square One*. 9th Edition. BMJ Publishing Group, 1997. Available via internet (series: Free Books for Doctors): <http://www.bmj.com/collections/statsbk/>.
12. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ* 2006; 332: 1080.
13. Bland JM, Altman DG. Matching. *BMJ* 1994; 309: 1128.
14. Altman DG, Bland JM. Presentation of numerical data. *BMJ* 1996; 312: 572.
15. Altman DG, Bland JM. Quartiles, quintiles, centiles, and other quantiles. *BMJ* 1994; 309: 996.
16. Altman DG, Bland JM. Statistics notes: the normal distribution. *BMJ* 1995; 310: 298.
17. Matthews JN, Altman DG. Statistics notes. Interaction 2: Compare effect sizes not P values. *BMJ* 1996; 313: 808.
18. Altman DG, Bland JM. Statistics notes: Absence of evidence is not evidence of absence. *BMJ* 1995; 311: 485.
19. Stanisław A. *Przystępny kurs statystyki w oparciu o program „STATISTICA PL” na przykładach z medycyny*. T. 1. Kraków, 1998.
20. Bland JM, Altman DG. *Statistic Notes: The odds ratio*. *BMJ* 2000; 320: 1468.
21. Bland JM, Altman DG. Multiple significance tests: the Bonfferoni method. *BMJ* 1995; 310: 170.
22. Mullner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Arch Intern Med* 2002; 136: 122-126.
23. Berger VW, Stefanescu C, Zhou YY. The analysis of stratified 2 x 2 contingency tables. *Biom J* 2006; 48: 992-1007.
24. Bland JM, Altman DG. Transforming data. *BMJ* 1996; 312: 770.

Komentarz

mgr inż. Andrzej Szafranek
Pracownia Informatyki, Śląskie Centrum Chorób Serca, Zabrze

Prawda jest zbyt złożona, by dawała się obejrzeć inaczej niż w przybliżeniu.
John von Neumann

Odkrywanie prawdy o cechach badanej populacji na drodze interpretacji i uogólniania wyników badania opartej na analizie próby losowej stanowi zasadniczy instrument stosowany w pracy badawczej we wszystkich dziedzinach nauki, również w medycynie.

Tak rozumiane wnioskowanie statystyczne zawiera dwie grupy metod interpretacji wyników. Jedną to *estymacja* (parametryczna i nieparametryczna), czyli szacowanie nieznanymi parametrów i określanie postaci rozkładu, a drugą to *weryfikacja hipotez statystycznych* w oparciu o założony model koncepcyjny. W każdym z przypadków mamy do czynienia z podejmowaniem zasadniczych decyzji na podstawie wybranej, reprezentatywnej próby.

Analizowana praca adresowana jest głównie do naukowców-badaczy z zakresu medycyny, wykorzystujących w swoich badaniach złożony aparat metod statystycznych. Autor przedstawia doniosłość kolejnych etapów realizacji medycznych prac badawczych, tzn. fazę *statystycznego planowania eksperymentu* oraz *analizy statystycznej danych eksperymentalnych*. Postawiony w pracy cel jest jasny i niezwykle ważki, albowiem, jak pokazują badania, właśnie etap opracowania danych i użyte do tego metody dyskwali-

fikują czasem bardzo ciekawie zapowiadające się prace naukowo-badawcze.

David S. Moore, profesor z *Department of Statistics Mathematical Sciences Building* w swoim eseju donosi, że kiedy prowadzono badania nad skutecznością metod operacyjnego leczenia wrzodu dwunastnicy, po otwarciu jamy brzusznej pacjenta komisja stwierdziła, które z czterech technik operacyjnych mogą być zastosowane w tym przypadku [1]. Po ustaleniu tego chirurg operujący dostawał przygotowaną przez statystyka kopertę z informacją określającą, która z możliwych technik ma być zastosowana. Postępując zgodnie z informacją, uzyskiwano dwa cele – właściwe leczenie pacjenta oraz zapewnienie pełnowartościowych wyników potrzebnych do dalszych badań statystycznych, tzn. takich, gdzie zminimalizowano nieprzewidziane czynniki zewnętrzne. Powyższy przykład świadczy o doniosłości problemu zapewnienia optymalnych warunków planowania eksperymentu jako wstępu do stawiania i weryfikowania hipotez.

Również, co podkreśla autor, w realizacji projektu badawczego bardzo ważne jest zapewnienie reprezentatywności próby, która zależy od dwóch czynników – sposobu

doboru (arbitralny lub losowy) oraz jej liczebności. Jednocześnie należy pamiętać, że zgodnie z *centralnym twierdzeniem granicznym* wraz ze wzrostem liczebności próby rozkład z próby dąży do rozkładu normalnego o średniej równej średniej w populacji i wariancji równej wariancji w populacji podzielonej przez pierwiastek z liczebności próby. Płyń z tego wniosek, że minimalizację błędu można osiągnąć poprzez zwiększanie liczebności próby, nie zapominając, że maleje on proporcjonalnie do pierwiastka kwadratowego liczebności, więc kierunek ten jest dosyć kosztowny.

Etap analizy danych statystycznych przedstawiony jest w pracy w sposób jasny i czytelny. Pokazano sposób dochodzenia do końcowych rezultatów w zależności od rodzaju zmiennej losowej (ciągła lub dyskretna), ich powiązania, charakteru rozkładu, jakiemu podlegają czy występowania (lub braku) jednorodności wariancji. Przedstawiono ogólne wskazania i kryteria do zastosowania określonego testu statystycznego oraz dobrej praktyki doboru poziomu istotności testu alfa.

Rozpatrując analizę statystyczną różnic pomiędzy grupami dla zmiennych jakościowych, większość autorów uważa za zasadne użycie poprawki Yatesa modyfikującej test χ^2 już w sytuacji, kiedy wartość oczekiwana w którymkolwiek polu tablicy czteropolowej jest mniejsza od 10, ale wciąż większa od 5, natomiast wystąpienie którejkolwiek wartości oczekiwanej ≤ 5 wymaga bezwzględnie zastosowania dokładnego testu Fischera [2, 3].

Wydaje się, że należy trochę szerzej, niż to przedstawiono w części *Analiza wielu zmiennych* wspomnieć o tzw. *data mining*, czyli procesie selekcji, eksploracji i modelowania dużej liczby wielowymiarowych danych, który służy odkrywaniu regularności i związków w nich występujących z wykorzystaniem komputerowych algorytmów teorii sztucznej inteligencji.

Przykładem takiego podejścia może być zastosowanie metody funkcji dyskryminacji (uczenia z nauczycielem) do wyprowadzenia równania pozwalającego jedynie na bazie analizy EKG prognozować efekt leczenia trombolitycznego w świeżym zawale mięśnia serca, osiągając bardzo wysoką czułość i specyficzność metody [5, 6].

Podobnie modyfikacja skali EUROSCORE poprzez wyliczenie nowych współczynników równania regresji logistycznej w oparciu o dane pacjentów operowanych w Śląskim Centrum Chorób Serca dała lepszą predykcję prognozowania ryzyka operacyjnego, co potwierdzono testem Hosmer-Lemeshowa oraz wartością pola pod krzywą ROC [7, 8].

Przedstawione w pracy zagadnienia można ująć w postaci schematu blokowego lub tablicy pozwalającej poprawnie wybrać *ścieżkę decyzyjną* w procesie wnioskowania statystycznego. Taką właśnie drogę wybrał autor, omawiając poszczególne sytuacje, a w końcu zestawiając poszczególne przypadki w tabeli. Podejście to z pewnością pozwoli czytelnikowi na zaznajomienie się z metodologią wykorzystania metod statystycznych oraz właściwe pokierowanie się nią we własnych pracach badawczych.

Reasumując, należy uznać, że cel pracy rozumiany jako przedstawienie kolejnych etapów poprawnej analizy statystycznej został w pełni osiągnięty, a praca może służyć badaczom jako poradnik statystyczny.

Piśmiennictwo

1. Lynn AS, Moore DS Mathematics Today Twelve Informal Essays. Springer-Verlag, New York 1979.
2. Gutman I, Wilks SS, Hunter JS. Introductory Engineering Statistics. John Wiley & Sons, 1982.
3. Mańczak K. Technika planowania eksperymentu. Wydawnictwa Naukowo-Techniczne, 1976.
4. Stanisz A. Przystępny kurs statystyki, Kraków 2007.
5. Buszman P, Szafranek A, Pasyk S. Non Invasive Prediction of Reperfusion on Basic of Computer Quantitative Analysis of ECG In AMI. American College of Cardiology, Dallas, Texas 1992.
6. Buszman P, Szafranek A. A complex analysis of electrocardiogram in patients undergoing thrombolysis for acute myocardial infarction: early electrocardiogram changes in relation to the result of treatment. 17th Congress of the European Heart Journal, 1995.
7. Szafron B, Szafranek A, Zembala M. Comparison of different risk stratification models in isolated CABG surgery, Interactive Cardiovascular and Thoracic Surgery, Ateny 2005.
8. Szafron B, Szafranek A, Zembala M. Risk stratification in cardiac surgery: the search between simplicity and precision, 55 International Congress St. Petersburg, Russian Federation 2006.