

Jerzy A. MOCZKO

Zmienne wikłające w badaniach obserwacyjnych i ich wpływ na uzyskiwane wyniki

Confounding variables in observational studies and their influence on obtained results

Katedra i Zakład Informatyki i Statystyki
Uniwersytetu Medycznego w Poznaniu
Kierownik: Prof. zw. dr hab. Jerzy A. Moczko

Dodatkowe słowa kluczowe:

badania obserwacyjne
wnioskowanie statystyczne
zmienne wikłające
metody nieparametryczne

Additional key words:

observational studies
statistical inference
confounding variables
nonparametric methods

Podczas badania wpływu jednej zmiennej na inną należy się upewnić, czy obserwowany efekt nie został spowodowany wpływem innych czynników nie wziętych pod uwagę w prowadzonym eksperymencie. Czynniki te zwane zmiennymi wikłającymi, wpływając jednocześnie na obie analizowane wielkości, mogą wygenerować sztuczne, nie istniejące w rzeczywistości związki lub zmienić kierunek związków rzeczywistych.

In estimating the influence of one variable on another one we must ensure that observed effect is not caused by factors other than these under investigation. These last, called confounders of analyzed variables, influencing simultaneously both analyzed quantities may generate spurious non existing in reality relationships or change direction of these really existing.

Analiza tablic kontyngencji – podejście naiwne

Jedną z najczęściej stosowanych przez eksperymentatorów narzędzi do badania zależności między zmiennymi mierzonymi w skali nominalnej jest test Chi-kwadrat. Test ten z uwagi na swoją prostą konstrukcję, jak i szeroką dostępność zdobył sobie szeroką popularność. Spróbujemy jednak przyrzeć się dokładniej niebezpieczeństwom, na jakie może natknąć się badacz stosując nieodpowiedzialnie tę „oczywistą” technikę statystyczną.

Jedną z podstawowych zasad podczas prowadzenia badań eksperymentalnych jest randomizacja przypadków mająca między innymi na celu zrównoważenie rozkładów badanych cech w analizowanych podgrupach. Całkiem inna sytuacja pojawia się w badaniach obserwacyjnych, takich jak obserwacyjne doświadczenia kliniczne lub studia epidemiologiczne. W tego typu analizach badacz rejestruje jedynie interesujące go dane nie mając żadnego wpływu na ich rozkład; rozkład ten jest uwarunkowany wieloma różnymi czynnikami, często trudnymi do obserwacji lub wręcz niemożliwymi do uwzględnienia w protokole badania.

Zilustrujmy to zjawisko wykorzystując dane zebrane w ramach kohortowego badania stanu zdrowia kobiet rozpoczętego w latach 1972-1974 i powtórnego po dwudziestu latach w Whickham w Wielkiej Brytanii [1]. W badaniu tym rejestrowano m.i. zgony kobiet w okresie kolejnych 20 lat. Próbowano znaleźć, czy występuje związek między paleniem tytoniu a wystąpieniem zgonu. Uzyskane dane przedstawione są w tabeli I. Jak widać, jest to prosta czteropolowa tablica kontyngencji, której kolumny zawierają przynależność do jednej z dwóch populacji (kobiety palące tytoń versus ko-

biety niepalące) a wiersze zmienną odpowiedzi (wystąpił zgon versus kobieta nadal żyje). Próbujemy udzielić odpowiedzi na pytanie, czy istnieje związek przyczynowy między wymienionymi uprzednio zmiennymi. Najczęstszym działaniem podejmowanym przez badaczy jest analiza danych zawartych w tabelicy I przy użyciu testu chi-kwadrat. Spróbujemy pójść tą drogą i dokonajmy obliczeń wykorzystując prosty pakiet statystyczny INSTAT firmy Graphpad, Inc. Wyniki zamieszczone są w tabeli II. Jak widać, obliczona wartość prawdopodobieństwa p jest bardzo niska i wynosi 0.0031. Oznacza to, że zakładając prawdziwość hipotezy zerowej (w naszym przypadku hipoteza zerowa oznacza brak związku między paleniem tytoniu a ryzykiem zgonu), prawdopodobieństwo uzyskania co najmniej takiej wartości statystyki Chi-kwadrat (8.752) jest bardzo niskie. Zastosowanie poziomu istotności $\alpha=0,05$ ogranicza liczbę wniosków fałszywie dodatnich do maksimum pięciu procent. W związku z tym będziemy próbowali wyciągnąć wniosek, że hipoteza zerowa jest mało prawdopodobna i należy ją odrzucić przyjmując jednocześnie hipotezę alternatywną stwierdzającą występowanie związku między badanymi zmiennymi. Wnioskujemy zatem, że palenie tytoniu wywiera istotny wpływ na częstość występowania zgonów.

Analizując dalej uzyskane wyniki zauważamy, że frakcja zgonów w grupie kobiet palących wyniosła w okresie 20 lat 0.2388, natomiast analogiczna frakcja w grupie kobiet niepalących aż 0.3142. Podobnie iloraz szans wynoszący 0.6848 wskazuje na fakt, że palenie tytoniu nie tylko nie jest szkodliwe, lecz więcej – obniża frakcję zgonów. Za ten wniosek moglibyśmy oczekiwać od firm tytoniowych sporej gratyfikacji. Czy za-

Adres do korespondencji:
Prof. dr hab. Jerzy A. Moczko
Katedra i Zakład Informatyki i Statystyki
ul. Dąbrowskiego 79
60-529 Poznań
Tel./fax. (+61) 841 11 16
e-mail: jmoczko@amp.edu.pl

tem eksperyment został źle zaplanowany, a może nieprawidłowo zebrano dane?

Analiza tablic kontyngencji – podejście zaawansowane

Przyjmując strategię działania przedstawionego w poprzedniej części pracy popełniliśmy niestety kilka błędów. Spróbujemy je teraz przeanalizować i zaprezentujemy prawidłowy sposób analizy przedstawionego problemu.

Błąd pierwszy dotyczy założenia, że w przypadku wykazania istotnej statystycznie zależności między zmiennymi udowodnimy jednocześnie istnienie związku przyczynowego łączącego te zmienne. Stwierdzenie to nie jest prawdziwe. Otóż nawet generowane w sposób losowy zmienne mogą czasami wykazywać istotną statystycznie zależność, jednakże nie ma ona żadnego związku przyczynowego. Aby określić związek przyczynowy w badaniach obserwacyjnych trzeba rozważyć kryteria opracowane przez *Hilla* [2]:

1. przyczyna musi poprzedzać skutek
2. wyniki muszą być sensowne pod względem biologicznym
3. wiele źródeł powinno potwierdzać uzyskane wyniki
4. statystyczny związek między przyczyną a skutkiem powinien być istotny i silny
5. usunięcie badanego czynnika powinno zmniejszyć ryzyko występowania badanego zjawiska

Pierwszy warunek jest rzeczywiście spełniony (palenie tytoniu musiało wystąpić lub nie mieć miejsca przed ewentualnie zarejestrowanym zgonem badanej kobiety). To samo możemy powiedzieć o spełnieniu warunku trzeciego – istotnie, wiele źródeł naukowych donosi o szkodliwości tego nalogu. Przeprowadzając obliczenia przy użyciu testu Chi-kwadrat wykazaliśmy istotność statystyczną istniejącego związku na poziomie co najmniej 0.05. Przyjrzyjmy się zatem drugiemu warunkowi. Porównujemy dwie subpopulacje kobiet: palących tytoń (582) oraz niepalących (732). W każdej z nich zmarło odpowiednio: 139 i 230 osób. Frakcja zmarłych wśród palących wynosi zatem 0.2388 a wśród niepalących 0.3142. Frakcje zgonów różnią się istotnie o wartość 0.07538 na niekorzyść kobiet niepalących. Podobnie wyznaczenie ilorazu szans daje w wyniku wartość 0.6848, co prowadzi do co najmniej dziwnego wyniku (sprzecznego z danymi uzyskiwanymi z innych źródeł), iż palenie tytoniu zmniejsza ryzyko zgonu. Być może źródłem tej „rewelacji” jest nieprawidłowo dobrany test statystyczny. Pamiętajmy bowiem, że test Chi-kwadrat daje jedynie wyniki asymptotyczne i są one prawidłowe w przypadku spełnienia warunków *Cochrana*:

1. wszystkie wartości oczekiwane są większe od 1
2. co najmniej 20% wartości oczekiwanych jest większa od 5

Łatwo sprawdzić, że analizowany przez nas układ danych spełnia nadmiarowo stawiane warunki (najmniejsza wartość oczekiwana przekracza 163) a więc to nie wy-

Tabela I

Tablica kontyngencji zestawiająca status (żyje, nie żyje) badanych kobiet z ich zachowaniami dotyczącymi palenia tytoniu.

Contingency table illustrating vital status of examined women by their smoking behavior.

	Zmarła	Żyje	Suma
Pali	139	443	582
	(11%)	(34%)	(44%)
Nie pali	230	502	732
	(18%)	(38%)	(56%)
Suma	369	945	1314
	(28%)	(72%)	(100%)

Tabela II

Wyniki obliczeń dla danych z tabeli I uzyskane programem Instat firmy Graphpad, Inc.

Results of calculations obtained with INSTAT 3.05 (GraphPad Software Inc.) .

Chi-square Test			
The two-sided P value is 0.0031, considered very significant.			
The row/column association is statistically significant.			
Calculation details:			
Chi-square statistic (with Yates correction) = 8.752			
Degrees of freedom = 1			
Relative Risk			
Relative risk = 0.7601			
95% Confidence Interval: 0.6347 to 0.9103 (using the approximation of Katz.)			
Difference between the two proportions			
Top row (pali):			
Fraction in the left column: 0.2388			
95% Confidence Interval of that fraction: 0.2049 to 0.2761			
Bottom row (nie pali):			
Fraction in the left column: 0.3142			
95% Confidence Interval of that fraction: 0.2813 to 0.3496			
Difference:			
Difference between the fractions: 0.07538			
Standard error of the difference: 0.02496			
95% confidence interval of difference: 0.02645 to 0.1243			
Odds Ratio			
Odds ratio= 0.6848			
95% Confidence Interval: 0.5353 to 0.8762 (using the approximation of Woolf.)			
Data analyzed			
	zmarła	żyje	Total
pali	139 (11%)	443 (34%)	582 (44%)
nie pali	230 (18%)	502 (38%)	732 (56%)

Total	369 (28%)	945 (72%)	1314 (100%)
* * *			

bór testu dającego wyniki asymptotyczne prawdopodobieństwa prowadzi nas do dziwiącego wniosku (użycie w miejscu testu Chi-kwadrat dokładnego testu *Fishera* daje w wyniku wartość $p=0.0030$ czyli minimalnie różniącą się od wyznaczonej uprzednio wartości asymptotycznej).

Cóż zatem powoduje, że uzyskane wyniki stoją w jawnej sprzeczności zarówno z wynikami uzyskanymi z innych źródeł, jak i zdrowym rozsądkiem? Czy należy podejrzewać, że badana populacja kobiet w Wic-kham wyróżnia się w jakiś szczególny spo-

sób na tle innych zbiorowości kobiet? Przy dokładniejszej analizie problemu okazuje się, że przyczyną zaobserwowanego zjawiska są tzw. zmienne wnikające (confounding variables) znane również pod nazwą zmiennych przyczajonych (*lurking variables*). Uwikłanie jest czynnikiem związanym z ryzykiem wystąpienia choroby w populacji zarówno eksponowanej, jak i nie eksponowanej na badany czynnik ryzyka (w naszym przypadku osób nie palących tytoniu), a jednocześnie nie jest uzależniony od czynnika ryzyka. W analizowanym przez nas ekspe-

Tabela III
Struktura danych zawartych w tabeli I po stratyfikacji według wieku badanych kobiet.

Structure of the data presented in table I after stratification according to the age of examined women.

Wiek 18-44	Zmarła	Żyje	Suma
Pali	19	13	32
	(3%)	(2%)	(5%)
Nie pali	269	327	596
	(43%)	(52%)	(95%)
Suma	288	340	628
	(46%)	(54%)	(100%)
Wiek 45-64	Zmarła	Żyje	Suma
Pali	78	52	130
	(18%)	(12%)	(29%)
Nie pali	167	147	314
	(38%)	(33%)	(71%)
Suma	245	199	444
	(55%)	(45%)	(100%)
Wiek > 64	Zmarła	Żyje	Suma
Pali	42	165	207
	(17%)	(68%)	(86%)
Nie pali	7	28	35
	(3%)	(12%)	(14%)
Suma	49	193	242
	(20%)	(80%)	(100%)

Tabela IV
Wyniki testów jednorodności ilorazów szans.

Results of test for homogeneity of odds ratios.

```

!Cytel Studio (8.0.0)
!Stratified 2 x 2 Tables:Homogeneity of Odds-Ratio Test
>>> binomial (type = independent, test_type = oddsratio, compute = ho,
method = exact, time_limit = none );
Datafile: <new>

TEST FOR HOMOGENEITY OF ODDS RATIOS
[ 3 2x2 informative tables ]

Observed Statistics:
  BD: Breslow and Day Statistic      =      0.9501
  BDT: B & D stat(With Tarone's corection ) =      0.9498
  ZE: Zelen Statistic                =      0.0203

Asymptotic p-value: (based on Chi-Square distribution with 2 df )
  Pr { BD .GE.      0.9501 }      =      0.6218
  Pr { BDT .GE.    0.9498 }      =      0.6220

Exact p-value:
  Pr { ZE .LE.     0.0203 }      =      0.5539

Elapsed Time is 0:0:0.16

```

rymencie przykładem takiego czynnika może być wiek badanego osobnika. Wiek jest jak najbardziej związany z ryzykiem zgonu w każdej populacji (zarówno palących, jak i niepalących), z drugiej zaś strony nie jest on zależny od faktu, czy osobnik pali czy nie pali tytoń.

Spróbujmy przeprowadzić ponowną analizę danych uwzględniając ich podział na trzy grupy wiekowe: 18-44, 45-64, >64 lat. Ten tzw. proces stratyfikacji powoduje on rozbięcie tabeli I na trzy tablice cząstkowe, których kształt przedstawia tabela III. Niestety, analiza stratyfikowanych tablic kontyngencji nie jest już tak powszechnie dostępna, jak analiza niestratyfikowana. Do

przeprowadzenia obliczeń wykorzystamy pakiet Statxact firmy Cytel, Inc., wersja 8.0. Pamiętając sposób obliczania ilorazu szans (OR) z tablicy kontyngencji 2 x 2 (stosunek iloczynu elementów macierzy diagonalnych do iloczynu elementów antydiagonalnych) łatwo obliczyć, że ilorazy te dla kolejnych trzech tablic wynoszą odpowiednio: 1.777, 1.320 oraz 1.018.

Wykażemy teraz testem jednorodności ilorazów szans *Breslowa-Day'a* oraz *Zelena* [3,4], że uzyskane wartości ilorazów szans nie różnią się od siebie statystycznie. Odrzucenie hipotezy zerowej o równości OR prowadziłoby do wniosku, że w różnych grupach wiekowych istnieje rozmaity wpływ

palenia tytoniu na śmiertelność. Zjawisko to znane jest pod nazwą interakcji lub efektu modyfikacji. Zarówno asymptotyczne wyniki testu *Breslowa-Day'a* (z poprawką *Tarona* ($p \leq 0.9498$) i bez niej ($p \leq 0.9501$)), jak i wynik dokładny testu *Zelena* ($p \leq 0.5539$) przedstawione w tabeli IV wskazują, że nie mamy podstaw do odrzucenia hipotezy zerowej (wszystkie wartości prawdopodobieństwa p znacznie przekraczają przyjęty próg decyzyjny poziomu istotności $\alpha = 0.05$).

Możemy zatem przejść do kolejnego kroku – oszacowania wspólnego, uśrednionego ilorazu szans dla analizowanych trzech warstw [4,5]. Wypadkowy estymator ilorazu szans dla trzech warstw oszacowany asymptotyczną metodą *Mantela-Haenszela* wynosi 1.357. Istotność statystyczna tego wyniku została oszacowana dwoma algorytmami: klasyczną metodą *Mantela-Haenszela* ($p \leq 0.0719$ (*M-H variance*)) oraz metodą *RBG Robinsa, Breslowa i Greenland* [6] ($p \leq 0.0738$ (*with RBG variance*)).

Obie techniki wskazują na marginalną istotność uzyskanych wyników, co potwierdza również fakt, iż 95.00 procentowy przedział ufności dla wypadkowego ilorazu szans oszacowany metodą *RBG* (0.9709, 1.897) zawiera w sobie wartość $OR = 1$. Dlatego wykonano dodatkowo obliczenia dokładne (metodami permutacyjnymi), dające estymator wypadkowy $OR = 1.3605$

Niestety, podobnie jak w przypadku metod asymptotycznych 95 procentowy przedział ufności uwzględniający poprawkę *mid-p* (0.9737, 1.909) również pokrywa wartość $OR = 1$.

Wnioski

Przytoczone w poprzednim rozdziale wyniki należy interpretować w sposób następujący.

Nie udało nam się wykazać na poziomie istotności statystycznej $\alpha = 0.05$ wpływu palenia tytoniu na wzrost liczby zgonów w badanej populacji kobiet. Uzyskane wyniki są jednak bardzo bliskie progu decyzyjnego a estymowany wypadkowy iloraz szans $OR = 1.3605$ wskazuje na istnienie tendencji zmniejszenia liczby zgonów we frakcji kobiet niepalących. Wiek badanych kobiet jest niezwykle istotną zmienną, bez której prawidłowa analiza danych jest niemożliwa i prowadzi do fałszywych wniosków. Po dokonaniu stratyfikacji danych względem wieku dodatni (lecz nieistotny statystycznie) związek między paleniem tytoniu a ryzykiem zgonu może zostać wytłumaczony poprzez różnicę proporcji kobiet palących w różnych grupach wiekowych (najwyższa frakcja kobiet palących – około 55% znajduje się w grupie wiekowej 45-64, w porównaniu do 46% w grupie 18-44 oraz 20% w grupie >64) oraz naturalny fakt wyższej śmiertelności w grupie kobiet starszych (odpowiednio 5%, 29% oraz 86% w podgrupach wiekowych). Gdy zignorujemy w analizie wiek kobiet, różnica w odsetkach śmiertelności zamiast ilustrować wpływ efektu palenia wskazuje na tendencję niższego wieku u osób palących i wyższego u osób niepalących.

Obserwujemy tzw. paradoks *Simpsona*, który polega na odwróceniu kierunku związku między zmiennymi na skutek połączenia danych pochodzących z kilku podgrup

Tabela V

Wyniki estymacji i testowania wypadkowej wartości ilorazu szans dla stratyfikowanych tablic kontyngencji.
Results of estimation and testing of common odds ratio for stratified contingency tables.

```
!Stratified 2 x 2 Tables:CI on Common-Odds-Ratio Test
>>> binomial (type = independent, test_type = oddsratio, compute = ci,
method = exact, time_limit = none );
Datafile: <new>

ESTIMATION AND TESTING OF COMMON ODDS RATIO
[ 3 2x2 informative tables ]

Summary of Exact Distribution of S [sum of counts in cell(1,2)]:
  Min      Max      Mean      Std-dev      Observed      Standardized
 158.0     355.0     240.7     5.933        230.0         -1.800

Exact p-values for testing that the Common odds Ratio is 1:
  One-sided: Pr { S .LE.      230.0 } = 0.0426
              Pr { S .EQ.      230.0 } = 0.0133
  Two-sided: Method 1: 2 * One-sided = 0.0852
              Method 2: (Sum of Probs .LE. 0.0133) = 0.0768
              Method 3: Pr{|S-Mean| .GE. | 230.0-Mean|} = 0.0768

Exact Estimation of Common Odds Ratio:
  Conditional maximum likelihood estimate: 1.3605
  95.00 % Conf. Intervals:
    Exact      : ( 0.9606 , 1.936)
    Mid-p corrected : ( 0.9737 , 1.909)

Mantel-Haenszel Inference:
  Common Odds Ratio estimate: 1.357
  Two-sided p-value: 0.0738 (with RBG variance)
                   0.0719 (M-H variance)
  95.00% CI with RBG variance: ( 0.9709 , 1.897)

Elapsed Time is 0:0:0.03
```

w jedną grupę.

Paradoks *Simpsona* i występowanie zmiennych wnikających jest jednym z trudności, na jakie może napotkać eksperymentator podczas analizy obserwacyjnych danych doświadczalnych i które musi wziąć koniecznie pod uwagę, aby uniknąć wysuwania nieprawidłowych wniosków. Istnieją trzy przydatne reguły diagnostyczne pozwalające na wykrycie potencjalnego efektu uwikłania:

1. potencjalna zmienna wnikająca musi być powiązana zarówno ze zmienną odpowiedzi, jak i badanym predyktorem
2. uwzględnienie potencjalnej zmiennej wnikającej musi zmieniać w sposób znaczący oszacowywany współczynnik predykcji dla badanego predyktora
3. powinien być wiarygodny i dopuszczalny z punktu widzenia merytorycznego jako czynnik modyfikujący badane zmienne.

Piśmiennictwo

1. **Vanderpump M.P., Turnbridge W.M., French J.M. et al.**: The development of ische-mic heart disease in relation to autoimmune thyroid disease in a 20-year follow-up study in English community. *Thyroid* 1996, 6, 155.
2. **Hill A.B.**: The environment and disease; association or causation? *Proc. Royal Soc. Med.* 1965, 58, 295.
3. **Breslow N.E., Day N.E.**: The analysis of case-control studies. IARC Scientific Publications No. 32 Lyon, France, 1980
4. **Zelen M.**: The analysis of several 2 x 2 contingency tables. *Biometrika* 1971, 58, 129.
5. **Gart J.**: Point and interval estimation of the common odds ratio in the combination of 2 x 2 tables with fixed marginals. *Biometrika* 1970, 57, 471.
6. **Robins J., Breslow N., Greenland S.**: Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models. *Biometrics* 1986, 42, 311.