

Halina Widła

Uniwersytet Śląski

O NIEPRAWIDŁOWYM STOSOWANIU NARZĘDZI STATYSTYCZNYCH W BADANIACH ANKIETOWYCH

On inaccurate use of statistical tools in questionnaire studies

The article explores common mistakes made while interpreting statistics in questionnaire studies, which may reflect insufficient training in research methodology. It reveals problems in qualitative and quantitative analysis in the published literature. It also discusses the possible reasons why errors occur in the process of constructing diagrammatic representations of the collected data. Examples of errors as they occur in published articles are also cited to prove that such problems are far from being hypothetical and do appear in research reports. They are mainly reflective of erroneous beliefs concerning statistical significance testing, chi-square analysis, covariance statistical corrections, as well as presentation of statistically inaccurate results.

1. Wstęp

Celem tego artykułu jest zwrócenie uwagi na kilka problemów, które często pojawiają się w toku wystąpień na wielu konferencjach naukowych, przy okazji prezentacji sprawozdań z badań ankiетowych. Niewątpliwie jest rzeczą chwalebną, że referenci nie chcą pozostać gołosłowni w relacjach i nie poprzestają na wrywkowych obserwacjach, lecz dążą do obiektywnej oceny rzeczywistości, uciekając się do narzędzi, jakimi dysponuje statystyka. Problem w tym, że są to tylko i wyłącznie narzędzia i jako takie mogą być stosowane z różnym efektem. Nieprawidłowe zastosowanie tych przydatnych narzędzi stwarza złudne wrażenie rzetelnego podejścia naukowego autora, gdy tymczasem problem często nie polega na wyborze narzędzi, a na jakości materiału, do którego obróbki one służą. Spośród za-

uważonych najczęściej „przewinień” referentów, przytaczam trzy, w moim przekonaniu najbardziej niebezpieczne. Skupię się szczególnie na:

- błędach na etapie konceptualizacji;
- nieprawidłowym grupowaniu danych oraz
- nieuprawnionych porównaniach.

2. Błędy konceptualizacji

Pierwszym napotkanym problemem jest określenie celu badań i dobór grupy badawczej. Już samo wyartykułowanie problemu badawczego zdradza stopień przygotowania metodologicznego autora badania.

Przykład 1

Teza badawcza dotycząca kariery życiowej młodych ludzi: studentki filologii są dokładniejsze i pracowitsze od studentów, a przez to lepiej radzą sobie w pracy zawodowej.

Czy jednak dysponujemy danymi potwierdzającymi różnice pod tym względem między płciami? Czy traktuje o tym nauka czy jest to obiegowa opinia? Czy wzięto pod uwagę zdolności i inne czynniki, przy założeniu, że wcześniej dysponowano na ten temat wynikami poważnych badań? Czy dokładność i pracowitość to jedyne cechy warunkujące zawodowy sukces? Jak zmierzyć ów sukces? Jak porównać karierę zawodową w poszczególnych dziedzinach? Co jest miarą sukcesu nauczyciela a co tłumacza? Zarobki? Prestiż zawodowy? A może własna satysfakcja? Jeśli tak, to jak ją zmierzyć? Czy i na ile stopień satysfakcji zawodowej jest wynikiem predyspozycji charakterologicznych i udanego życia osobistego? Czy nie ma wreszcie innych czynników które by jeszcze lepiej tłumaczyły kobiecą zaradność w pracy?

Przykład 2

Problem badawczy: tematem badań jest interferencja w wypowiedziach osób wielojęzycznych. W tym celu badaniom poddano interpretacje tłumaczeń z języka obcego na ojczysty i porównano te wypowiedzi pod kątem przenikania się struktur i słownictwa. W badaniu wzięło udział 180 osób trójjęzycznych, reprezentujących rozmaite „triady”, np. polski-włoski-rosyjski, niemiecki-angielski-hiszpański itp.

Czy w ogóle można porównywać, a jeśli tak, to pod jakim kątem, wypowiedzi osób o tak różnych biografiach językowych? Czy porównywalna jest ilość i jakość badanych faktów w wypadku osób znających języki mniej lub bardziej typologicznie od siebie odległe i przyswojone w nieporównywalnej kolejności, w różnych warunkach nauczania i w dodatku w nierównym stopniu? Jaka jest wartość uogólnienia takich porównań i czy w ogóle mają one sens? Badacz nie może stawiać przytoczonej tezy, zanim nie ustosunkuje się do powyższych kwestii.

Przykład 3

Hipoteza badawcza: nauczyciele z Dolnego Śląska nie mają dostępu do dobrze wyposażonych laboratoriów językowych.

Czy rzeczywiście stwierdzenie takiej bolączki wymaga naukowej obudowy? Czy nie wystarczy przyjąć takiej informacji, pozyskiwanej w banalnie prosty sposób, za fakt służący do rozważań na temat jakości pracy i wyników, ale nic ponadto?

Przykład 4

Konstruowanie pytań.

Na temat sposobu konstruowania pytań w kwestionariuszach wylano już przysłowiowe morze atramentu. Okazuje się jednak, że mimo wielu rad i przestróg, wciąż do głównych przewinień badaczy należą:

- pytania dwuznaczne, niezrozumiałe i prowadzące do nieporozumień;
- pytania sugerujące.

Choć nie ma w zasadzie pytań całkiem neutralnych, choćby przez to, że każą zwrócić uwagę na problemy, które często nigdy przedtem badanych nie nurtowały, a przez to wymuszają powierzchowne sądy, to przecież istnieje zdroworozsądkowe podejście nie tolerujące choćby wartościujących pytań w stylu:

- *Czy jesteś za ograniczeniem suwerenności i zgodą na monitoring w klasie?*
- *Czy jesteś za zwalczaniem chuligaństwa i zgodą na monitoring w klasie?*

Osobnym i częstym problemem jest nie do końca przemyślana liczba możliwości do wyboru, że nie wspomnę o sugestyności takich propozycji. Dlatego zawsze w wyliczeniu rozmaitych możliwości, oczywiście, po wcześniejszym dochowaniu należytej dbałości o stworzenie zbioru rozłącznych zakresów, należy dodać zabezpieczenie w rodzaju:

- *Inne:*
- *Inne (jakie?):*

Należy też wspomnieć o braku wrażliwości niektórych autorów. Pytania kłopotliwe, nie dość, że winny być stawiane na końcu, gdy ankietowany wyrobi sobie pogląd na temat samych badań i zdecyduje na stopień szczerości, to powinny dawać margines ochrony osobistych danych:

- *Wiek: najlepiej w przedziałach*
- *Zarobki: wyłącznie w przedziałach*

Nawet anonimowa ankieta może powodować duży dyskomfort badanych (zrodzić obawy o zidentyfikowanie) i nastroić nieufnie i niechętnie do idei przystępowania do badania czy eksperymentu.

Takie tematy można by mnożyć, a przecież wystarczy dokładnie przeczytać w literaturze przedmiotu, jakiego typu sądy spełniają kryteria naukowości, a także jak ustrzec się przed ryzykiem spowodowanym przez zmienne zakłócające. Nie da

się, oczywiście, potraktować problemów poruszanych w naukach humanistycznych jak obszaru zainteresowania biologa czy chemika. Dlatego mówi się o naturalnym marginesie błędu (około 20%), co w naukach ścisłych byłoby niepodobieństwem. Niemniej jednak, zawężenie problemu do w miarę zhomogenizowanej grupy badawczej, jasne określenie celów, czynników, które w literaturze przedmiotu uchodzą za sprzyjające lub zakłócające wyniki badań, to podstawowy obowiązek każdego, kto chce sięgnąć po narzędzia do obróbki statystycznej danych. Zdobądźmy najpierw prawidłowo te dane i niech będą one rzetelne.

3. Nieprawidłowe grupowanie danych (manipulacja)

Kolejnym grzechem, i to zatrwajająco częstym, jest manipulacja danymi: od najprostszej po najbardziej wyrafinowaną, więc i w tej kolejności zacytuję przykłady:

Przykład 1

Średnia arytmetyczna.

Średnia arytmetyczna to przykład oklepany, ale przytaczam go nie tylko dla porządku. Jest nagminną procedurą i bardzo niebezpieczną z dwóch powodów.

$$\bar{x} = \frac{\sum fx}{n}$$

Po pierwsze, co wszystkim wiadomo, zaciemnia obraz i to tak skutecznie, że nie pozostaje nic innego, jak posądzenie osoby uciekającej się do takich wyliczeń, bez stosownych zabezpieczeń, o manipulację danymi. Mało kto pamięta, że średnia arytmetyczna nie jest jedynym miernikiem tendencji centralnej. O wiele częściej winno się sięgać po *medianę* czy *modalną*. Ta pierwsza jest wartością środkową, powyżej i poniżej której znajduje się jednakowa liczba obserwacji. Warto ją znać, gdy porównujemy np. zarobki, tu średnia niewiele nam wnosi. A jeszcze lepiej znać też wtedy *modalną* (*dominantę*), czyli najczęstszą wartość w danym zbiorze wyników. Wreszcie przykład z dydaktycznego podwórka: średniej ocen grup o różnej liczebności nie da się prawidłowo obliczyć przy pomocy wzoru na średnią arytmetyczną, konieczna jest tu formuła obliczania tzw. *średniej ważonej*. Przykładowo, średnia ważona dla dwóch grup, które osiągnęły odpowiednio: 40 pkt. w grupie 25-osobowej i 50 pkt. w grupie 20-osobowej wynosi:

$$\bar{x} = \frac{(25) \cdot 40 + (20) \cdot 50}{25 + 20} = \frac{2000}{45} \approx 44,45$$

Kolejnym problemem jest nieuwzględnienie *rozrzutu wyników wokół średniej*. Najczęściej autorzy porównują wyniki nauczania mierzone w ocenach lub punktach. Nawet jeżeli są to dane pozyskane z jednakowych testów w dwóch grupach uczonych tą samą metodą i przez tę samą osobę, w podobnych warunkach,

przez tę samą ilość czasu, to przecież wzrost średniej, bez uwzględnienia miernika dyspersji, nie jest żadnym dowodem postępu, a tylko złudnym jego wrażeniem. Wyobraźmy sobie dla ułatwienia dwie dwudziestoosobowe grupy uczące się niemieckiego, które osiągnęły tę samą średnią na sprawdzianie, a więc teoretycznie zaznaczyły jednakowy postęp w nauce gramatyki. Tyle tylko, że na średnią w pierwszej grupie złożyło się dziesięć czwórek i dziesięć trojek, a na drugą dwadzieścia trojek. Obie grupy po roku nauczania poprawiły średnie o pół stopnia. Ale czy rzeczywiście można powiedzieć, że ta różnica jest istotna i oddaje rzeczywisty postęp uczniów? Jeśli chcemy się tego dowiedzieć, musimy uwzględnić różnice wokół średniej i policzyć *odchylenie standardowe*. Stosując miernik dyspersji uwzględniamy następującą relację: odchylenie standardowe określa różnicę między danymi a średnią arytmetyczną; innymi słowy pokazuje, w jaki sposób dane rozmieszczone są wokół średniej:

$$\delta = \sqrt{V}$$

$$V = \frac{(x - \bar{x})^2}{n}$$

To jeszcze nie wszystko. Trzeba przecież sprawdzić istotność zaobserwowanych różnic, do czego służą specjalne statystyczne wzory. Przytaczam tu jeden z częściej stosowanych przez dydaktyków test parametryczny służący do porównywania grup równolicznych, tzw. *test t Studenta*:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\text{odchylenie} \sqrt{\frac{1}{n_1} - \frac{1}{n_2}}}$$

gdzie

$$\text{odchylenie} = \sqrt{\frac{\delta_1^2(n_1 - 1) + \delta_2^2(n_2 - 1)}{(n_1 + n_2) - 2}}$$

\bar{x}_1 = średnia arytmetyczna pierwszej grupy

\bar{x}_2 = średnia arytmetyczna drugiej grupy

δ_1 = odchylenie standardowe pierwszej grupy

δ_2 = odchylenie standardowe drugiej grupy

n = liczebność grupy

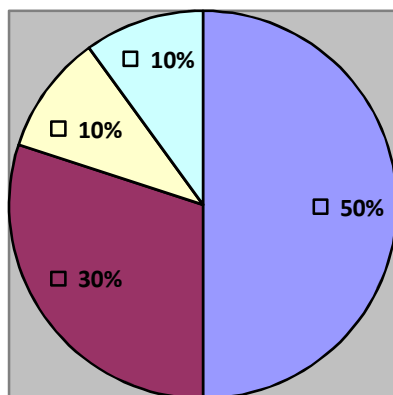
Po obliczeniu wartości (t) istotności różnic między porównywanymi średnimi, należy tę wartość porównać z wartością krytyczną (t_c) z tablicy statystycznej,

z uwzględnieniem stopnia prawdopodobieństwa (α) oraz liczby stopni swobody obliczanych według wzoru ($n_1 + n_2 - 2$), a także typu testu (jednostronnego lub dwustronnego). I znowu pojawia się problem interpretacji wyników na podstawie tablic statystycznych. Zdarza się, że możemy mówić o istotności lub jej braku przy różnych marginesach błędu i może należałoby to w raportach po prostu uczciwie zaznaczyć?

Przykład 2

Procenty a liczby.

Prawdą jest, że pozyskiwanie danych do najłatwiejszych zadań nie należy. Ale też prawdą jest, że nieraz lepiej pozyskać ich mniej niż sztucznie nadmuchiwać ilość probantów, by nie być posądzonym o badanie na zbyt małej próbie. Pomijam już dylematy związane z doбором grupy reprezentatywnej. Zauważam, że coraz częściej przedstawia się wyniki w formie diagramów kołowych lub słupkowych pokazujących tylko procentowy rozkład danych, bez ujawnienia wielkości próby. Ładniej przecież wygląda rozkład 50%/30%/10%/10%, jak to zilustrowano na rysunku 1, niż 5/3/1/1. Nieodparcie nasuwa się podejrzenie o próbę wywarcia lepszego wrażenia na odbiorcy:



Rysunek 1: Repartycja probantów w poszczególnych grupach wiekowych.

Przykład 3

Dobór przedziałów.

To wyjątkowo niebezpieczna i o wiele bardziej wyrafinowana metoda zaciemniania obrazu badanej sytuacji. Cel bywa osiągnięty na dwa różne sposoby: na etapie konceptualizacji i na etapie obróbki danych.

Na etapie konceptualizacji wiąże się to z reguły z trudnościami w pozyskiwaniu danych. Załóżmy, że autora opracowania interesują przedziały wiekowe,

a do reprezentantów tych przedziałów ma nierówny dostęp. Tworzy więc przedziały sztuczne, z pozoru logiczne, ale w istocie nie pozwalające na rzetelny ogłęd faktów. Jeśli probant „siedzi za ławką”, łatwiej zebrać dane na jego temat niż gdy studiuje, z kolei gdy studiuje, i tak prościej do niego dotrzeć niż gdy pracuje. A więc przedziały: dzieci-młodzież-dorośli nie pokrywają się np. ze szkołą podstawową, średnią i studiami tylko tworzy się je w przedziałach wiekowych: 7-12 13-18 18-23. Interwały pięcioletnie brzmią logicznie, ale należy zapytać, czy doświadczenia 19-latka z LO współgrają bardziej z doświadczeniami innego licealisty czy studenta? Przy czym to z reguły osoby z klas maturalnych wypełniają trzecią grupę, bo nie było jak zdobyć danych od starszych osób. I oto mamy doświadczenia i poglądy dorosłych referowane przez badacza ustami tych, co w dorosłość dopiero wkraczają. Podobne sytuacje zauważam w wypadku porównań studentów i osób pracujących do lat trzydziestu, gdzie luki wypełniane są danymi pochodzącymi od studentów ostatnich lat. Górę biorą ambicje badacza, który nie chcąc zrezygnować z jednej grupy, stwarza ją sztucznie.

Drugim sposobem, na etapie już samej obróbki danych, jest poszerzenie lub zawężenie interwałów tak, by nie pokazać niekorzystnych tendencji. Jest to dość prosty, ale trudny do wychwycenia zabieg. Załóżmy, że mamy wybory rektora i chcemy sprawdzić, ile osób z uprawnionych jednostek do nich przystąpiło. Zauważamy, że 5 instytutów mieści się w przedziale od 0-9,9% głosujących wśród uprawnionych, 12 instytutów mieści się w przedziale od 10%-19,9% głosujących wśród uprawnionych, 14 instytutów mieści się w przedziale od 20%-29,9% głosujących wśród uprawnionych, a 28 instytutów mieści się w przedziale od 30%-39,9% głosujących wśród uprawnionych. Czy nie lepiej będzie napisać, że w przedziale od 0%-19,9% mieszczą się głosujący wśród uprawnionych z 17 instytutów, a w przedziale od 20-39,9% mieszczą się głosujący wśród uprawnionych z 42 instytutów. Czy rzeczywiście jest to właściwy kompromis między zbytnią szczegółowością i nieuchronną stratą informacji przy małej liczbie przedziałów? Ten kompromis przyjmuje się w myśl zasady, że szerokość przedziału nie może przekraczać takiej różnicy, którą w naszym przekonaniu można pominąć.

4. Nieuprawnione porównania (korelacja)

Nieuprawnione porównania to przewinienie zwłaszcza tych, którzy decydują się na rachunek korelacji. Bo i skorelować da się niejedno. Anegdotyczny już wpływ spożycia ziemniaków na wzrost demograficzny w pewnym województwie da się łatwo przecież udowodnić statystycznie: oba wskaźniki sukcesywnie i proporcjonalnie wzrastają. A może te ziemniaki badano po przetworzeniu w stan płynny? Nie chcę, by brzmiało to okrutnie, ale to w fałszywych korelacjach wychodzi na jaw brak odczytania w literaturze przedmiotu. Badamy rzeczy zbadane, czynimy coś wyrzykowo, nie dość obserwujemy rzeczywistość i za słabo wprzegamy w rozumowanie... zdrowy rozsądek. A rachunek korelacji da efekty tylko pod warunkiem, że podłożymy do wzoru właściwe dane. Statystyka to nie matematyka,

w której nie ma miejsca na prognozowanie z założonym marginesem błędu. Wiele osób, niestety, obie te nauki błędnie utożsamia z powodu stosowanych w obu tych samych pojęć i symboli. Przytoczony wzór na rachunek korelacji:

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{(n-1) s_x s_y}$$

uwzględnia następujące czynniki:

n = liczba par $(x; y)$

$(x_i; y_i)$ = dane i pary

$$\sum_{i=1}^n x_i y_i = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$$

\bar{x} = średnia arytmetyczna zmiennej niezależnej

\bar{y} = średnia arytmetyczna zmiennej zależnej

s_x = odchylenie standardowe zmiennej niezależnej

s_y = odchylenie standardowe zmiennej zależnej

Wartość współczynnika korelacji interpretowana jest automatycznie, należy jednak pamiętać, że inaczej dla badań ścisłych, a inaczej dla badań społecznych.

Wreszcie sprawa wniosków. Często bywa, że nawet po lekturze serii bardzo dokładnych badań czytelnik zastanawia się, czemu służyła aż tak drobiazgowo diagnostyka, a ma prawo dowiedzieć się tego przed usłyszeniem referatu lub lekturą tekstu. Jeśli autor, dzięki wnikliwemu podejściu, zdola hipotezę potwierdzić lub odrzucić, zyskujemy pewną wiedzę i drobiazgowo przytoczenie kolejnych etapów postępowania ma głęboki sens – daje on szansę innym na powtórzenie eksperymentu w porównywalnych warunkach. Najlepiej jednak, gdy na podstawie wyników można coś odbiorcy zaproponować. W końcu koncepcja bez wdrożenia to dopiero połowa drogi, choć w badaniach podstawowych bywa, i to często, że teorię tworzy kto inny, a kto inny ją wciela w życie.

Czy należy się w tej sytuacji dziwić, że cytując wyniki badań statystycznych traktuje się nas potem dowcipami o sąsiadach-tyranach, z których każdy bijal żonę półtora raza w tygodniu lub o dziadku, który na wieczór kładł nogi na przemian do misek z gorącą i zimną wodą, więc można stwierdzić, że trzymał nogi w letniej wodzie? Albo że badamy zagadnienia wagi tak doniosłej, jak wpływ zaćmienia słońca na rozwój dorożkarstwa w Chinach? No i wreszcie, że wszyscy znamy cechy przeciętnego Polaka, ale jako żywo nikt się do takiej przeciętności nie przyznaje?

5. Wnioski

Przytoczone w niniejszym artykule przykłady, dotyczące typowych błędów pojawiających się w badaniach ankietowych, mogą posłużyć do zaprezentowania następujących wniosków:

1. Nie możemy, jak to często wyznają statystycy, udawać się do nich z pogrupowanymi danymi na dany temat i zapytaniem, co się z tego da przebadać. I choć może to zabrzmieć obrazoburczo, uwagi te nie są, niestety, głośne. Nie chcę przez to odwozić nikogo od konsultacji ze statystykami. Nie mogą i nie będą oni jednak nikogo *wyręczać na etapie konceptualizacji problemu, postawienia pytania badawczego, ułożenia kwestionariusza czy ankiety, rzetelnego zebrania danych, umiejętne ich pogrupowania.*
2. Specjalista, dysponując tak czytelną informacją, może wtedy rzeczywiście podpowiedzieć, *które testy i jakie formuły da się zastosować w konkretnym przypadku.* A wtedy kolejne etapy będą już mechaniczne do czasu *interpretacji wyników, w której znów nikt nas nie zastąpi.*
3. Wielkim ułatwieniem jest możliwość obliczeń statystycznych w arkuszach kalkulacyjnych, kalkulatorach i pakietach oprogramowania jak np. *Statistica.* One zrobią to za nas prawidłowo *pod przytoczonymi warunkami.*

Zapewniam, że wszystkie przykłady są autentyczne, a kosmetycznych przeróbek, nie mających wpływu na ich wydźwięk, dokonałam jedynie w trosce o zapobieżenie identyfikacji autorów. Zasmucające jest to, że w ogóle nie szukałam powyższych przykładów, tylko przytoczyłam te, z którymi ostatnio się zetknęłam. Zainteresowanych zachęcam do korzystania ze stosownej literatury, a w szczególności z dostępnych w języku polskim publikacji Brzezińskiego (2000), Fergusona i Takane (2002), Luszniwicza i Słabego (2001), Niemierki (2004), Openheima (2004), Pawłowskiego (1999) oraz Tadeusiewicza (2000). Źródła te, w moim przekonaniu, przyśpieszają tłumaczą zasady, na które zwróciłam uwagę z tym artykule.

BIBLIOGRAFIA

- Brzeziński, J., 2000. *Badania eksperymentalne w psychologii i pedagogice.* Warszawa: Scholar.
- Ferguson, G. A. i Takane, Y. 2002. *Analiza statystyczna w psychologii i pedagogice.* Warszawa: Wydawnictwo Naukowe PWN.
- Gajdzińska A. i Krzyżanowski P. (red). 1999. *Przeszłość w językowym obrazie świata.* Lublin: Wydawnictwo UMCS.
- Luszniwicz, A. i Słaby, T.: 2001. *Statystyka z pakietem komputerowym Statistica™ PL. Teoria i zastosowania.* Warszawa, C. H. Beck.
- Niemierko, B. 2004. *Pomiar wyników kształcenia.* Warszawa: WSiP.
- Openheim, A. N. 2004. *Kwestionariusze, wywiady, pomiary postaw.* Poznań: Zysk i S-ka.

- Pawłowski, A. 1999. „Metodologiczne podstawy wykorzystania słowników frekwencyjnych w badaniu językowego obrazu świata”, w: Gajdzińska A. i Krzyżanowski P. (red). 1999. 81-99.
- Tadeusiewicz, R. 2000. *Drogi i bezdroża statystyki w badaniach naukowych*. (http://www.statsoft.pl/czytelnia/badania_naukowe/d0ogol/nadrogi6.pdf).