

TYTUS SOSNOWSKI

## DOCEŃMY BADANIA EKSPŁORACYJNE\*

Mankamentem wielu artykułów polemicznych jest to, że są to dzieła *de omnibus rebus et quibusdam aliis*, czyli: „o wszystkich rzeczach i niektórych innych”, co sprawia, że problematyka dyskusji staje się bardzo rozproszona. Mankamentu tego nie ustrzegł się również prof. Brzeziński. Pozwolę więc sobie ograniczyć się tylko do jednego z wątków poruszonych w jego obszernym artykule.

„Zacznę od wyrażenia mojego przekonania, że badanie naukowe (a w przypadku psychologii ma ono charakter **empiryczny**) nie polega na wrzucaniu do komputera różnych wyników i dopasowywaniu do nich testów istotności oraz miar korelacji tak długo, aż «coś» z tego wyjdzie” – zaczyna swój artykuł prof. Brzeziński, i nieco dalej pisze: „Problemem jest bowiem powstawanie coraz większej liczby prac przyziemnie empirycznych, bez oddechu teoretycznego, bez wizji”.

Na czym może polegać „oddech teoretyczny” w badaniach psychologicznych? W najprostszym przypadku – na podejmowaniu badań mających na celu testowanie teorii. Ich przeciwieństwem są badania eksploracyjne, nastawione na poszukiwanie nowych zależności empirycznych. Tym dwom modelom badań odpowiadają w analizie wariancji (choć nie tylko tam) dwie metody testowania hipotez statystycznych: analiza kontrastów planowanych i analiza kontrastów *post hoc*.

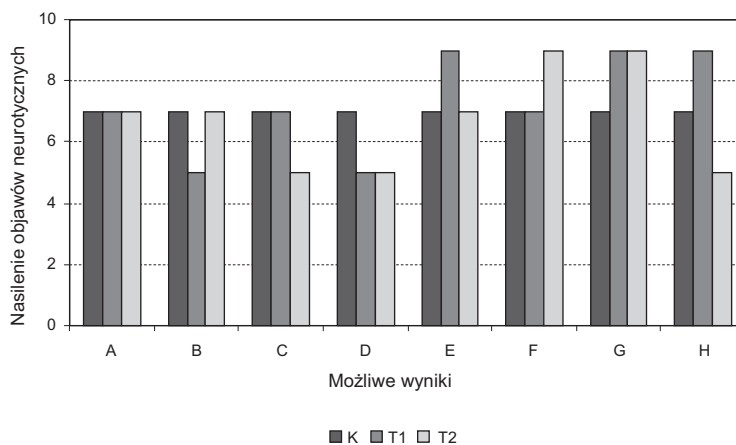
Przypuśćmy, że badacz zaproponował teorię wyjaśniającą mechanizm pewnego rodzaju zaburzeń neurotycznych i postawił na tej podstawie hipotezę, że

---

PROF. DR HAB. TYTUS SOSNOWSKI, Wydział Psychologii Uniwersytetu Warszawskiego, ul. Stawki 5/7, 00-183 Warszawa; e-mail: Tytus@psych.uw.edu.pl

\* Praca finansowana z programu badawczego BST 154526.

określony rodzaj terapii (T1) będzie łagodzić objawy neurotyczne, natomiast inny, konkurencyjny rodzaj terapii (T2) efektu takiego w ogóle nie będzie wywierać. Dla sprawdzenia trafności swej teorii badacz przeprowadził eksperyment w trzech grupach zrandomizowanych, z których jedna, kontrolna (K), nie została poddana żadnemu oddziaływaniu, druga – została poddana terapii metodą T1, a trzecia – terapii metodą T2. Rysunek 1 przedstawia osiem różnych wyników takiego eksperymentu (dla uproszczenia założmy, że każda różnica między średnimi, widoczna na rysunku, jest istotna statystycznie). Tylko jeden wynik (A) wskazuje na brak jakichkolwiek różnic między średnimi, ale też tylko jeden (B) stanowi potwierdzenie hipotezy badacza; pozostałe są z nią niezgodne. Przeglądając się rysunkowi widać doskonale, jak mało przydatny do testowania hipotez teoretycznych jest uogólniony test F (omnibus F test). Test ten służy do oszacowania hipotezy zerowej mówiącej, że wszystkie średnie są w populacji równe, w naszym przypadku – hipotezy  $H_0: \mu_1 = \mu_2 = \mu_3$ . Istotność testu F mówi nam więc tylko tyle, że istnieją jakieś istotne różnice między średnimi, nie odpowie natomiast na pytanie, czy różnice te są zgodne z hipotezą czy też z nią niezgodne.



Rysunek 1. Osiem układów wyników (A do H) w przykładowym eksperymencie trzygrupowym

Chcąc ocenić słuszność naszej teorii musimy sprawdzić, czy efekt każdej z dwóch metod terapii jest taki, jak wynika to z teorii. Do oceny trafności takich

przewidywań przeznaczone są kontrasty planowane<sup>1</sup> (Ferguson, Takane, 1997; King, Minium, 2009; Kirk, 1995; Rosenthal, Rosnow, Rubin, 2000; Winer, Brown, Michels, 1991; por. też Sosnowski, 2004, 2010). W naszym wypadku moglibyśmy zdefiniować dwa takie kontrasty: (1) porównanie grupy T1 z dwiema pozostałymi grupami dla sprawdzenia, czy metoda T1 daje wyniki lepsze niż brak terapii lub metoda T2, oraz (2) porównanie grupy K z grupą T2 dla sprawdzenia, czy efekt terapii T2 jest taki sam jak efekt braku terapii. Definiowanie kontrastów planowanych ma tę zaletę, że wymusza bardzo precyzyjne formułowanie przewidywań. Kontrasty takie mają też większą moc statystyczną niż kontrasty *ex post*. No i wreszcie, stosując kontrasty planowane, możemy zrezygnować z testu F. Dyrektywa ta była przez długi czas przyjmowana z dużą nieufnością przez badaczy, ale dziś można ją spotkać nawet w podręcznikach statystyki na poziomie podstawowym, żeby wymienić chociażby podręcznik Kinga i Minium (2009), przetłumaczony na język polski pod redakcją naukową prof. Brzezińskiego.

Przy okazji warto zastanowić się nad jeszcze jedną kwestią: co faktycznie testujemy w opisanym wyżej eksperymencie? Stosując test F, testujemy efekt trójkategorialnej zmiennej „metoda terapii”, stosując zaś kontrasty planowane, testujemy dwa efekty oddziaływań eksperymentalnych (metod terapii): T1 i T2. Zauważmy, że gdybyśmy analizę tych samych danych przeprowadzili za pomocą wielokrotnej analizy regresji (a analiza wariancji może być traktowana jako szczególny przypadek analizy regresji), oba kontrasty byłyby włączone do równania regresji jako dwa odrębne predyktory, a więc, *de facto*, jako dwie zmienne niezależne (por. Cohen i in., 2003; Pedhazur, 1982). Podobne ujęcie efektów zmiennych kategorialnych znajdziemy w modelu strukturalnym ANOVA (por. Winer i in., 1991).

Opisaną wyżej procedurę statystyczną można by uznać za wzorcową dla testowania hipotez teoretycznych w analizie wariancji. Można ją też polecić do analizy danych w badaniach replikacyjnych; tam również interesuje nas pytanie, czy otrzymane wyniki są dokładnie zgodne z oczekiwaniami, tj. wynikami wcześniejszych badań. W praktyce jednak opisany wyżej model analizy stosowany jest niezbyt często. Przyczyn takiego stanu rzeczy należy szukać jednak nie tylko w kompetencji statystycznej psychologów, ale również w stanie teorii psychologicznych.

---

<sup>1</sup> Analiza kontrastów ma zastosowanie wtedy, gdy zmienna niezależna ma więcej niż dwa poziomy. W przeciwnym wypadku wystarcza test F lub test t.

Aby można było przetestować teorię w opisany wyżej sposób, musi być ona bardzo jasno sformułowana i mieć jednoznaczną operacjonalizację, tak aby było dokładnie wiadomo (przed rozpoczęciem badania), jakie dane empiryczne będą zgodne z teorią, a jakie będą z nią niezgodne. Gdybyśmy mieli do czynienia z taką sytuacją, szersza dyskusja wyników byłaby właściwie niepotrzebna. Wystarczyłoby stwierdzenie, że otrzymane wyniki potwierdziły teorię lub też jej nie potwierdziły. Taki przypadek nieczęsto się jednak zdarza<sup>2</sup>. Przeciwnie, często okazuje się, że wyniki są „tylko częściowo zgodne z hipotezą”. Albo też, że chociaż wyniki nie są zgodne z przewidywaniami, to niekoniecznie są sprzeczne z teorią, bo na przykład nie kontrolowaliśmy dobrze jakichś zmiennych ubocznych, nie zoperacjonalizowaliśmy właściwie zmiennych eksperymentalnych lub popełniliśmy jakieś inne błędy metodologiczne. Gdyby nie te błędy, wynik mógłby być zgodny z teorią, a tak... Takim wynikiom towarzyszy zwykle obszerna dyskusja, która jest po prostu interpretacją wyników *ex post*. Postępujemy więc wówczas tak, jakbyśmy przeprowadzili badania eksploracyjne, uzyskali nieprzewidywany wcześniej wynik empiryczny i zastanawiali się po fakcie nad jego interpretacją teoretyczną.

Autorzy takich badań unikają jednak zazwyczaj prezentowania ich *explicite* jako badań eksploracyjnych. Wynika to, jak sądzę, z szeroko rozpowszechnionego przekonania, że badania eksploracyjne są metodologicznie „gorsze” od testowania hipotez. Przekonanie to z kolei jest źródłem presji, aby za wszelką cenę formułować jakieś hipotezy teoretyczne lub przynajmniej operacyjne, nawet gdyby miały to być hipotezy trywialne, w rodzaju: „zgodnie z naszą hipotezą, zmienna X wpływa na zmienną Y”.

Tymczasem nie ma powodu, aby deprecjonować badania eksploracyjne. Ich wynikiem może być opis ważnych faktów lub odkrycie ciekawych zależności empirycznych. Zależności takie mogą być z kolei inspiracją do budowy teorii – coś może być silniejszym impulsem do poszukiwania teorii, jeśli nie dobrze potwierdzona zależność empiryczna, która nie da się wywnioskować z żadnej z istniejących teorii lub jest wręcz z tymi teoriami sprzeczna? Ogromne znaczenie do tego typu faktów, nazwanych anomaliami, przywiązywał – jak wiadomo – Kuhn (2001). Warto też pamiętać, że wiele ważnych odkryć naukowych powstało przypadkowo, przy okazji badań nastawionych na inne cele. Moim ulubionym przykładem jest tu przypadek Marii Skłodowskiej-Curie, która szukając

---

<sup>2</sup> Gwoli sprawiedliwości należy przyznać, że psychologowie stosują często złożone modele badawcze i testują jednocześnie wiele różnych hipotez; w takim wypadku krótkie i jednoznaczne podsumowanie wyników może być rzeczywiście trudne.

w smółce uranowej pierwiastka promieniotwórczego, odkryła nieoczekiwanie dwa takie pierwiastki: rad i polon (Curie, 1979).

Warto więc przyjąć do akceptującej wiadomości, że wobec ogromnej złożoności zjawisk, jakimi zajmuje się psychologia, i przy obecnym stanie jej teorii, jest w psychologii miejsce na badania eksploracyjne (por. też Sternberg, 2011) i nie ma powodu, aby badania takie sztucznie „stroić w szaty” testowania hipotez. Co oczywiście nie umniejsza w niczym doniosłości tworzeniu teorii i ich testowaniu.

#### LITERATURA CYTOWANA

- Cohen, J., Cohen, P., West, S. G., Aiken, L. S. (2003<sup>3</sup>). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Curie, E. (1979). *Maria Curie*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Ferguson, G. A., Takane, Y. (1997). *Analiza statystyczna w psychologii i pedagogice*. Warszawa: Wydawnictwo Naukowe PWN.
- King, B. M., Minium, E. W. (2009). *Statystyka dla psychologów i pedagogów*. Warszawa: Wydawnictwo Naukowe PWN.
- Kirk, R. E. (1995<sup>3</sup>). *Experimental design: Procedures for the behavioral sciences*. Pacific Grove: Brooks/Cole Publishing Company.
- Kuhn, T. S. (2001). *Struktura rewolucji naukowych*. Warszawa: Fundacja Aletheia.
- Pedhazur, E. J. (1982<sup>2</sup>). *Multiple regression in behavioral research*. New York: Holt, Rinehart and Winston.
- Rosenthal, R., Rosnow, R. L., Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge: Cambridge University Press.
- Sosnowski, T. (2004). Analiza kontrastów: między eksploracją a testowaniem hipotez. *Przegląd Psychologiczny*, 47, 4, 367-378.
- Sosnowski, T. (2010). Analiza interakcji zmiennych kategoryalnych i ciągłych. *Psychologia Społeczna*, 5, 2-3, 162-190.
- Sternberg, R. J. (2011). *Recenzowanie prac naukowych w psychologii*. Warszawa: Paradygmat.
- Winer, B. J., Brown, D. R., Michels, K. M. (1991<sup>3</sup>). *Statistical principles in experimental design*. Boston, MA: McGraw Hill.