

**PIOTR MACIEJ SKORUPIŃSKI**

Warszawa

## **KORELACJA WYNIKÓW POMIARÓW**

W 1897 r. – w 187. tomie renomowanego periodyku założonego w 1660 r. Królewskiego Towarzystwa w Londynie (Royal Society of London) – Karl Pearson (1857–1936) określił korelację następująco: *Powiada się, że dwa organy u tej samej jednostki albo u związanej pary jednostek są skorelowane, kiedy przy serii pierwszego organu o określonej wielkości, średnia rozmiarów korespondującego drugiego organu jest funkcją rozmiaru organu pierwszego*<sup>1</sup>. Tej zgrzebnej formule, powstałej w kręgu idei Francisca Galtona (1822–1911), towarzyszył elegancki wzór (1.1), umożliwiający matematyczne sprecyzowanie korelacji pomiędzy dwoma organami rozumianymi jako każda mierzalna właściwość organizmu:

$$r = S(xy)/(n\sigma_1\sigma_2), \quad (1.1)$$

gdzie:

$r$  – symbol współczynnika korelacji (*coefficient of correlation*);  $S$  – znak sumy;  $x$  i  $y$  – odchylenia od średniej wartości zmiennych  $X$  i  $Y$ ;  $n$  – liczba par w próbie;  $\sigma_1$  i  $\sigma_2$  – odchylenia standardowe (*standard deviation*) wartości zmiennych  $X$  i  $Y$ <sup>2</sup>.

W 1904 r. tę samą formułę określania relacji pomiędzy zmiennymi wykorzysta Edward L. Thorndike (1874–1949) w pierwszym podręczniku teorii pomiaru w naukach społecznych, poprzedzając jej wprowadzenie interesującym komentarzem: *[...] w studiach nad dziedzicznością potrzeba miary podobieństwa między synami i ojcami, pomiędzy synami i dziadkami, a także pomiędzy synami a innymi osobnikami płci męskiej, wylosowanymi z tej samej rasy, oraz ilościowego porównania tych trzech podobieństw. Tak samo w diagnostyce edukacyjnej [...] potrzeba miary podobieństwa pomiędzy inteligencją i zdolnościami chłop-*

---

<sup>1</sup> Pearson K., *Mathematical contributions to the theory of evolution – III. Regression, heredity, and panmixia*, „Philosophical transactions of the Royal Society of London. Series A” 1897, t. 187., s. 256–257.

<sup>2</sup> Por. ibidem, s. 265.

ców a ich osiągnięciami szkolnymi [...], a także pomiędzy innymi ich charakteryzującymi faktami<sup>3</sup>. Od tego czasu tzw. współczynnik Pearsona według momentu iloczynowego (mieszanego)<sup>4</sup> – *the product-moment correlation coefficient* – stanowi standardowy składnik klasycznego wykładu metodologii pomiaru edukacyjnego<sup>5</sup>. Co więcej, okaże się szczególnie użyteczny w połowie ubiegłego stulecia – w czasach dominacji kryterialnego (korelacyjnego) podejścia do problematyki trafności – mimo sugestii Stanleya S. Stevensa (1906–1973), iż *statystyczne manipulacje, którym zasadnie można poddać dane empiryczne, zależą od typu skali zastosowanej do uporządkowania tych danych*<sup>6</sup>.

Taksonomia pomiarów i związana z nią hierarchia czterech skal pomiarowych (skale: nominalna, porządkowa, przedziałowa i stosunkowa), przedstawione przez Stevensa w jego *Teorii skal pomiarowych (On the theory of scales of measurement)*, również stały się ważnym – i kontrowersyjnym – składnikiem tradycyjnej metodologii pomiarowej<sup>7</sup>. Kontrowersyjnym, ponieważ stanowiącym, iż jedynie pomiar z wykorzystaniem skali przedziałowej (bądź stosunkowej) upoważnia do manipulowania uzyskanymi w ten sposób danymi za pomocą takich formuł, jak np.: średnia (arytmetyczna), odchylenie standardowe czy współczynnik Pearsona według momentu iloczynowego (mieszanego)<sup>8</sup>. To zasadnicze ograniczenie działań analitycznych i interpretacyjnych jest wyjątkowo kłopotliwe w dziedzinie diagnostyki edukacyjnej, gdyż zarówno stopnie szkolne, jak wyniki testów osiągnięć są – wedle wczesnych nauk Lordowskiego Profesora X – liczbami jedynie porządkowymi<sup>9</sup>.

Jednakże Profesor X – jak pamiętamy – zmienił zdanie, ponieważ zrozumiał, że *liczby nie pamiętają, skąd pochodzą, toteż zawsze zachowują się w ten sam sposób*<sup>10</sup>, i od tamtej pory nie zamykał się już w swoim gabinecie, aby w ukryciu przed studentami rozkoszować się obliczaniem średnich i odchyłeń standardowych. Nie znaczy to oczywiście, iż głos Frederica M. Lorda (1912–2000) w dyskusji toczącej się wokół taksonomii pomiarów i skal zakończył ten wieloletni spór. Trzydzieści lat później John Gaito na łamach „Biuletynu Psychologicznego” Amerykańskiego Towarzystwa Psychologicznego (American Psychological Association) ubolewać będzie, że wciąż w podręcznikach metodologii

<sup>3</sup> Thorndike E.L., *An introduction to the theory of mental and social measurements*, II wyd. popr., New York 1916, s. 156.

<sup>4</sup> Por. Ferguson G.A., Takane Y., *Analiza statystyczna w psychologii i pedagogice*, wyd. II, PWN, Warszawa 1999, s. 133.

<sup>5</sup> Por. Popham W.J., *Educational statistics. Use and interpretation*, New York 1967, s. 68–78; Niemierko B., *Pomiar wyników kształcenia*, Warszawa 1999, s. 155 i 200–201; Konarzewski K., *Jak uprawiać badania oświatowe. Metodologia praktyczna*, Warszawa 2000, s. 186–189.

<sup>6</sup> Stevens S.S., *On the theory of scales of measurement*, „Science” 1946, t. 103., s. 677.

<sup>7</sup> Por. Popham W.J., op. cit., s. 271–273; Niemierko B., op. cit., s. 51; Brzeziński J., *Metodologia badań psychologicznych*, Warszawa 1997, s. 188–189.

<sup>8</sup> Por. Stevens S.S., op. cit., s. 678.

<sup>9</sup> Por. Lord F.M., *On the statistical treatment of football numbers*, „American Psychologist” 1953, t. 8., s. 750.

<sup>10</sup> Ibidem, s. 751.

i statystyki wykorzystywana jest myląca koncepcja Stevensa<sup>11</sup>. I mimo że Gaito przypomniał dzieje krytyki tej idei i kilkakrotnie podkreślił, iż *skale pomiarowe nie są powiązane z technikami statystycznymi*<sup>12</sup> i że specyficzne właściwości skali nie stanowią warunku wykorzystania określonej procedury statystycznej<sup>13</sup>, to prawie 15 lat później Paul F. Velleman i Leland Wilkinson w ogniu tej samej polemiki powtórzą: *Chociaż proponujemy wiele argumentów [na przykład dane, które nie pasują do żadnej ze skal z typologii Stevensa, czy alternatywne taksonomie skal pomiarowych – przyp. aut.], jedynym jednoczącym argumentem przeciw przypisywaniu statystyk typowi skali jest fakt, że to nie działa*<sup>14</sup>.

A zatem czy tzw. współczynnik Pearsona według momentu iloczynowego (mieszanego) – chociażby w formule (1.2)<sup>15</sup>:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} \quad (1.2)$$

(gdzie:  $x$  i  $y$  – odchylenia od średniej wartości zmiennych  $X$  i  $Y$ ) – można zasadnie stosować w celu poznania związku pomiędzy wynikami pomiarów edukacyjnych, dokonanych z użyciem różnych skal? Np. skali stopni szkolnych i skal egzaminacyjnych, funkcjonujących podczas egzaminów doniosłych: gimnazjalnego i maturalnego? I czy uzasadnione jest porównywanie wartości współczynnika Pearsona, charakteryzujących relacje pomiędzy wynikami egzaminu, np. maturalnego z języka polskiego a stopniami z tego przedmiotu, uzyskanymi przez zdających w poprzedzających ich egzamin szkolnych klasyfikacjach? Porównywanie, które chociażby umożliwiłoby zaprezentowane w tab. 1. zestawienie?

Tab. 1. przedstawia wartości współczynnika Pearsona według momentu iloczynowego (mieszanego)<sup>16</sup>, otrzymane z porównania 11 zbiorów wyników uzyskanych przez 246 badanych<sup>17</sup>: najpierw kandydatów ubiegających się – w czerwcu 2002 r. – o miejsce w renomowanej warszawskiej szkole ponadgimnazjalnej<sup>18</sup>, potem jej uczniów i – od 15 kwietnia 2005 r. – absolwentów. Na wspomniane zbiory złożyły się:

<sup>11</sup> Por. Gaito J., *Measurement scales and statistics: resurgence of an old misconception*, „Psychological Bulletin” 1980, t. 87., nr 3, s. 564.

<sup>12</sup> Ibidem.

<sup>13</sup> Por. ibidem, s. 566.

<sup>14</sup> Velleman P.F., Wilkinson L., *Nominal, ordinal, interval, and ratio typologies are misleading*, „The American Statistician” 1993, t. 47., nr 1, s. 68.

<sup>15</sup> Por. Thorndike E.L., op. cit., s. 173.

<sup>16</sup> Obliczenia wykonano w programach SPSS 11.5.0 oraz Microsoft Excel 2002.

<sup>17</sup> Por. przyp. 13. W każdej komórce tab. 1. została podana liczba zestawów wziętych pod uwagę w obliczeniach.

<sup>18</sup> Wspomniane liceum ogólnokształcące zostało scharakteryzowane w pierwszym z cyklu artykułów poświęconych problemowi korelacji wyników egzaminu zewnętrznego ze stopniami szkolnymi, rozpatrywanemu na podstawie studium przypadku tej właśnie szkoły {por. Skorupiński P.M., *Trafność interpretacji wyników egzaminu gimnazjalnego w rekrutacji kandydatów do publicznej szkoły ponadgimnazjalnej*, [w:] B. Niemierko, J. Brzdąk (red.), *Dwa rodzaje oceniania szkolnego. Ocenianie wewnątrzszkolne i zewnętrzne a jakość pracy szkoły*, Katowice

Tab. 1. Macierz korelacji pomiędzy wynikami pomiarów

			Wynik egz. gimnazjalnego		Stopnie z j. pol. w kolejnych klasyfikacjach								Wynik egz. mat. z j. pol.	
			część		III kl. gim.	I klasa liceum		II klasa liceum		III klasa liceum		część		
			hum	mat.-przyr.	II sem.	I sem.	II sem.	I sem.	II sem.	I sem.	II sem.	ustna	pisemna	
Wynik egz. gimn.	część	hum.	<b>1</b> n=246	<b>0,21</b> n=246	<b>0,26</b> n=244	<b>0,26</b> n=217	<b>0,34</b> n=246	<b>0,37</b> n=245	<b>0,27</b> n=245	<b>0,39</b> n=246	<b>0,37</b> n=246	<b>0,26</b> n=246	<b>0,31</b> n=246	
		mat.-przyr.	<b>1</b> n=246	<b>0,28</b> n=244	<b>0,25</b> n=217	<b>0,23</b> n=246	<b>0,29</b> n=245	<b>0,27</b> n=245	<b>0,17</b> n=246	<b>0,17</b> n=246	<b>0,09</b> n=246	<b>0,24</b> n=246		
Stopnie z j. pol. w kolejnych klasyfikacjach	III kl. gimn.	II sem.	<b>1</b> n=244	<b>0,44</b> n=215	<b>0,45</b> n=244	<b>0,45</b> n=243	<b>0,40</b> n=243	<b>0,36</b> n=244	<b>0,45</b> n=244	<b>0,44</b> n=244	<b>0,44</b> n=244	<b>0,33</b> n=244		
		I sem.	<b>1</b> n=217	<b>0,65</b> n=217	<b>0,50</b> n=217	<b>0,52</b> n=217	<b>0,44</b> n=217	<b>0,49</b> n=217	<b>0,46</b> n=217	<b>0,46</b> n=217	<b>0,29</b> n=217			
	I klasa liceum	I sem.	<b>1</b> n=246	<b>0,71</b> n=245	<b>0,70</b> n=245	<b>0,62</b> n=246	<b>0,67</b> n=246	<b>0,50</b> n=246	<b>0,29</b> n=246					
		II sem.	<b>1</b> n=245	<b>0,77</b> n=245	<b>0,66</b> n=245	<b>0,68</b> n=245	<b>0,51</b> n=245	<b>0,26</b> n=245						
	II klasa liceum	I sem.	<b>1</b> n=245	<b>0,77</b> n=245	<b>0,66</b> n=245	<b>0,68</b> n=245	<b>0,51</b> n=245	<b>0,26</b> n=245						
		II sem.	<b>1</b> n=245	<b>0,73</b> n=245	<b>0,73</b> n=245	<b>0,51</b> n=245	<b>0,29</b> n=245							
	III klasa liceum	I sem.	<b>1</b> n=246	<b>0,79</b> n=246	<b>0,47</b> n=246	<b>0,32</b> n=246								
		II sem.	<b>1</b> n=246	<b>0,56</b> n=246	<b>0,37</b> n=246									
	Wynik egz. mat. z j. pol.	część	ustna	<b>1</b> n=246	<b>0,32</b> n=246									
			pisemna	<b>1</b> n=246										

- roczne oceny klasyfikacyjne z języka polskiego, ustalone – według 6-stopniowej skali, wskazanej w rozporządzeniu ministra właściwego do spraw oświaty i wychowania, regulującym między innymi warunki oraz sposób oceniania, klasyfikowania i promowania – po trzeciej klasie gimnazjum w roku szkolnym 2001/2002,

2002, s. 45. Także por. Skorupiński P.M., *Kryterialny wymiar trafności interpretacji wyników egzaminu gimnazjalnego*, [w:] B. Niemierko (red.), *Trafność pomiaru jako podstawa obiektywizacji egzaminów szkolnych*, Łódź 2003, s. 220}. W celu uzupełnienia tej charakterystyki należy dodać, że w roku szkolnym 2004/2005 do 9 klas trzecich uczęszczało 296 uczniów, natomiast do egzaminu maturalnego nie przystąpiło – z powodu nieukończenia szkoły – 2 spośród nich. W tab. 1. nie uwzględniono 48 absolwentów liceum z powodu braku dostępu do ich wyników z egzaminu gimnazjalnego (por. Skorupiński P.M., *Trafność interpretacji...*, op. cit., s. 45–46).

- wyniki egzaminu gimnazjalnego 2002 r.<sup>19</sup> (wyrażone w 50-punktowej skali surowych wyników testowania),
- śródroczne i roczne oceny klasyfikacyjne z języka polskiego, ustalone – według 6-stopniowej skali, scharakteryzowanej w przywołanym wyżej rozporządzeniu – w kolejnych latach szkolnych: 2002/2003, 2003/2004 i 2004/2005,
- wyniki egzaminu maturalnego z języka polskiego (według nowej formuły)<sup>20</sup> – wyrażone w dwóch skalach wyników surowych: 20-punktowej (część ustna) oraz 70-punktowej (część pisemna).

W tab. 1. podkreślone zostały te wartości współczynnika Pearsona, które oznaczać mogą wysoką (powyżej 0,7) lub umiarkowaną (powyżej 0,5) korelację między analizowanymi zmiennymi. W takiej właśnie relacji znajdują się stopnie szkolne, wystawione w kolejnych klasyfikacjach w szkole ponadgimnazjalnej. Co więcej, z semestru na semestr związek ten znacząco się zacieśnia, osiągając poziom współczynnika o wartości 0,79 w przypadku stopni z I oraz II semestru trzeciej klasy. Taki układ miar korelacji stanowi – być może – odbicie **procesu ustalania (stabilizacji) oceny szkolnej**. W tym kontekście niepokoi niska współzależność między tą oceną a wynikiem egzaminu zewnętrznego (poniżej wartości 0,4). Ustna część egzaminu maturalnego, skorelowana ze stopniami ze świadectwa ukończenia szkoły na poziomie współczynnika o wartości 0,56, traci przecież swój doniosły charakter również z powodów zilustrowanych w tab. 2.

Tab. 2. Stopnie i wyniki egzaminu z języka polskiego

	Stopień na zakończenie III kl.	Wynik ustnego egzaminu maturalnego	Wynik pisemnego egzaminu maturalnego
Uczeń 1	2	12	31
Uczeń 2	2	20	35
Uczeń 3	3	12	39
Uczeń 4	3	20	33
Uczeń 5	4	11	24
Uczeń 6	4	20	32
Uczeń 7	5	12	39
Uczeń 8	5	20	25

<sup>19</sup> Egzamin ten odbył się 14 maja [egzamin z zakresu przedmiotów humanistycznych – arkusz *Lądy archeologów*, o symbolu GH-A1-021; por. Centralna Komisja Egzaminacyjna, *Egzamin w trzeciej klasie gimnazjum z zakresu przedmiotów humanistycznych. Lądy archeologów*, dokument dostępny w internecie pod adresem: [www.cke.edu.pl/podstrony/archiwum/gimn\\_2002/a1h.pdf](http://www.cke.edu.pl/podstrony/archiwum/gimn_2002/a1h.pdf) (20 lipca 2005 r.)] oraz 15 maja (egzamin z zakresu przedmiotów matematyczno-przyrodniczych) 2002 r.

<sup>20</sup> Egzamin ten odbył się w dniach od 18 kwietnia do 11 maja (część ustna) oraz 5 maja (część pisemna) 2005 r. Dla celów niniejszego studium wykorzystano jedynie wyniki tzw. poziomu podstawowego pisemnej części egzaminu [*Arkusz I*, o symbolu MPO-P1A1P-052, zawierający tekst K. Szymborskiego pt.: *Dziki słówka* – por. Centralna Komisja Egzaminacyjna, *Egzamin maturalny z języka polskiego. Arkusz I. Poziom podstawowy*, dokument dostępny w internecie pod adresem: [www.cke.edu.pl/images/stories/Matura2005/polsi\\_podst\\_a1.pdf](http://www.cke.edu.pl/images/stories/Matura2005/polsi_podst_a1.pdf) (20 lipca 2005 r.)].

W świetle tab. 2. oraz – przede wszystkim – przedstawionego powyżej układu relacji pomiędzy stopniami szkolnymi a wynikami egzaminów zewnętrznych nasuwa się oczywiście pytanie o właściwą miarę osiągnięć edukacyjnych ucznia – pytanie szczególnie doniosłe w trakcie procesów decyzyjnych, skutkujących selekcją na kolejnych progach szkolnych. Podstawą oceny jakości tej odpowiedniej miary mogłaby się stać m.in. pogłębiona analiza korelacyjna, uwzględniająca wyniki trafnych i rzetelnych pomiarów kryterialnych. Co jednak począć, gdy właściwości dostępnych narzędzi, a więc i znaczenie zebranych dzięki nim danych, pozostają w istotnej części niejasne? Jak w takiej sytuacji interpretować uzyskane wartości współczynnika Pearsona? Wydaje się, że bez wiedzy o przedmiotach, które – wedle naszego mniemania – mierzymy, nie jesteśmy w stanie stosownie wykorzystać prostych – jak to ujął Thorndike – miar podobieństwa.