

Sławomir Sapanowski

Okręgowa Komisja Egzaminacyjna w Łodzi

Dlaczego nie lubię skali staninowej?

Podczas poprzedniej Konferencji Diagnostyki Edukacyjnej w Kielcach (2009) postawiłem przewrotne pytanie: *Czy staniny muszą odejść?*, które zresztą stało się tematem jednego z paneli dyskusyjnych. Formułując to pytanie, miałem zamiar przedyskutować potrzebę wprowadzenia doskonalszego narzędzia do komunikowania wyników i do ich porównywania w kolejnych latach. Niestety, przeżyłem zawód i rozczarowanie. Większość osób obecnych na panelu dyskusyjnym była przekonana o braku potrzeby zmian. Skala staninowa postrzegana była bez wad, które mogłyby ją wyeliminować jako główny sposób zrównywania wyników egzaminów zewnętrznych. Nawet więcej, podkreślano jej niezliczone zalety zarówno prawdziwe, jak fikcyjne. Wszystko to spowodowało dylemat: a może się mylę, a może nie mam racji, postulując odsunięcie staninów do lamusa zużytych i niepotrzebnych już wskaźników osiągnięć egzaminacyjnych.

Odkąd istnieją badania statystyczne, odtąd pojawiały się próby normalizacji i standaryzacji wyników tychże badań. Owocem jednej z takich prób jest skala dziewięciostopniowa, która opisywała cechę populacji podlegającą rozkładowi normalnemu. Jeśli rozkład odbiegał od rozkładu Gaussa, skala staninowa nie mogła być stosowana. A jakie są rozkłady wyników egzaminów zewnętrznych? Jeśli chodzi o sprawdzian i egzamin gimnazjalny, to owe rozkłady są zawsze mniej lub bardziej skośne, a zatem odbiegają od normalności. Dlaczego tak się dzieje? Przyczyn jest oczywiście dużo, ale jedną z głównych jest to, że suma rozkładów normalnych¹ nie musi dać rozkładu Gaussa. A przecież wynik egzaminu jest sumą uzyskanych rezultatów za rozwiązanie poszczególnych zadań.

Niestety, pojawiają się kolejne sygnały świadczące o niedoskonałości omawianej skali. Rozważmy przypadek czwórki uczniów: Adama, Basi, Celiny i Daniela. W 2005 roku pisali oni sprawdzian w szóstej klasie szkoły podstawowej i uzyskali wyniki przedstawione w poniższej tabeli.

Uczeń	Wynik
Adam	24
Basia	28
Celina	37
Daniel	39

Osoba zainteresowana (rodzic, uczeń) może rozumować następująco: Basia uzyskała wynik lepszy od Adama o 4 punkty, a z kolei Daniel uzyskał o 2 punkty więcej od Celiny, w związku z tym Basia „bardziej” wyprzedziła Adama niż Daniel Celinę. Być może taki sposób podejścia do wyników nie zachwyca finezją, ale dla większości beneficjentów systemu oświaty będzie sensowny.

¹ Łatwo sprawdzić, że rozkład wyników za pojedyncze zadanie jest rozkładem zbliżonym do normalnego.

Czyż nie? Nic bardziej mylnego! Spójrzmy, jak wyglądają te wyniki w opublikowanej dla tego rocznika² skali staninowej.

Uczeń	Stanin
Adam	4
Basia	4
Celina	7
Daniel	9

Zdumiewające! Wyniki Adama i Basi znalazły się w tym samym staninie, a więc są wynikami porównywalnymi! Z kolei wyniki Daniela i Celiny różnią się aż o dwa staniny³, czyli o całe odchylenie standardowe. Toż to cała przepaść.

Podobne zestawienie dla średniej szkół dla Sprawdzianu 2007 przynosi równie niespodziewane wnioski.

	Średnia	Stanin
Szkoła A	20,9	2
Szkoła B	22,6	2
Szkoła C	25,4	4
Szkoła D	26,9	6

Szkoła B uzyskała wyższą średnią od szkoły A o 1,7 punktu i znalazła się w tym samym staninie (2), a więc obie szkoły mają porównywalne wyniki. Z kolei szkoła D osiągnęła średnią tylko o 1,5 punktu wyższą od szkoły C, a mimo to w skali staninowej dzieli je różnica dwóch stopni – to dużo. Takie sposoby interpretacji i analizy średniej oraz pozycji na skali staninowej mogą budzić zastrzeżenia, ale niewątpliwie mają prawo pojawić się w praktyce. Nasuwa się pytanie, czy podane przykłady nie wynikają przypadkiem ze stronniczości autora i jego niechęci do standardowej dziewiątki? W związku z czym dobrane zostały tendencyjnie, a cały problem jest marginalny i niewart uwagi?

Przyjrzyjmy się, jak ma się położenie szkoły na skali dziewięciostopniowej z uwzględnieniem przedziału ufności dla średniej.

Przedmiotem mojej analizy były wyniki szkół podstawowych (Sprawdzian 2009) z terenu Okręgowej Komisji Egzaminacyjnej w Łodzi (woj. łódzkie i świętokrzyskie). Wstępnie odrzuciłem szkoły z liczbą uczniów mniejszą⁴ od 10. Dla pozostałych szkół w liczbie ok. 1100 obliczyłem szerokość przedziału ufności dla średniej. Najmniejsza wartość wynosiła ± 1 , największa zaś ± 6 . Uporządkowany rozkład szerokości przedziałów ufności dla średniej szkoły przedstawia poniższy wykres.

² Przypomnę jedynie, że w roku 2005 Sprawdzian był najłatwiejszy z dotychczas przeprowadzonych i oczywiście bardzo lewoskośny.

³ W skali staninowej przyjmuje się, że wartość średnia wynosi 5, a odchylenie standardowe – 2.

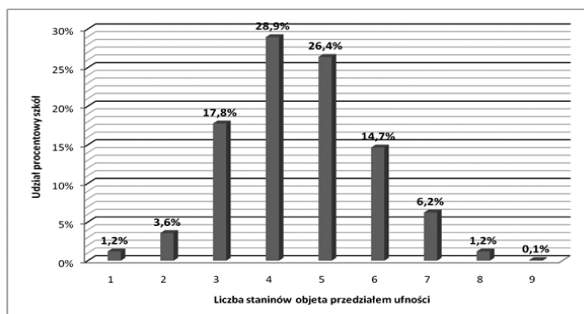
⁴ Dla szkół z tak małą liczbą uczniów przedział ufności jest tak duży, że określenie pozycji szkoły na jakiegokolwiek skali jest praktycznie niemożliwe.



Średnia dla wszystkich szkół wynosiła⁵ ponad 2,5 punktu. Oznacza to, że lokalizacja wyniku szkoły o przeciętnym przedziale ufności jest określona z dokładnością $\pm 2,5$ punktu (dla $\alpha=0,05$). W naturalny sposób pojawia się pytanie, ile jest szkół, dla których cały przedział ufności zawiera się w obszarze stanina, do którego przypisana jest szkoła. Jeśli tak jest w istocie, to standardowa dziewiątka dobrze opisuje rzeczywistość i nie ma konieczności dyskusowania o jej zmianie. Ale jeśli tak nie jest?

Przyjmijmy następującą umowę: jeśli cały przedział ufności dla średniej zawiera się w obszarze jednego stanina, wtedy szkole przypisujemy liczbę 1, jeśli początek przedziału znajduje się w pewnym staninie, a koniec w następnym – 2 itd. Przypisana liczba będzie informowała, jaki obszar na skali staninowej zajmuje przedział, w którym z 95% pewnością będziemy mogli lokalizować szkołę.

Kolejny wykres ilustruje rozkład badanych szkół pod względem opisanej cechy.



Na 1100 szkół jedynie 13 (1,2%) „zmieściło” się w obszarze jednego stanina. Co więcej, były to wyłącznie szkoły z wynikami najniższymi i najwyższymi (1. i 9. stanin) i to dlatego, że w zasadzie pierwszy stanin nie ma ograniczenia z dołu, a dziewiąty – z góry.

Spróbujmy złagodzić nasze warunki. Uznajmy, że pozycja szkoły określona z dokładnością do jednego stanina satysfakcjonuje nas w pełni. Inaczej mówiąc,

⁵ okładnie była równa 2,71 punktu.

przyjmujemy, że jeśli określamy położenie szkoły na czwartym stopniu skali, to stwierdzamy, że z 95% pewnością szkoła znajduje się w 3., 4. lub 5. staninie. Niestety nawet takie założenie nie zapewnia zadowolających rezultatów. Wyżej wymienione warunki spełnia co piąta szkoła. To zbyt mało, aby można twierdzić, iż skala staninowa dla szkół jest przydatna do porównywania osiągnięć placówek. Co więcej, położenie prawie 80% szkół na skali jest bardzo mało pewne, obarczone błędem ± 2 staniny lub więcej. Łatwo zauważyć, że przy tak dużych niedokładnościach szkoły różniące się o dwa lub trzy stopnie na tej skali wcale nie muszą różnić się w stopniu istotnym statystycznie!

Czy można temu zaradzić? Oczywiście, że tak! Wystarczy do komunikowania wyników zastosować dłuższą skalę standardową, np. skalę akademicką lub probabilistyczną skalę theta (Szaleniec, 2005; Sapanowski, 2008, 2009), a także komunikować średnią szkoły wraz z przedziałem ufności. Pierwsze kroki w tym kierunku poczynił już zespół pracujący nad wskaźnikami Edukacyjnej Wartości Dodanej (EWD), który wykorzystał skalę IQ i prezentuje wyniki placówek oświatowych w granicach ufności na poziomie 95%.

W wielu opracowaniach wyników i raportach można znaleźć stwierdzenie: *Średniej wyników egzaminu ... nie należy porównywać ze średnimi wyników egzaminu ... z lat poprzednich*. Przy czym w wolne miejsce należy wstawić nazwę odpowiedniego egzaminu. I nie koniec na tym. Okazuje się, że skala staninowa dla szkół również nie pozwala na takie porównania.

Bibliografia:

1. Szaleniec H., *Wykorzystanie probabilistycznych modeli zadania testowego do zrównania wyników sprawdzianu 2003-2005 i budowania banku zadań*, [w:] *Holistyczne i analityczne metody diagnostyki edukacyjnej*, red. B. Niemierko, G. Szyling, Gdańsk 2005.
2. Sapanowski S., *Oszacowanie umiejętności „tetha” oraz wyskalowanie osi w metodzie IRT dla potrzeb obliczania parametrów zadań*, [w:] *XIV KDE – Uczenie się i egzamin w oczach nauczycieli*, Opole 2008.
3. Sapanowski S., *Porównywanie wyników egzaminów – propozycja metody*, [w:] *XV KDE – Badania międzynarodowe i wzory zagraniczne w diagnostyce edukacyjnej*, Kielce 2009.