KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

Projekt współfinansowany przez Unię Europejską
w ramach Europejskiego Funduszu Społecznego

EUROPEJSKI
FUNDUSZ SPOŁECZNY

Podręcznik użytkownika bazy EWD-CKE

baza w wersji 1.0

dokumentacja w wersji 1.0

Maciej Jakubowski
grudzień 2008

Centralna Komisja Egzaminacyjna
00-842 Warszawa
ul. Łucka 11
tel. (022) 656 38 00
www.cke.edu.pl

publikacja jest dystrybuowana
bezpłatnie

www.efs.gov.pl

Zespół realizacji projektów
tel. (022) 658 64 20
fax (022) 658 61 32
efs@cke.edu.pl

Spis treści

1. Opis bazy danych EWD-CKE

A. Pochodzenie danych zawartych w bazie

Udostępniona baza danych, która zawiera wyniki egzaminów zewnętrznych umożliwiające oszacowanie edukacyjnej wartości dodanej, powstała jako efekt prac prowadzonych przez grupę badawczą kierowaną przez doktora Romana Dolatę przy Centralnej Komisji Egzaminacyjnej. Grupa ta, w składzie (kolejność alfabetyczna): Roman Dolata, Maciej Jakubowski, Ewa Kędracka, Przemysław Majkut, Artur Pokropek, Anna Rappe, Ewa Stożek, Krystyna Szmigiel, odpowiada za przygotowanie danych, modeli statystycznych oraz szkoleń z analizy efektywności pracy szkół metodą edukacyjnej wartości dodanej. Wyniki prac tej grupy opublikowane zostały w książce pod red. R. Dolaty oraz specjalnych numerach biuletynów CKE. Powstała też praca dokumentująca prace nad modelem EWD w Polsce w języku angielskim. Publikacje te można znaleźć na stronie http://www.ewd.edu.pl/publikacje_ewd.html. Istnieje też bardzo bogata literatura dotycząca metod modelowania EWD w oparciu o doświadczenia innych krajów. Głównym źródłem odniesienia są tu raporty OECD (2008) oraz RAND (2005). Podstawowe metody modelowania EWD, a także przykłady badań, opisano też poniżej, w ostatniej części podręcznika.

Wyjątkową cechą udostępnianych zbiorów danych jest to, że dla każdego ucznia zawierają połączone wyniki sprawdzianu oraz egzaminu gimnazjalnego. Łączenie zostało dokonane na poziomie Okręgowych Komisji Egzaminacyjnych, które co roku przekazywały grupie badawczej dane uczniów, dla których scalenie wyników obu egzaminów zewnętrznych było możliwe. Ze względu na to, że brakuje w polskim systemie egzaminacyjnym indywidualnego identyfikatora dla każdego ucznia, łączenie wykonano głównie w oparciu o imię i nazwisko oraz płeć uczniów, które nierzadko podawane były błędnie. Z tego względu dla części uczniów danych nie połączono. Trzeba też wziąć pod uwagę, że łączenie na poziomie OKE powoduje, że uczniowie, którzy między sprawdzianem szóstoklasistów a egzaminem gimnazjalnym zmienili miejsce zamieszkania na leżące w innej OKE, nie mogli zostać uwzględnieni. Przy łączeniu nie brano też pod uwagę uczniów piszących inne niż standardowe arkusze egzaminacyjne, bowiem wyniki tych uczniów nie są w pełni porównywalne. Powoduje to, że łącznie przez 3 lata dla mniej niż 10% uczniów danych nie zdołano połączyć. Nie jest to duża liczba, biorąc pod uwagę, że większość uczniów, dla których wyników nie udało się połączyć, pisała arkusze niestandardowe i pominięcie tych uczniów jest uzasadnione. Braki danych można też uznać w większości za losowe, bowiem wynikają z błędnie zapisanych danych uczniów w systemie. Jedynie kwestia braku uczniów, którzy zmienili rejon OKE, powoduje nielosowy dobór związany m.in. z migracją. Jest to jednak znikomy procent uczniów i dla analiz na poziomie szkół nie powinien mieć większego znaczenia. Z ogólnej liczby gimnazjów w Polsce (włączając np. szkoły przyszpitalne) baza przygotowana do liczenia EWD zawiera ponad 98% szkół.

Baza, z której wylosowano udostępniane dane, została też ograniczona do szkół, dla których możliwe jest prowadzenie porównywalnych, merytorycznie i statystycznie uprawionych analiz. Po pierwsze, wykluczono gimnazja niepubliczne. W bazie dla całej populacji szkół tych jest niewiele (ok. 7%), a uczęszcza do nich jedynie ok. 1,5% uczniów (są to szkoły stosunkowo niewielkie). Po ograniczeniach co do wielkości szkół i klas (patrz niżej) wylosowana baza zawierałaby 1-2 takie szkoły, co z pewnością uniemożliwia jakiekolwiek porównania między placówkami publicznymi a niepublicznymi. Ponadto zmienne dołączane do bazy, np. wydatki na gimnazja, dotyczą przede wszystkim zasobów dostępnych szkołom publicznym, które utrzymują się głównie z publicznych funduszy (dane statystyczne nie zawierają wydatków prywatnych). Stąd uwzględnienie szkół niepublicznych było bezcelowe. Po drugie, ze względu na to, że losowano nie tylko szkoły, lecz także klasy, dla których potencjalne analizy powinny być również wiarygodne, wykluczono klasy małe, z mniej niż 15 uczniów (ok. 2% uczniów), oraz klasy bardzo duże, z więcej niż 35 uczniów (mniej niż 1% uczniów). Klasy bardzo małe muszą być nietypowe w szkołach dużych, jakie zawiera wylosowana próba (patrz następny punkt), co było dodatkowym argumentem, żeby nie brać ich pod uwagę. Klasy z więcej niż 35 uczniów są w rzeczywistości bardzo rzadkie i istnieje duże prawdopodobieństwo, że tak duża liczba uczniów wynika z błędnego kodu klasy. Po trzecie, założono, że udostępnione zostaną dane jedynie dla szkół, dla których posiadamy odpowiednią liczbę obserwacji z każdego roku. Takie kryterium jest niezbędne, o ile dane mają być wykorzystywane do oceny zmian między latami w poszczególnych szkołach. Za minimalną liczbę przyjęto 10 obserwacji w każdym roku[1]. Jeśli w którymś roku gimnazjum nie posiadało tylu zdających na obu częściach egzaminu gimnazjalnego, to nie brano go pod uwagę przy losowaniu opisywanej tu próby. Po czwarte, przyjęto, że ze wszystkich 3 lat każde gimnazjum powinno posiadać co najmniej 50 obserwacji. Gwarantuje to, że analizy dla szkół są miarodajne, bowiem przy mniejszej liczbie obserwacji statystyki na poziomie szkół są mało odporne na obserwacje nietypowe, dość częste w przypadku wyników egzaminów zewnętrznych[2]. Przyjęte kryteria wykluczają z analizy gimnazja niewielkie, jednak są to najczęściej szkoły nietypowe (pomijając szkoły niepubliczne, gimnazja w odróżnieniu od szkół podstawowych są najczęściej szkołami dużymi). Po piąte, wykluczono z bazy niewielką liczbę obserwacji dla uczniów z brakiem wyniku lub wynikiem równym zero. Dla tych uczniów nie posiadamy informacji odnośnie ich poziomu umiejętności (wynik zero często oznacza unieważnienie egzaminu lub oddanie pustego testu z innych przyczyn). Liczba takich obserwacji jest niewielka i w żaden sposób nie może zniekształcać wyników analiz (w zależności od roku jest to kilkudziesięciu uczniów z ponad 400 tysięcy). Przyjęcie wszystkich powyższych kryteriów zmniejszyło znacząco liczbę szkół (o 20%), jednak liczba uczniów zmniejszyła się o mniej niż 8%. Najwięcej szkół odrzucono ze względu na zbyt małą liczbę obserwacji lub ich brak w niektórych latach (ok. 16% szkół i 4% uczniów), a szkoły niepubliczne, które sprostały temu kryterium, stanowiły kolejne 2% gimnazjów i 1% uczniów odrzuconych przed losowaniem bazy. Widać więc, że baza przygotowana do losowania próby została ograniczona głównie przez te dwa kryteria.

Tabela poniżej podsumowuje różnice między: (a) całą populacją gimnazjów i uczniów zdających egzaminy gimnazjalne w każdym z analizowanych lat[3]; (b) liczbą gimnazjów i uczniów dostępnych w bazach scalonych przez OKE; (c) liczbą gimnazjów i uczniów po wykluczeniu szkół: niepublicznych, z niewielką liczbą lub brakiem obserwacji w niektórych latach (w sposób opisany powyżej); oraz

---

[1] Liczbę obserwacji (liczbę uczniów zdających w danym roku egzamin gimnazjalny) obliczono jako średnią z liczby prawidłowych wyników w części humanistycznej oraz w części matematyczno-przyrodniczej. W większości szkół liczba zdających obie części egzaminu była identyczna, jednak w kilku przypadkach różniła się. Jeśli w jednej z części liczba zdających wynosiła w danym roku 10, ale w innej 9, to taka szkoła nie była brana pod uwagę przy losowaniu prezentowanej bazy danych, ponieważ średnia liczba uczniów wyniosła w tej szkole 9,5.

[2] Na przykład uczeń, który uzyskał wysoki wynik na sprawdzianie, jednak z różnych względów napisał egzamin gimnazjalny na bardzo niskim poziomie, zaburza ocenę efektów pracy gimnazjum, o ile liczba innych uczniów jest niewielka. Łatwo to zrozumieć, jeśli wyobrazimy sobie ucznia, który uzyskał maksimum punktów na sprawdzianie, ale zupełnie nie przyłożył się do egzaminu gimnazjalnego, uzyskując liczbę punktów bliską zeru. Uczeń tego rodzaju może spowodować znaczne zaniżenie oceny EWD szkoły, jeśli liczba innych zdających była nieduża. Przy znacznej liczbie zdających jego wynik będzie miał znacznie mniejsze znaczenie dla oceny szkoły.

[3] Wg oficjalnych raportów CKE dla egzaminu gimnazjalnego z lat 2005, 2006, 2007 oraz dla sprawdzianu szóstoklasistów z lat 2002, 2003, 2004.

(d) ostateczną liczbą gimnazjów i uczniów w wylosowanej celem upublicznienia próbie. W dalszych punktach przeanalizowano reprezentatywność tak utworzonej bazy względem podstawowych zmiennych: wyników egzaminów, płci oraz lokalizacji szkoły.

Tabela 1. Liczba szkół i uczniów w populacji, bazach do dyspozycji CKE, oraz wylosowanej próbie.

Egzamin gimnazjalny				Sprawdzian szóstoklasistów	
Rok	Liczba szkół	Liczba zdających w części humanistycznej	Liczba zdających w części matematyczno-przyrodniczej	Rok	Liczba zdających
Cała populacja zdających (wg raportów CKE)					
2005	6319	539408	538211	2002	549319
2006	6447	522474	522321	2003	530577
2007	6513	517489	517124	2004	527245
Baza dla całej populacji dostępna po połączeniu danych w OKE					
2005	6256	483673	483624	2002	483692
2006	6361	481150	481150	2003	481150
2007	6347	460760	460626	2004	460825
Baza dla szkół publicznych z odpowiednią liczbą uczniów w ciągu 3 lat					
2005	5036	443886	443845	2002	443901
2006	5036	440323	440323	2003	440323
2007	5036	423463	423354	2004	423518
Wylosowana próba (baza publicznie dostępna)					
2005	200	9197	9197	2002	9197
2006	200	9122	9122	2003	9122
2007	200	9051	9049	2004	9051

Do bazy z wynikami egzaminacyjnymi dołączono zmienne z Systemu Informacji Oświatowej (SIO) Ministerstwa Edukacji Narodowej oraz Banku Danych Regionalnych Głównego Urzędu Statystycznego (BDR GUS). Dane SIO łączone były za pomocą kodu gimnazjum przypisanego przez OKE. Kod ten powinien znajdować się także w bazach SIO, niestety w wielu przypadkach nie był tam jednak zapisany lub zapisany był błędnie. Dla takich szkół dane łączono po nazwie szkoły i miejscowości, w której znajduje się placówka. Dane GUS połączono, wykorzystując kod nadawany przez GUS gminom, który zawarty jest także w kodzie szkoły przypisywanym przez OKE. W tym przypadku także występowały różnice wynikające z innego kodowania OKE w dużych miastach oraz zmian kodów GUS nie zawsze odzwierciedlonych w kodach OKE. Dla wszystkich szkół i gmin różnice te udało się jednak wyjaśnić i dane zostały połączone. Trzeba jeszcze raz podkreślić, że dane SIO dotyczą poziomu szkoły, a dane BDR GUS poziomu gminy (lub miasta-powiatu). Wysoki poziom agregacji danych BDR GUS powoduje, że nie zawsze są to zmienne dobrze określające sytuację danej szkoły. Jeśli np. gmina przeznacza bardzo różne kwoty podlegającym jej placówkom, to średnie wydatki na poziomie gminy przypisane szkole są kiepskim wskaźnikiem jej rzeczywistych zasobów. Można jednak uznać, że w większości przypadków gminy rozdysponowują środki na podobnych zasadach dla wszystkich szkół. Można też traktować te zmienne jako odzwierciedlające poziom finansowania lub inne cechy na poziomie lokalnym, czyli jako wskaźniki otoczenia szkoły, w jakim funkcjonuje. Ostateczna interpretacja zależy od użytkowników danych, a przede wszystkim od tego, jakie zmienne i w jakim celu wykorzystywano w konkretnym badaniu.

B. Charakterystyka egzaminów zewnętrznych

W bazie zawarto wyniki egzaminów zewnętrznych dla wylosowanych uczniów i szkół. Baza zawiera zarówno sumaryczny wynik sprawdzianu ucznia, jak i osobne wyniki egzaminu gimnazjalnego w obu częściach. Uczniom przypisano losowe identyfikatory, podobnie jak szkołom. Niemożliwe jest zidentyfikowanie ucznia i szkoły, gdyż ani zanimizowanie identyfikatorów ani procedura losowania nie umożliwiają określenia, z wynikami jakiego gimnazjum mamy do czynienia. Mimo pełnej anonimowości wyniki zawarte w bazie są całkowicie reprezentatywne dla populacji uczniów, dla których udało się połączenie wyników sprawdzianu i egzaminu gimnazjalnego, a przy tym chodzących do gimnazjów publicznych z odpowiednio licznymi klasami w analizowanych latach. Można uznać, że baza jest reprezentatywna dla populacji uczniów „typowych" gimnazjów w Polsce. Przedstawiona w kolejnym punkcie procedura losowania gwarantuje taką reprezentatywność, a potwierdzają to testy na różnice w statystykach między populacją, z której losowano, a udostępnioną próbą – je także przedstawiono poniżej.

Dokładny opis systemu egzaminów zewnętrznych, poszczególnych egzaminów oraz raporty z wynikami i podsumowaniem egzaminów z każdego roku, można znaleźć na stronie Centralnej Komisji Egzaminacyjnej: www.cke.edu.pl. Dokładne wyniki w regionach prezentowane są również na stronach OKE (patrz linki na stronie CKE). Poniżej opisujemy też pokrótce cechy egzaminów zewnętrznych, które mają znaczenie dla modeli edukacyjnej wartości dodanej, a także modyfikacje wprowadzone celem ułatwienia i uprawomocnienia analizy.

Egzaminy zewnętrzne zostały po raz pierwszy przeprowadzone w 2002 roku, kiedy niemal wszyscy uczniowie kończący szkołę podstawową napisali **sprawdzian**. Sprawdzian szóstoklasistów jest egzaminem obowiązkowym, zawierającym te same zadania w całej Polsce i ocenianym zewnętrznie wobec szkoły. Do dziś jest to stosunkowo prosty, ponadprzedmiotowy egzamin, którego wyniki raportowane są na skali od 0 do 40 stanowiącej prostą sumę punktów uzyskanych przez danego ucznia. Wyniki egzaminu tylko w 2002 roku zostały umieszczone na świadectwie szkolnym i od początku nie powinny być wykorzystywane do selekcji uczniów przy naborze do gimnazjów. Tak więc sprawdzian ma z natury jedynie charakter informacyjny, określający, na ile uczniowie kończący szkołę podstawową posiedli wiedzę umożliwiającą naukę na kolejnym poziomie kształcenia.

W 2002 roku po raz pierwszy przeprowadzono także **egzamin gimnazjalny**, który jest obowiązkowy dla uczniów kończących gimnazja. Egzamin ten składa się z dwóch części: humanistycznej i matematyczno-przyrodniczej, z których wyniki indywidualne raportowane są osobno na nieprzetworzonej skali od 0 do 50. Wyniki egzaminu gimnazjalnego mogą być wykorzystywane przy naborze uczniów do szkół średnich[4].

W prezentowanej bazie zawarto wyniki egzaminu gimnazjalnego z lat 2005, 2006 i 2007, a także wyniki sprawdzianu z lat 2002, 2003 i 2004, które mogą służyć jako zmienne kontrolne mierzące zasób wiedzy ucznia na progu gimnazjum. Warto podkreślić kilka cech egzaminów zewnętrznych istotnych dla dalszych analiz statystycznych. Po pierwsze, wyniki indywidualne raportowane są na nieprzetworzonej skali, sumującej punkty za poprawne odpowiedzi, przez co wyniki te nie są bezpośrednio porównywalne między latami i przedmiotami. Ich rozkład w kilku przypadkach znacząco odbiega od normalnego. Po drugie, przeprowadzone dotąd badania pokazują, że wyniki egzaminów zewnętrznych okazują się zarówno dobrymi predyktorami osiągnięć na kolejnych etapach kształcenia

[4] Szczegółowe informacje o systemie egzaminów zewnętrznych, a także raporty o wynikach z poszczególnych lat można znaleźć na stronie CKE: www.cke.edu.pl, lub na: http://www.wynikiegzaminow.pl/.

(patrz analizy EWD w kolejnych rozdziałach), jak i powiązane są silnie z cechami uczniów warunkującymi sukcesy edukacyjne oraz charakterystykami środowiska szkolnego i społecznego[5]. Sprawdzono także, że wyniki egzaminów silnie korelują z wynikami międzynarodowego badania PISA. Inaczej mówiąc, choć niestandaryzowane wyniki egzaminów są trudne do bezpośredniej interpretacji, to stanowią dobre miary poziomu wiedzy i umiejętności uczniów na poszczególnych etapach kształcenia. Tabela poniżej prezentuje współczynniki korelacji między wynikami sprawdzianu oraz egzaminu gimnazjalnego (dla całej populacji z połączonymi wynikami), a także wynikami egzaminu gimnazjalnego oraz wynikami PISA (dla próby uczniów wylosowanych do badania PISA 2006)[6]. Jak widać, wynik sprawdzianu niezwykle silnie koreluje z wynikami egzaminów zewnętrznych w obydwu częściach. Podobnie silne związki dotyczą wyników egzaminu gimnazjalnego z wynikami PISA. Dodatkowo wyniki PISA w matematyce i naukach ścisłych są silniej powiązane z częścią matematyczno-przyrodniczą egzaminu gimnazjalnego, a w czytaniu – z częścią humanistyczną. Wszystko to pokazuje, że egzaminy zewnętrzne, mimo swojej prostoty, są dobrym pomiarem umiejętności uczniów dotyczących podobnych konstruktów psychologicznych. Co więcej, biorąc pod uwagę, że poniższe korelacje odzwierciedlają jedynie związek liniowy i są dodatkowo zaniżone ze względu na losowy błąd pomiaru (występujący przy każdym pomiarze umiejętności uczniów) można stwierdzić, że rzeczywiste związki są jeszcze silniejsze.

Tabela 2. Korelacja między wynikami egzaminów zewnętrznych oraz wynikami PISA 2006.

	2002/2005		2003/2006		2004/2007		PISA 2006		
	spr	*hum*	*spr*	*hum*	*spr*	*hum*	*nauki ścisłe*	*czytanie*	*matematyka*
hum	0,73		0,72		0,77		0,69	0,68	0,68
mat	0,73	0,71	0,73	0,69	0,74	0,71	0,74	0,66	0,75

W bazie zawarto wyniki egzaminów gimnazjalnych i sprawdzianu dla 3 kohort uczniów: pierwszego rocznika, który pisał sprawdzian szóstoklasistów w 2002 roku, a następnie egzamin gimnazjalny w 2005 roku, oraz kolejnych roczników, które pisały sprawdzian w latach 2003 i 2004, a egzamin gimnazjalny odpowiednio w latach 2006 i 2007. W momencie losowania próby tylko dla tych trzech kohort uczniów dostępne były zarówno wyniki sprawdzianu, jak i egzaminu gimnazjalnego. Oprócz wyników egzaminu dołączono także cechy uczniów, które gromadzone są w bazach OKE w sposób spójny dla całego kraju. Są to: płeć ucznia, informacja, czy pisał wersję dla dyslektyków w szkole podstawowej oraz jaką wersję pisał w gimnazjum, oznaczenie uczniów będących laureatami olimpiad. Skonstruowano także zmienne podsumowujące cechy szkoły dotyczące egzaminów zewnętrznych: średni wynik sprawdzianu uczniów gimnazjum, średni wynik egzaminu gimnazjalnego szkoły, miary zróżnicowania wyników itp. Zmienne te opisano w odpowiedniej części podręcznika.

[5] Wyniki badań publikowane są na bieżąco w Biuletynie Badawczym CKE „Egzamin" do ściągnięcia ze strony CKE (http://www.cke.edu.pl/index.php?option=content&task=view&id=365&Itemid=179).

[6] Korelacje dla egzaminów zewnętrznych obliczono dla pełnej bazy połączonych wyników egzaminów, bez żadnych wykluczeń i standaryzacji wyników. Korelacje z wynikami PISA obliczono dla uczniów, dla których udało się do danych PISA dołączyć wyniki egzaminu gimnazjalnego, czyli dla 3 798 osób. Dane te połączył zespół PISA przy Instytucie Filozofii i Socjologii PAN. Korelacja dotyczy pierwszej „plausible value" oraz wyniku egzaminu standaryzowanego do średniej 500 i odchylenia standardowego 100, podobnie jak wyniki PISA.

C. Charakterystyka źródeł innych zmiennych zawartych w bazie

Ustawa z dnia 19 lutego 2004 r. o systemie informacji oświatowej dała podstawę do gromadzenia danych ze wszystkich placówek systemu oświaty w Polsce. System Informacji Oświatowej (SIO) od kilku lat stanowi główne źródło danych dla statystyki publicznej. Zarządzany jest przez Ministerstwo Edukacji Narodowej i zawiera dokładne dane dotyczące organizacji oraz kadry placówek oświatowych w całym kraju. Więcej informacji dotyczących SIO można uzyskać na stronach MEN, np.: http://sio.men.gov.pl.

W tej wersji bazy zmiennych SIO nie dołączono.

Innym ważnym źródłem informacji na temat szkolnictwa i samorządów jest Bank Danych Regionalnych Głównego Urzędu Statystycznego (BDR GUS). Rolę GUS w gromadzeniu danych oświatowych przejął w ostatnich latach SIO, jednak wcześniej placówki oświatowe przesyłały odpowiednie formularze do GUS. Wciąż też GUS odpowiedzialny jest za prowadzenie statystyk oświatowych dla jednostek samorządu terytorialnego. Te właśnie statystyki gromadzone są w BDR GUS i udostępniane w Internecie na stronach GUS. BDR GUS zawiera m.in. liczbę uczniów oraz placówek oświatowych podległych jednostkom samorządowym, a także innym podmiotom działającym na ich terenie. BDR GUS zawiera także dane budżetowe samorządów, w tym dotyczące wydatków na oświatę. W chwili konstruowania niniejszej bazy danych dostępne były dane z lat 1996-2006. Więcej informacji można uzyskać pod adresem www.stat.gov.pl/bdr_n/app/strona.indeks.

Dane finansowe są najczęściej analizowane w jednostkach realnych, po uwzględnieniu inflacji. W bazie udostępniono wydatki gmin zarówno w ujęciu nominalnym, jak i realnym, po uwzględnieniu wskaźnika HICP mierzącego inflację wg standardów europejskich nałożonych przez Eurostat. Więcej informacji o konstrukcji tego wskaźnika, jak i same dane, można znaleźć na stronach Eurostat[7]. Możliwe jest także wykorzystanie innych wskaźników inflacji i z tego względu w bazie pozostawiono wydatki nominalne, umożliwiając ich dowolne przeliczenie przez użytkowników. Trzeba podkreślić, że nieuwzględnienie inflacji w analizach między latami może prowadzić do błędnych wniosków, dlatego też zachęca się użytkowników do korzystania z wydatków wyrażonych w terminach realnych.

[7] Opis metodologii i dane HICP można znaleźć na stronie: http://epp.eurostat.ec.europa.eu/portal/page?_pageid=2714,1,2714_61582043&_dad=portal&_schema=PORTAL

2. Opis losowania próby gimnazjów do bazy

A. Sposób losowania

Schemat doboru próby zastosowany w przypadku omawianych danych można określić jako warstwowy, wielopoziomowy dobór zespołowy. Najpierw ustalone zostały warstwy, wedle których szkoły były losowane. W poszczególnych warstwach szkoły wylosowane zostały z prawdopodobieństwem proporcjonalnym do liczby oddziałów szkolnych klas trzecich uczących się w szkole w latach: 2005-2007. Następnie w wylosowanych szkołach, losując w sposób prosty, wybrano po 2 oddziały z każdego rocznika. Co daje maksymalnie 6 oddziałów z każdej szkoły. Jeżeli w danym roku w szkole występowały jedynie dwa bądź jeden oddział – wybierane były one automatycznie do badania.

Warstwy utworzone zostały na podstawie wyników egzaminu gimnazjalnego (sumy) z 2007 roku. Szkoły zostały podzielone na 100 warstw zawierających 50 lub 51 szkół każda. W pierwszej warstwie zgrupowanych zostało 50 szkół o najniższych wynikach średnich egzaminu gimnazjalnego. W warstwie o numerze 100 znalazło się natomiast 50 szkół gimnazjalnych o najwyższych wynikach średnich egzaminu gimnazjalnego. Jako iż ilości szkół nie dało się podzielić na 100 równych warstw, 39 warstw zawierało 51 szkół, w pozostałych 61 warstwach liczebność wynosiła 50 szkół. Populacja szkół w operacie liczy 5 039, wylosowano 200 w 100 warstwach, zatem w przypadku prostego podziału na 100 warstw po 50 szkół pozostawałoby 39 szkół.

Zanim przyjęto schemat losowania i dokonano losowania, przeprowadzono symulacje, które służyły ocenie schematu losowania oraz warstwowania. Za podstawowe kryterium oceny przyjęto kwadratową funkcję błędu estymatora (MSE – *mean squared error*) dla wyników egzaminu gimnazjalnego z obu części oraz sprawdzianu po szkole podstawowej. Dla każdego schematu losowania wykonano 1 500 symulacji (patrz załącznik). Ostatecznie wybrano sposób losowania dający największa precyzję oceny parametrów w populacji: losowanie proporcjonalnie do liczby klas w szkole w warstwach ustalonych ze względu na wynik gimnazjum na egzaminie.

Wagi, które stosujemy w tych zbiorach, to wagi analityczne określane za pomocą odwrotności prawdopodobieństwa wylosowania jednostki z próby. Prawdopodobieństwo *p* wylosowania jednostki *i* określamy jako:

$$p_i = \frac{n}{N}$$

Gdzie *n* oznacza liczbę szkół, które mają być wylosowane, a *N* całkowitą liczbę szkół. Waga *w* dla wylosowanej jednostki obserwacji *i* określona jest zatem jako:

$$w_i = \frac{1}{p_i} = \frac{N}{n}$$

Odnosząc się do konkretnego przykładu: gdy losujemy 100 szkół z populacji liczącej 5 000 szkół, wtedy prawdopodobieństwo wylosowania danej szkoły wynosi 1/50, a waga, jako odwrócone prawdopodobieństwo, wynosi:

$$w_i = \frac{1}{p_i} = \frac{1}{(1/50)} = 50 \; = \; w_i = \frac{N}{n} = \frac{5000}{100} = 50$$

Jeżeli losowanie ma charakter wielopoziomowy i np. najpierw losowana jest określona liczba szkół, a później z każdej szkoły losowana jest określona liczba uczniów, to określane są niezależne wagi (i prawdopodobieństwa wylosowania danej szkoły) dla obydwu poziomów losowania. Załóżmy, iż w danej populacji istnieje 5 000 szkół, a w każdej ze szkół uczy się dokładnie po 200 uczniów. Schemat losowania zakłada, iż z każdej szkoły losujemy po 20 uczniów. Wtedy prawdopodobieństwo wylosowania szkoły *i* jest oczywiście znowu równe 1/50, ale w każdej z wylosowanych szkół *i* prawdopodobieństwo wylosowania ucznia *j* wynosi:

$$p_{ji} = \frac{20}{200} = \frac{1}{10}$$

Zatem – co wiemy dzięki rachunkowi prawdopodobieństwa – łączne prawdopodobieństwo wylosowania danego ucznia *j* wynosi:

$$p_j = p_i \cdot p_{ji} = \frac{1}{50} \cdot \frac{1}{10} = \frac{1}{500}$$

Natomiast waga przyporządkowana uczniowi:

$$w_{ji} = \frac{1}{p_{ji}} = \frac{1}{(1/500)} = 500$$

Tak zdefiniowane wagi mają dość prostą interpretację, mówią nam, ile jednostek obserwacji z badanej populacji reprezentowanych jest przez jedną *j* wylosowaną jednostkę. Inaczej mówiąc, waga $w_{ji} = 500$ oznacza, iż jeden wylosowany przez nas uczeń odpowiada 500 uczniom w całej populacji. Jeden uczeń reprezentuje 500 innych uczniów.

Aby przybliżyć sposób, w jaki liczone były wagi w tym badaniu, w tabeli 3. przedstawiono hipotetyczny operat losowania składający się, dla przejrzystości, jedynie z dwóch warstw. Schemat losowania jest identyczny jak w rzeczywistym badaniu. Najpierw losowane są szkoły (w przykładzie po dwie z każdej warstwy), potem losowane są oddziały – po dwa z każdej szkoły (jeżeli nie ma dwóch, wybierany jest 1 oddział). Dla przejrzystości zakładamy tutaj, iż w każdej szkole liczebności klas są jednakowe. Rozpatrzmy np. przypadek szkoły nr 3. Należy ona do warstwy A., zawiera w sobie 4 oddziały, w każdym z oddziałów uczy się po 25 uczniów (razem 100). Jako że losujemy po dwie szkoły z każdej warstwy, to prawdopodobieństwo wylosowania szkoły nr 3 jest równe ilorazowi wylosowanych szkół z danej warstwy do całkowitej ilości szkół w warstwie:

$$p_3 = \frac{n_A}{N_A} = \frac{2}{3} = 0,67$$

Waga dla szkoły jest odwróceniem tego prawdopodobieństwa i wynosi:

$$w_3 = \frac{1}{p_3} = \frac{1}{(2/3)} = 1,5$$

Prawdopodobieństwo wylosowania 2 spośród 4 klas w tej szkole jest oczywiście równe

$$p_{3k} = \frac{n_{3k}}{N_{3k}} = \frac{2}{4} = 0,5$$

Waga dla tej klasy jest oczywiście równa 2. Całkowite prawdopodobieństwo wylosowania ucznia należącego do tej szkoły do badania podstawowego jest równe iloczynowi prawdopodobieństwa wylosowania szkoły i prawdopodobieństwa wylosowania klasy:

$$p_{bad.p.} = p_3 \cdot p_{3k} = \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3} \approx 0,33$$

Waga dla badania podstawowego dla ucznia z tej szkoły, jako że jest odwróceniem prawdopodobieństwa, równa jest $w_{bad.p.} = \frac{1}{p_{bad.p}}$ i dla szkoły nr 3 wynosi 3 (po wcześniejszym zaokrągleniu 3,03).[8]

Tabela 3. Hipotetyczny operat losowania. Dwie warstwy A i B, z każdej warstwy losowane są dwie szkoły, z każdej szkoły losowane są dwa oddziały, jeżeli w szkole jest jeden oddział, wybierany jest jeden oddział.

Szkoła	Warstwa	Liczba oddział.	Ilość uczniów w oddział.	Ilość uczniów w szkole	Prawd. wylos. szkoły	Waga dla szkoły	Prawd. wylos. klasy	Prawd. wylos. ucznia	Waga dla ucznia
1	A	1	20	20	0,67	1,50	1,00	0,67	1,49
2	A	2	20	40	0,67	1,50	1,00	0,67	1,49
3	A	4	25	100	0,67	1,50	0,50	0,33	3,03
4	B	2	20	40	0,29	3,50	1,00	0,29	3,45
5	B	5	25	125	0,29	3,50	0,40	0,11	9,09
6	B	1	30	30	0,29	3,50	1,00	0,29	3,45
7	B	4	25	100	0,29	3,50	0,50	0,14	7,14
8	B	2	20	40	0,29	3,50	1,00	0,29	3,45
9	B	5	30	150	0,29	3,50	0,40	0,11	9,09
10	B	3	20	60	0,29	3,50	0,67	0,19	5,26

8 W tabeli wynosi 3,03, ponieważ obliczona została z wyniku zaokrąglonego: p = 0,33 zatem w = 1/p = 1/(0,33) = 3,03

Ważenie dla wylosowanej próby jest trochę bardziej skomplikowane niż przedstawiono to na przykładach. Co do zasady jednak, liczenie wag nie różni się tu znacząco. Waga dla wylosowanej jednostki obserwacji jest odwróconym prawdopodobieństwem jej wylosowania. Podstawowe różnice zawierają się w 3 punktach:

- znacznie większa liczba warstw
- losowanie odbywa się z prawdopodobieństwem proporcjonalnym do liczby oddziałów
- w szkole losowane są po dwa oddziały z każdego roku

W rozdziale „Opis losowania próby gimnazjów do bazy" opisano zmienne zawierające wagi i inne informacje dotyczące schematu losowania. W rozdziale „Jak analizować dane w bazie" przedstawiono, w jaki sposób poprawnie analizować wylosowaną próbę, uwzględniając zarówno wagi, jak i korygując oszacowania ze względu na schemat losowania.

B. Reprezentatywność próby (udostępnianej bazy danych)

Odpowiedni schemat losowania gwarantuje, że obliczenia statystyczne parametrów dla całej populacji będą możliwie precyzyjne. Losowanie udostępnianej bazy opisaną powyżej metodą daje najmniejszy błąd standardowy przy ocenie wyników egzaminów w populacji. Przyjęty schemat daje też próbę, którą można nazwać „autoważoną", czyli taką, dla której nie jest niezbędna korekta poprzez wagi. Prawidłowa analiza statystyczna powinna jednak wagi uwzględniać, co gwarantuje, że średni wynik w próbie nie będzie odbiegał od średniej w populacji.

Tabela poniżej przedstawia średnie wyniki w obu częściach egzaminu gimnazjalnego oraz sprawdzianu szóstoklasistów dla trzech różnych populacji oraz wylosowanej próby w wersji bez wag i z ważeniem obserwacji. Wylosowana próba stanowi udostępnioną publicznie bazę EWD-CKE, której dotyczy niniejszy podręcznik. Podział w tabeli 4. odpowiada przedstawionemu w tabeli 1., z tą różnicą, że najszersza populacja obejmuje uczniów, którzy w interesujących nas latach rozwiązywali arkusze standardowe (w tabeli 1. przedstawiono liczbę wszystkich uczniów zdających egzaminy). Dla tej populacji wyniki cytujemy za oficjalnymi raportami CKE. Druga populacja dotyczy szkół i uczniów, dla których udało się połączenie wyników egzaminu gimnazjalnego oraz sprawdzianu. Trzecia populacja dotyczy uczniów szkół publicznych spełniających kryteria odpowiedniej liczby obserwacji w każdym roku. Ostatnie dwie grupy w tabeli dotyczą uczniów i szkół wylosowanych wg schematu opisanego powyżej, bez wag, oraz próbę przeważoną. Oceniając reprezentatywność wylosowanej próby można porównać średnie wyniki z próby z rzeczywistymi wynikami w każdej z trzech populacji. Niewątpliwie jednak wyniki będą się różnić między próbą a populacją wszystkich zdających oraz populacją uczniów o połączonych wynikach egzaminu i sprawdzianu, bowiem nie te populacje stanowiły operat losowania. Wyniki nie powinny się natomiast różnić od populacji szkół publicznych posiadających odpowiednio dużą liczbą uczniów, która stanowiła podstawę losowania (operat). Porównania, osobno dla każdego roku, przedstawiono poniżej.

Tabela 4. Wyniki egzaminów w różnych populacjach oraz w wylosowanej próbie.

Rok	Egzamin gimnazjalny – Średni wynik w części humanistycznej	Egzamin gimnazjalny – Średni wynik w części matematyczno-przyrodniczej	Sprawdzian szóstoklasistów – Rok	Sprawdzian szóstoklasistów – Średni wynik
Cała populacja zdających (wg raportów CKE)				
2005	33,18	24,26	2002	29,49
2006	31,39	23,90	2003	28,61
2007	31,48	25,31	2004	25,55
Baza dla całej populacji dostępna po połączeniu danych w OKE				
2005	33,80	24,92	2002	30,11
2006	31,91	24,47	2003	29,18
2007	32,19	25,84	2004	26,24
Baza dla szkół publicznych z odpowiednią liczbą uczniów w ciągu 3 lat				
2005	33,87	24,97	2002	30,16
2006	31,96	24,49	2003	29,22
2007	32,23	25,81	2004	26,27
Wylosowana próba – wyniki nieważone				
2005	33,72	24,92	2002	30,00
2006	32,06	24,61	2003	29,20
2007	32,03	25,45	2004	26,04
Wylosowana próba – wyniki ważone				
2005	33,89	25,08	2002	30,09
2006	32,16	24,77	2003	29,29
2007	32,29	25,67	2004	26,19

Tabela ukazuje dość niewielkie różnice w średnich wynikach egzaminów między trzema omawianymi populacjami. Kolejne populacje, coraz bardziej zawężane, charakteryzują nieco wyższe wyniki na wszystkich egzaminach. Jest to zrozumiałe, o ile weźmie się pod uwagę, że kolejne populacje są uboższe o szkoły najmniejsze, często nietypowe (np. przyszpitalne czy małe szkółki wiejskie). Trzeba jednak podkreślić, że różnice są niewielkie. Największe zmiany wynikają z braku możliwości połączenia wyników uczniów ze sprawdzianu i egzaminu gimnazjalnego, jednak oscylują one w granicach 0,5-0,7 punktu egzaminacyjnego. Minimalne natomiast różnice, mniejsze niż 0,1 punktu egzaminacyjnego, związane są z wykluczeniem szkół niepublicznych, niewielkich lub bardzo dużych klas, a także szkół z brakiem lub małą liczbą obserwacji w pojedynczych latach i w przeciągu 3 lat. Można więc uznać, że operat losowania był bliski populacji gimnazjów i ich uczniów w całym kraju.

Porównanie oszacowań średnich dla próby, zarówno w wersji bez ważenia, jak i po zastosowaniu wag, pokazuje, że wylosowana próba jest w pełni reprezentatywna pod względem przeciętnych wyników egzaminacyjnych uczniów. Różnice między operatem („Baza dla szkół publicznych z odpowiednią liczbą uczniów w ciągu 3 lat”) a wylosowanymi próbami są niewielkie, przeciętnie ok. 0,15 punktu dla próby bez ważenia i 0,11 dla próby z zastosowaniem wag. Można więc uznać, że stosowanie wag, przynajmniej w przypadku analizy wyników egzaminacyjnych, nie jest konieczne, choć ich uwzględnienie zwiększa precyzję. Trzeba jednak pamiętać, że w przypadku bardziej szczegółowych analiz zastosowanie wag może mieć znacznie większe znaczenie, szczególnie jeśli liczymy w grupach zróżnicowanych pod względem wyników egzaminów. O ile oprogramowanie statystyczne, z którego korzystamy, pozwala na zastosowanie wag, to powinniśmy je wziąć pod uwagę, podobnie jak inne zmienne określające schemat losowania (warstwy, identyfikatory jednostek losowania – szkół i klas, korekty na skończoną populację). W rozdziale „Jak analizować dane w bazie" pokazujemy, jak informacje te wykorzystać w programie Stata. Wagi można zastosować w niemal każdym współczesnym pakiecie statystycznym, a najnowsze wersje wielu programów potrafią także wykorzystać pozostałe udostępnione w bazie informacje dotyczące losowania, przynajmniej dla podstawowych procedur statystycznych (umożliwiają to np. SAS czy SPSS).

3. Charakterystyka zmiennych w bazie

A. Wyniki egzaminów

Surowe, nieprzetworzone wyniki egzaminów zewnętrznych zawarto w zmiennych *spr*, *hum* i *mat*, opisanych w tabeli poniżej, gdzie podano także podstawowe ich statystyki (oszacowane bez oraz po uwzględnieniu schematu losowania). Jak już wielokrotnie podkreślano, wyniki egzaminów nie są standaryzowane, a przez to nie są bezpośrednio porównywalne ani między latami, ani między przedmiotami i egzaminami. Inaczej mówiąc, nie są one publikowane na jednej, niezmiennej skali. Z tego względu przy analizie oryginalnych wyników niezbędne są odpowiednie korekty, szczególnie jeśli wykorzystujemy w jednym modelu statystycznym wyniki z kilku lat. Przykład, w jaki sposób takiej korekty można dokonać w modelu regresji EWD, podano w rozdziale „Jak analizować dane w bazie".

Aby ułatwić analizę z wykorzystaniem wyników egzaminów z kilku lat, sprowadzono je do rozkładu wyników dla ostatniego rocznika uczniów[9]. W ten sposób wyniki egzaminów z każdego roku mają tę samą średnią i odchylenie standardowe równe średniej i odchyleniu standardowemu wyników egzaminu gimnazjalnego z 2007 roku oraz sprawdzianu z 2004 roku. Te same egzaminy mają w kolejnych latach podobny rozkład, choć między egzaminami różnice w rozkładach pozostały. Trzeba jednak dodać, że przy tak „krótkiej" skali, obejmującej 40 wartości dla sprawdzianu i 50 dla egzaminu gimnazjalnego, trudno jest uzyskać rozkłady idealnie „gładkie". Nierozwiązany pozostaje też problem „efektu sufitowego". Niewątpliwie jednak przełożenie wyników na jedną skalę ułatwia analizę i korekty różnic między latami nie są już konieczne. Tabele poniżej zawierają nazwy zmiennych, w jakich zawarto wyniki egzaminów oraz statystyki dla oryginalnych oraz przekształconych wyników. Dla każdego rocznika podano statystyki dla całej populacji (operatu losowania) oraz dla wylosowanej bazy danych. W załącznikach można znaleźć wykresy rozkładów każdego egzaminu z poszczególnych lat, zarówno w wersji oryginalnej, jak i po przekształceniu.

[9] Przełożenia wyników na skalę dla ostatniego rocznika dokonała pani Anna Rappe z Okręgowej Komisji Egzaminacyjnej w Krakowie. Wyniki egzaminacyjne z lat 2005, 2006 zostały zrównane z wynikami bazowymi, za jakie przyjęto rok 2007. Dokonano zrównania „wyniku surowego na wynik surowy", bez skalowania. Wykorzystano metodę zrównywania ekwicentylowego, która została dokładnie opisana w Biuletynie Badawczym nr 10, wydanym przez CKE, w artykule prof. Samuela A. Livingstona z Educational Testing Service, USA: „Zrównywanie wyników testów", str. 8-59. W metodzie tej stosuje się procedurę dostosowania, w której wyniki bazowe i zrównywane zachowują takie same rangi centylowe. Przy obliczaniu zastosowano program RAGE-RGEQUATE, który jest ogólnie dostępny na stronie Uniwersytetu Iowa: http://www.education.uiowa.edu/casma/EquatingLinkingPrograms.htm.

Tabela 5. Nazwy zmiennych zawierających oryginalne oraz przetworzone wyniki egzaminów zewnętrznych.

Wynik egzaminu	Nazwa zmiennej
Sprawdzian	*Spr*
Sprawdzian na skali 2004	*Spr_2004*
Część humanistyczna	*Hum*
Część humanistyczna na skali 2007	*Hum_2007*
Część matematyczno-przyrodnicza	*Mat*
Część matematyczno-przyrodnicza na skali 2007	*Mat_2007*

Tabela 6. Statystyki dla wyników egzaminów zewnętrznych.

Rocznik	Populacja/próba	Statystyka	Nazwa zmiennej z wynikiem egzaminu zewnętrznego					
			Spr	*Spr_2004*	*Hum*	*Hum_2007*	*Mat*	*Mat_2007*
2002/ 2005	populacja	średnia	30,16	26,29	33,87	32,26	24,97	25,89
		odch. std.	6,32	7,39	8,24	9,27	10,04	10,09
	próba	średnia	30,03	26,14	33,93	32,34	24,79	25,71
		odch. std.	6,30	7,33	8,24	9,26	9,96	9,99
2003/ 2006	populacja	średnia	29,22	26,27	31,96	32,25	24,49	25,85
		odch. std.	6,29	7,38	7,99	9,29	10,10	10,09
	próba	średnia	29,21	26,24	32,03	32,34	24,43	25,79
		odch. std.	6,33	7,41	7,95	9,24	9,98	9,98
2004/ 2007	populacja	średnia	26,27		32,23		25,81	
		odch. std.	7,39		9,29		10,07	
	próba	średnia	26,12		32,16		25,44	
		odch. std.	7,50		9,29		10,10	

Jak widać, różnice między populacją a wylosowaną próbą (bazą) są nieduże i pomijalne z praktycznego punktu widzenia, zarówno pod względem średniej, jak i zróżnicowania wyników uczniów (odchylenie standardowe). Co więcej, przełożenie wyników dla roczników 2002/2005 oraz 2003/2006 na skalę wyników rocznika 2004/2007 powoduje, że w każdym roku średnia i odchylenie standardowe wyników danego egzaminu są podobne. Widać więc, że można porównywać wyniki egzaminów między rocznikami bez dodatkowych korekt, choć tylko w ramach danego egzaminu (lub części egzaminu gimnazjalnego).

Na podstawie wyników egzaminów w całej populacji, jeszcze przed losowaniem, obliczono statystyki na poziomie szkół, które mogą być wykorzystane w analizie. Dla oryginalnych wyników egzaminów obliczono dla każdej szkoły, osobno dla każdego rocznika, rozstęp ćwiartkowy wyników egzaminu gimnazjalnego w obu częściach, a także rozstęp ćwiartkowy wyników sprawdzianu wszystkich uczniów gimnazjum. Statystyki te mierzą zróżnicowanie końcowych wyników w szkole, a także zróżnicowanie poziomu uczniów na wejściu. Podobnie obliczono medianę wyników uczniów na wejściu oraz medianę wyników końcowych. Dla wyników przełożonych na skalę wyników rocznika 2004/2007 obliczono podobne statystyki, ale tym razem zamiast rozstępu ćwiartkowego wykorzystano odchylenie standardowe, a zamiast mediany średnią. Zastosowanie tych statystyk jest w tym przypadku możliwe ze względu na to, że wyniki mają wspólną skalę i ich porównanie zakładające tę samą jednostkę pomiaru ma sens. Trzeba jednak pamiętać, że w obydwu przypadkach zmienne te należy traktować jako nieodporne na zmiany skali pomiaru i traktować je z ostrożnością. Nazwy zmiennych wraz z opisem przedstawiono w tabeli poniżej. Użytkownicy bazy mogą łatwo obliczyć podobne statystyki dla wylosowanych klas. Mogą one służyć włączaniu do analiz jako zmiennych kontrolnych cech szkoły (klasy) dotyczących przeciętnego poziomu i zróżnicowania wyników uczniów szkoły (klasy) na wejściu do gimnazjum i pod koniec nauki.

Tabela 7. Zmienne zawierające statystyki dla wszystkich uczniów wylosowanych gimnazjów.

Statystyki dla wyników na oryginalnych skalach	
iqr_gim_hum	rozstęp ćwiartkowy wyników uczniów gimnazjum w części humanistycznej
iqr_gim_mat	rozstęp ćwiartkowy wyników uczniów gimnazjum w części matematyczno-przyrodniczej
iqr_gim_spr	rozstęp ćwiartkowy wyników sprawdzianu uczniów gimnazjum
med_gim_hum	mediana wyników uczniów gimnazjum w części humanistycznej
med_gim_mat	mediana wyników uczniów gimnazjum w części matematyczno-przyrodniczej
med_gim_spr	mediana wyników sprawdzianu uczniów gimnazjum
Statystyki dla wyników na skali dla rocznika 2004/2007	
sd_gim_hum	odchylenie standardowe wyników uczniów gimnazjum w części humanistycznej na skali 2007 roku
sd_gim_mat	odchylenie standardowe wyników uczniów gimnazjum w części matematyczno-przyrodniczej na skali 2007 roku
sd_gim_spr	odchylenie standardowe wyników sprawdzianu uczniów gimnazjum na skali 2004 roku
sr_gim_hum	średni wynik w części humanistycznej uczniów gimnazjum na skali 2004 roku
sr_gim_mat	średni wynik w części matematyczno-przyrodniczej uczniów gimnazjum na skali 2004 roku
sr_gim_spr	średni wynik sprawdzianu uczniów gimnazjum na skali 2004 roku

B. Identyfikatory szkół, klas i uczniów

Szkołom oraz klasom zostały losowo przypisane liczbowe identyfikatory. Zmienna *szkola* jednoznacznie identyfikuje szkołę, przyjmując losowe wartości od 1 do 200. Zmienna *klasa* jednoznacznie identyfikuje zarówno szkołę, jak i rocznik oraz klasę. Inaczej mówiąc, zmienna ta identyfikuje konkretną grupę uczniów zdających egzamin jako jeden oddział. Wylosowanych zostało 1 181 oddziałów. Teoretycznie możliwe było wylosowanie 1 200 klas (2 klasy z każdej szkoły i każdego z 3 roczników), jednak niektóre z 200 wylosowanych szkół posiadały tylko jedną klasę w danym roku. Nie mniej zdecydowana większość gimnazjów posiada dwie klasy w każdym roczniku. Ucznia jednoznacznie identyfikują zmienne *klasa* oraz *uczen_id*. Ta ostatnia zmienna porządkuje uczniów w klasie, przypisując im losowy numer od 1 do 35 (maksymalna wielkość oddziału w bazie).

Trzeba jeszcze raz podkreślić, że zmienne identyfikujące szkoły, oddziały oraz uczniów przypisane zostały w sposób losowy. W pierwszym kroku uczniom przypisano wartości dwóch niezależnych zmiennych utworzonych przez generator liczb losowych. Następnie obliczono średnie z tych zmiennych, osobno dla szkoły i dla klasy. W kolejnym kroku uporządkowano wg tych wartości zbiór, przypisując kolejne liczby jako identyfikatory szkoły oraz klasy. Dla uczniów zbiór posortowano wg liczby losowej osobno dla każdej klasy, a następnie przypisano im kolejne liczby porządkowe jako identyfikator.

Podobnie stworzono losowy identyfikator szkoły podstawowej ucznia. Procedura ta powoduje, że identyfikatory nie są związane ani ze sobą nawzajem, ani z żadnymi innymi zmiennymi w zbiorze, pochodzą bowiem z niezależnych losowych rozkładów. Ewentualny ich związek z jakąkolwiek zmienną musi mieć charakter przypadkowy. W przekazywanej do analiz bazie nie korelują one z żadną zmienną. Nazwy, interpretację i wartości identyfikatorów podsumowuje tabela poniżej.

Tabela 8. Identyfikatory gimnazjum, klasy, ucznia oraz szkoły podstawowej.

Jednostka do zidentyfikowania	Zmienna	Opis zmiennej i jej zastosowania
Gimnazjum	*szkola*	Jednoznacznie identyfikuje gimnazjum. Ma tę samą wartość dla wszystkich uczniów 3 roczników uczęszczających do danego gimnazjum.
Oddział	*klasa*	Jednoznacznie identyfikuje oddział, do jakiego przypisany był uczeń w czasie egzaminu gimnazjalnego. Zmienna ta przyjmuje inne wartości dla oddziałów tego samego gimnazjum z różnych lat.
Uczeń	*uczen_id*	Identyfikuje ucznia w danym oddziale. Różni uczniowie mogą mieć tę samą wartość tej zmiennej. Jednoznaczny identyfikator każdego ucznia tworzy kombinacja zmiennych *klasa* oraz *uczen_id*.
Szkoła podstawowa	*szkola_sp*	Jednoznacznie identyfikuje szkołę podstawową, do której uczęszczał uczeń gimnazjum. Przyjmuje tę samą wartość dla wszystkich uczniów, którzy uczęszczali do tej samej szkoły podstawowej, niezależnie od rocznika i gimnazjum, w jakim się znaleźli.

C. Wagi i inne zmienne umożliwiające szacowanie parametrów w populacji

W tabeli 9. przedstawiono nazwy i opisy wag policzonych dla wylosowanej próby oraz nazwy i opisy innych niezbędnych zmiennych potrzebnych do estymacji. W pierwszej kolumnie podano nazwę zmiennej znajdującej się w bazie danych, w drugiej jej etykietę, a w 3 krótki opis.

Tabela 9. Zmienne i ich opisy niezbędne dla uwzględnienia wielopoziomowego charakteru próby.

Zmienna	Etykieta	Opis
warstwa	Warstwa główna: wynik oraz lokalizacja	Zmienna opisująca warstwy losowania
pw_szkola	Waga: poziom szkoły	Waga dla szkół; odwrócone prawdopodobieństwo wylosowania danej szkoły
pw_klasa	Waga: poziom oddziału	Waga dla oddziałów; odwrócone prawdopodobieństwo wylosowania danego oddziału (z danego roku)
pw_ucznia	Waga finalna dla ucznia	Waga końcowa = *pw_pps1*pw_klas*
fpc1	Korekta dla skończonej populacji	Liczba szkół w populacji
fpc2	Korekta dla skończonej populacji: poziom oddziałów	Liczba oddziałów w kolejnych rocznikach w danej szkole

D. Zmienne z SIO

W obecnej wersji bazy danych zmiennych SIO nie dołączono.

E. Zmienne z BDR GUS

Ważnym źródłem danych o finansach i organizacji oświaty jest Bank Danych Regionalnych (BDR) Głównego Urzędu Statystycznego zawierający bogaty zestaw charakterystyk na poziomie jednostek samorządu terytorialnego. Przy budowie zmiennych do bazy wykorzystano dane o wydatkach i dochodach gmin oraz liczbie uczniów w gimnazjach, szkołach podstawowych i przedszkolach z lat 1996-2006. W chwili budowania bazy pełne dane dla gmin były dostępne w BDR GUS tylko dla tego okresu.

Zmienne określające wydatki na oświatę skonstruowano tak, aby odzwierciedlały przeciętną kwotę, jaką dana gmina przeznaczała na sfinansowanie nauki ucznia przez cały okres jego edukacji na danym etapie. Tak więc dla gimnazjum odpowiada to przeciętnym wydatkom na ucznia przez 3 lata nauki w gimnazjum, a dla szkół podstawowych przeciętnym wydatkom na szkoły podstawowe podczas 6 lat nauki ucznia w danej gminie. Nie rozróżnia się przy tym wydatków na kształcenie w różnych klasach, ani wydatków na płace czy utrzymanie budynku (takie dane docelowo dostępne mają być w bazach SIO). Wszystkie wydatki uwzględniają inflację i zostały wyrażone w pieniądzach realnych z 2005 roku po zastosowaniu wskaźnika inflacji HICP Eurostatu.

Warto zauważyć, że ze względu na dostępność danych oraz przyjęte założenia, zmienne określające wydatki zakładają, że wszyscy uczniowie w gminie finansowani byli na podobnym poziomie i wszyscy uczniowie zdający egzamin odbyli swoją edukację na danym etapie w tej samej gminie. Pierwsze założenie wymuszone jest tym, że w momencie tworzenia bazy nie dysponowaliśmy danymi o kosztach prowadzenia pojedynczych szkół. Jest ono jednak bliskie realiom polskiego szkolnictwa. Gminy starają się w większości przypadków dzielić środki równo między podległe im placówki. Po spełnieniu obowiązku opłacenia pensji nauczycielskich i kosztów prowadzenia budynków władze samorządowe rzadko kiedy stosują mechanizmy różnicujące dodatkowe, będące w ich pełnej dyspozycji, wypłaty. Różnice w wydatkach na szkoły związane są głównie z różnicami w stopniu awansu zawodowego kadry nauczycielskiej. Po pokryciu przez gminy obowiązkowej części płac nauczycielskich szkoły, nauczyciele i uczniowie są w większości przypadków finansowani na podobnym poziomie. Trzeba jednak pamiętać, że jeśli w rzeczywistości szkoły różnią się stopniem finansowania, to skonstruowane zmienne obarczone będą dużym błędem pomiaru. W przyszłych wersjach bazy postaramy się dostarczyć wskaźniki finansowania dla każdej szkoły. Także drugie założenie podyktowane jest brakiem odpowiednich danych. Musimy przyjąć, że każdy, a przynajmniej zdecydowana większość uczniów odbyła naukę na danym etapie w jednej gminie, tej samej, w której uczeń zdawał egzamin zewnętrzny. Nie posiadamy bowiem żadnej wiedzy co do ścieżki edukacyjnej ucznia i jedyna informacja, jaką mamy, dotyczy momentu zdawania egzaminu. Biorąc pod uwagę dość niewielki stopień migracji między gminami rodzin z dziećmi w Polsce, takie założenie wydaje się w zdecydowanej większości przypadków prawomocne.

Kwestią istotną przy konstruowaniu zmiennych określających wydatki na ucznia jest okres trwania roku budżetowego, pokrywający się z rokiem kalendarzowym, który jest inny niż okres trwania roku szkolnego, za który uznaliśmy czas od początku września danego roku do końca sierpnia kolejnego roku. Trzeba przyjąć kryteria rozdzielenia ogólnej kwoty wydatkowanej z budżetu gminy w latach kalendarzowych na roczniki uczniów, którzy uczęszczali w tym czasie do szkół. Dodatkowy problem związany jest z brakiem danych o wydatkach na gimnazja w 2007 roku, które przybliżone muszą zostać przez wydatki z roku 2006 (przypomnijmy, że wydatki z 2007 roku nie były dostępne w BDR GUS w momencie tworzenia bazy). I tak dla każdego rocznika zsumowano wydatki w pieniądzu realnym od września roku kalendarzowego, w którym rozpoczął naukę na danym poziomie, do sierpnia roku, w którym odbył się egzamin zewnętrzny. Tę kwotę podzielono przez sumę uczniów z wszystkich roczników uczęszczających w tym okresie do szkół na danym poziomie i w danej gminie. Dla przykładu, dla kohorty 2002/2005:

- wydatki na ucznia gimnazjum określono przez zsumowanie 1/3 wydatków na gimnazja z 2002 roku (co odpowiada nakładom od września 2002 roku), całości wydatków na gimnazja z lat 2003 i 2004 oraz 2/3 wydatków z 2005 roku; sumę tę podzielono przez liczbę uczniów uczęszczających do gimnazjów w danej gminie od roku szkolnego 2002/2003 do roku szkolnego 2004/2005;

- wydatki na ucznia szkoły podstawowej danej gminy obliczono w podobny sposób, sumując wydatki od września 1996 do sierpnia 2002.

Podobnie postąpiono dla kohorty 2003/2006 oraz 2004/2007 z tą różnicą, że w tym ostatnim przypadku wydatki na gimnazja z 2007 zastąpiono wydatkami z 2006[10].

Dla każdego ucznia – biorąc pod uwagę kod nadany przez OKE szkole podstawowej, do której uczęszczał, zawierający w sobie kod gminy – dołączono też informacje o opiece przedszkolnej. Informacje te pochodzą z roku, w którym dany uczeń powinien uczęszczać do klasy zero. Stworzono przy tym trzy wskaźniki, z których jedynie udział 6-latków w przygotowaniu przedszkolnym odpowiada rzeczywistym danym z roku poprzedzającego rozpoczęcie nauki w szkole podstawowej. Dla dwóch pozostałych wskaźników, procentu 3-5-latków w przedszkolach oraz procentu 3-6-latków w przedszkolach lub klasach zero, także wykorzystano dane tylko z jednej roku, poprzedzającego naukę w szkole podstawowej. Tak więc dla tych wskaźników założono, że w ciągu kilku lat dostępność opieki przedszkolnej była podobna. Założenie to wymuszone jest brakiem danych w BDR GUS o opiece przedszkolnej przed 1995 rokiem. Tak więc dla pierwszego z analizowanych roczników niemożliwe było określenie partycypacji w opiece przedszkolnej dla wcześniejszych lat niż 1995 rok (rocznik ten w 1996 roku rozpoczął naukę w szkole podstawowej). Dla kolejnych lat dostępność danych jest większa, jednak przyjęto podobną zasadę konstrukcji wskaźników, aby utrzymać ich wspólną definicję. Założono też, że uczeń mieszkał w tym okresie w gminie, w której zdawał sprawdzian szóstoklasistów. Wszystkie wskaźniki partycypacji w opiece przedszkolnej obliczono, dzieląc liczbę dzieci ją objętych przez liczebność odpowiednich roczników zawartą w bazach BDR GUS.

Użytkownicy bazy mogą korzystać z dowolnego wskaźnika partycypacji w opiece przedszkolnej, warto jednak zauważyć, że o ile procent 6-latków odpowiada rzeczywistemu udziałowi uczniów sześcioletnich w przygotowaniu przedszkolnym, to jest też najmniej zróżnicowany między gminami. Największe zróżnicowanie dotyczy tu opieki nad 3-5-latkami, gdzie wiele gmin zupełnie wycofało się z dostarczania tej usługi, a niektóre umożliwiają naukę w przedszkolach większości dzieci.

[10] Dla sprawdzenia, czy tego rodzaju przybliżenie jest prawomocne, zastosowano podobną procedurę dla pełnych danych z lat 2002-2005, przyjmując dla 2005 sumę wydatków z 2004 roku. Nie zmieniło to jednak wyników żadnych analiz prowadzonych z tymi zmiennymi (patrz przykłady w ostatniej części podręcznika), co daje podstawę do zastosowania tego rodzaju przybliżenia dla innych lat. Faktem jest, że sumy wydatków w gminach zmieniają się między latami w znacznie mniejszym stopniu niż wydatki na ucznia, na które wpływ mają trendy demograficzne.

Do bazy dołączono też wskaźnik dotyczący sytuacji finansowej gminy i jej mieszkańców. Obliczono dochody całkowite gminy przypadające na jednego mieszkańca, zarówno dla gminy, w której uczeń odbywał naukę w gimnazjum, jak i dla gminy, w której odbywał naukę w szkole podstawowej i prawdopodobnie korzystał też z opieki przedszkolnej. Podobnie jak w przypadku wydatków na szkoły, dochody na mieszkańca obliczono, sumując wydatki z okresu pokrywającego się z latami szkolnymi całego okresu nauki ucznia na danym etapie (1/3 dochodów z roku kalendarzowego, w którym uczeń rozpoczął naukę, po 2/3 dochodów z roku, w którym naukę skończył). Kwotę tę podzielono przez sumę liczby mieszkańców ze wszystkich tych lat, udostępnianą także w BDR GUS. Dochody gminy na mieszkańca mogą być wykorzystane jako wskaźnik zasobności budżetu do dyspozycji władz gminy, ale i jako przybliżony wskaźnik zamożności mieszkańców gminy, bowiem dochody gminy w dużym stopniu składają się z podatków PIT i CIT opłacanych przez mieszkańców i firmy mające siedzibę w danej gminie. Dla pojedynczych przypadków wskaźniki te są dalekie od doskonałości. Budżet gminy jest często mało elastyczny i wysoki dochód na mieszkańca niekoniecznie odzwierciedla środki, jakie gminy mogą swobodnie wydatkować na cele oświatowe czy społeczne, nie mówiąc o efektywności ich wydatkowania. Tym bardziej dochody gminy są jedynie przybliżonym wskaźnikiem zasobności mieszkańców, bowiem samorządy posiadają też spore źródła dochodów niezależnych od zamożności ludności, a ponadto polski system transferów zawiera w sobie mechanizm wyrównywania budżetów gmin. Jednak dla większej próby wskaźniki te mogą być stosowane, choć interpretacja musi być ostrożna (patrz przykłady analiz z ich wykorzystaniem w ostatnim rozdziale).

Tak więc wszystkie wskaźniki dotyczące wydatków na szkoły oraz dochodów gmin na mieszkańca odpowiadają okresom dłuższym niż jeden rok. O ile możliwe jest skonstruowanie wydatków i dochodów dla pojedynczych lat, to z dwóch względów wydaje się to bezcelowe. Po pierwsze, wskaźniki jednoroczne mają dużo większe zróżnicowanie, wynikające np. z jednorazowych inwestycji (znacznie większe wydatki w danym roku) czy grantów (znacznie większe dochody w danym roku). Po drugie, takie zmienne, mimo jednostkowych silnych wahań, są silnie skorelowane między latami, co powoduje, że niemożliwe jest ich wykorzystanie np. w modelu regresji, ze względu na współliniowość. Wydaje się, że wskaźniki „ciągnione" dochodów i wydatków dużo lepiej opisują realną sytuację szkół i gmin, szczególnie w dłuższym okresie.

Tak opracowane dane BDR połączono ze zbiorem CKE za pomocą kodu GUS zawierającego numer województwa, powiatu oraz gminy. W przypadku kilkudziesięciu gmin zmieniono kody na obowiązujące w latach poddanych analizie, aby np. móc połączyć dane z lat 90. z danymi z lat 2000-2006. Kilka gmin, które zostały podzielone lub scalone z innymi, wykluczono z analiz ze względu na niemożność ścisłego określenia, do której z nich uczęszczali uczniowie. Dane Warszawy z okresu, gdy dzielnice były samodzielnymi gminami zagregowano do poziomu całego miasta. Okazało się, że żadna z wylosowanych szkół nie należy do gminy, dla której nie udało się dołączyć danych o wydatkach dla wszystkich lat. Dla uczniów kilku gmin nie udało się natomiast dołączyć danych o partycypacji w opiece przedszkolnej. Do bazy dołączono także typ gminy, w której znajduje się gimnazjum ucznia, wg klasyfikacji GUS. Opis wszystkich zmiennych skonstruowanych z danych BDR GUS przedstawia tabela poniżej.

Tabela 10. Zmienne skonstruowane z danych BDR GUS.

Zmienna	Opis
naklady_gim	Nakłady z ostatnich 3 lat na ucznia gimnazjum w gminie, w której uczeń zdawał egzamin gimnazjalny
naklady_sp	Nakłady z ostatnich 6 lat na ucznia szkoły podstawowej w gminie, w której uczeń zdawał sprawdzian szóstoklasistów
udzial6	Procent 6-latków w przedszkolach lub klasach zero w gminie, w której uczeń zdawał sprawdzian szóstoklasistów (dane z roku poprzedzającego naukę w szkole)
udzial	Procent 3-6-latków objętych opieką przedszkolną lub w klasach zero w gminie, w której uczeń zdawał sprawdzian szóstoklasistów (dane z roku poprzedzającego naukę w szkole)
udzial35	Procent 3-5-latków objętych opieką przedszkolną w gminie, w której uczeń zdawał sprawdzian szóstoklasistów (dane z roku poprzedzającego naukę w szkole)
doch_mieszk	Dochody budżetu na mieszkańca w gminie, w której uczeń zdawał egzamin gimnazjalny
doch_mieszk_sp	Dochody budżetu na mieszkańca w gminie szkoły podstawowej ucznia, w której zdawał sprawdzian szóstoklasistów
typ	Typ GUS gminy, w której uczeń zdawał egzamin gimnazjalny: wiejska, miejska (w tym miasta-powiaty), miejsko-wiejska

4. Jak analizować dane w bazie

A. Uwzględnienie wag i schematu losowania: estymacja za pomocą pakietu statystycznego Stata

Dla estymacji w złożonych schematach losowania niezbędne jest uwzględnienie schematu losowania. Chociaż poprawne wartości podstawowych statystyk, takich jak średnia, mediana czy odchylenie standardowe, można uzyskać, stosując jedynie wagi, to już wszelkie testy statystyczne czy oszacowanie estymatorów przedziałowych, np. przedziału ufności dla średniej, wymagają uwzględnienia schematu losowania. Poniżej przedstawiamy, w jaki sposób zdefiniować schemat losowania w programie Stata w wersji 10, który posiada rozbudowane narzędzia do analizy prób pochodzących ze złożonych schematów losowania (a takimi są zawsze dane edukacyjne). Jednak inne pakiety, np. SAS czy SPSS, także umożliwiają uwzględnienie struktury losowania w wielu procedurach statystycznych. Użytkownik znający te pakiety powinien łatwo wykorzystać poniższe przykłady do przeprowadzenia podobnych analiz.

Program Stata wymaga zdefiniowania struktury losowania komendą **svyset** o poniższej składni:

```
svyset psu [weight] [, design_options] [|| ssu , design_options] [options]
```

gdzie:

psu	oznacza **primary sampling unit**, czyli jednostkę, która losowana była jako pierwsza w schemacie losowania
weight	oznacza wagę finalną; pakiet Stata dopuszcza 4 rodzaje wag, jednak tutaj należy podać tzw. **probability weight** dla ostatniego poziomu losowania, czyli wagę finalną ucznia zawartą w bazie
design_options	po przecinku definiujemy opcje dotyczące schematu losowania; najważniejsze opcje to: **fpc** – czyli wartość potrzebna do obliczenia korekty błędu standardowego dla skończonej populacji, oraz **strata**, czyli opcja definiująca zmienną użytą do warstwowania
||	oddziela zmienne definiujące kolejny poziom losowania
ssu	oznacza **second sampling unit**, czyli element, który losowany jest w obrębie wcześniej wylosowanej jednostki obserwacji z wyższego poziomu
options	oznaczają dodatkowe opcje

Dla bazy danych EWD-CKE poprawne zdefiniowanie schematu losowania wygląda następująco:

```
svyset szkola [pw=pw_ucznia], fpc(fpc1) strata(warstwa) || klasa, fpc(fpc2)
strata(rok) singleunit(scaled)
```

Warto podkreślić, iż wagę definiujemy tu jedynie raz, dla poziomu ucznia. Zmienne określające warstwy i poprawkę na skończoną populację definiujemy osobno dla kolejnych stopni losowania. Zmienną określającą warstwę na drugim stopniu została zmienna *rok*, określająca rok zdawania egzaminu przez uczniów. Według schematu losowania z każdego roku (rocznika), niezależnie wylosowywane były przynajmniej po dwie szkoły. Opcję **singleunit(scaled)** dodano po to, by można było obliczyć błędy standardowe przy założeniu jednoelementowych warstw.

Po zdefiniowaniu schematu możemy wykorzystać kilkadziesiąt procedur estymacji uwzględniających złożone schematy losowania. Wystarczy przed każdą z nich dodać prefiks **svy**, a przy estymacji zdefiniowanej wcześniej schemat zostanie w pełni uwzględniony. Na przykład, aby poprawnie oszacować średnie z przedziałami ufności, osobno dla każdego roku, wystarczy wpisać komendę:

```
svy: mean mat hum spr, over(rok)
```

Poniżej zaprezentowano wynik oszacowań uzyskany w programie Stata 10:

```
Survey: Mean estimation

Number of strata =     100         Number of obs    =    27259
Number of PSUs   =     200         Population size  =  1302379
Design df        =     100

2005: rok = 2005
2006: rok = 2006
2007: rok = 2007

Linearized
Over        Mean   Std. Err.     [95% Conf. Interval]

mat
2005    24.79096   .2274684      24.33967    25.24225
2006     24.4314   .1865269      24.06134    24.80147
2007    25.44326   .1784265      25.08927    25.79725

hum
2005    33.92983   .2041425      33.52482    34.33484
2006    32.02564   .1773533      31.67377     32.3775
2007    32.15991   .1441616       31.8739    32.44592

spr
2005     30.0339   .1423735      29.75143    30.31636
2006    29.20525   .1307743       28.9458    29.46471
2007    26.11809   .1351333      25.84999    26.38619
```

Po tej komendzie możemy uzyskać dodatkowe statystyki, w tym oszacowania wariancji zmiennych w populacji przez komendę **estat sd**. Po jej wpisaniu uzyskujemy oceny odchyleń standardowych:

```
Over   Mean          Std. Dev.

mat
2005   24.79096      9.958292
2006   24.4314       9.98488
2007   25.44326      10.0985

hum
2005   33.92983      8.241717
2006   32.02564      7.953628
2007   32.15991      9.289629

spr
2005   30.0339       6.295349
2006   29.20525      6.331303
2007   26.11809      7.499112
```

Dociekliwy czytelnik może sprawdzić, że uzyskane tak oszacowania są bliskie wartościom w populacji (proszę porównać z tabelą 6.). W szczególności przedziały ufności dla średnich wyników egzaminów zawierają prawdziwe wartości średnich wyników w próbie, z której dokonano losowania, oprócz części matematyczno-przyrodniczej z 2007 roku, gdzie średnia w populacji jest o 0,02 większa niż górna granica przedziału ufności. Są to jednak wartości pomijalne z praktycznego punktu widzenia.

W większości przypadków zadowalające efekty z praktycznego punktu widzenia da uwzględnienie podstawowych informacji dotyczących schematu losowania, przede wszystkim podstawowej jednostki losowania, czyli „PSU", oraz wag na poziomie ucznia. Poniższe komendy pokazują, że podanie tylko tych informacji daje identyczne oceny średnich (podobnie będzie dla odchyleń standardowych), jednak nieco inne oceny błędów standardowych i przedziały ufności. Czytelnik może porównać te wyniki z podanymi powyżej, gdzie uwzględniono wszystkie informacje o schemacie losowania, żeby stwierdzić, że różnice w ocenach błędów standardowych średnich są co mniejsze niż 0,1 punktu. Są więc w tym przypadku pomijalne ze względów praktycznych. Tak więc, o ile pakiet statystyczny wykorzystywany do analiz przez użytkownika nie daje możliwości zdefiniowania wszystkich aspektów schematu losowania, wystarczy zdefiniować PSU oraz wagi dla uczniów, aby uzyskać bardzo bliskie poprawnym rezultaty. Takie możliwości dają wszystkie popularne pakiety statystyczne. Oczywiście, ponieważ rezultaty te nie są w pełni poprawne, należy je interpretować z ostrożnością, a w miarę możliwości wykorzystywać wszystkie zawarte w bazie informacje dotyczące schematu losowania.

```
svyset szkola [pw=pw_ucznia]

       pweight: pw_ucznia
       VCE: linearized
       Single unit: missing
       Strata 1: <one>
       SU 1: szkola
       FPC 1: <zero>

svy: mean mat hum spr, over(rok)

       (………………)

       Over        Mean   Std. Err.     [95% Conf. Interval]

       mat
       2005    24.79096   .2927988      24.21357    25.36834
       2006     24.4314   .2693666      23.90022    24.96258
       2007    25.44326    .276535      24.89794    25.98857

       hum
       2005    33.92983    .254771      33.42743    34.43223
       2006    32.02564   .2142713       31.6031    32.44817
       2007    32.15991    .242742      31.68124    32.63859

       spr
       2005     30.0339   .1705966      29.69749    30.37031
       2006    29.20525   .1671325      28.87567    29.53483
       2007    26.11809   .1841096      25.75504    26.48115
```

B. Regresja liniowa z korektą błędów standardowych ze względu na pogrupowanie obserwacji

Dane edukacyjne zawierające dane na poziomie ucznia mają zawsze strukturę, w której występują grupy o podobnych cechach. Najczęściej są to uczniowie tej samej szkoły i klasy, tak jak w przypadku omawianej bazy. Pominięcie tego faktu przy estymacji jakiegokolwiek modelu statystycznego może prowadzić do błędnej oceny statystyk, przede wszystkim do niepoprawnej oceny błędów standardowych interesujących nas estymatorów. Inaczej mówiąc, jeśli nie weźmiemy pod uwagę pogrupowania uczniów, to możemy np. zbyt optymistycznie wnioskować o stwierdzonych różnicach czy zależnościach między zmiennymi, będąc mylnie przekonanym, że wykonane testy statystyczne mające je potwierdzić są prawidłowe. Będzie tak np. gdy przeprowadzimy zwykły t-test na różnice średnich czy oszacujemy korelacje jakichkolwiek zmiennych na poziomie uczniów. Podobnie gdy oszacujemy model regresji liniowej, stwierdzając wpływ jakiejś zmiennej na wyniki uczniów, gdyż statystyka t przy tej zmiennej wskazuje na jej istotność. Jednak bez uwzględnienia pogrupowania danych zarówno wartość statystyki t, jak i tzw. „p-value" są najczęściej niepoprawne. Mogą być one zaniżone lub zawyżone w zależności od charakteru podobieństwa wartości zmiennych w grupach. Większość podręczników statystycznych kwestie te pomija, choć dla badań edukacyjnych są one kluczowe. Z tego względu przestrzegamy użytkowników bazy przed stosowaniem opisywanych w podręcznikach procedur, o ile nie są one dostosowane do analiz danych pochodzących ze złożonych schematów losowań.

Jak opisano w poprzednim punkcie, pakiet statystyczny Stata w wersji 10 pozwala na uwzględnienie nawet bardzo złożonych schematów losowań. Większość pakietów statystycznych posiada podobne procedury. Poniżej pokazano, w jaki sposób przeprowadzić regresję liniową w pakiecie Stata, uwzględniając schemat losowania, oraz jakie niesie ze sobą konsekwencje pominięcie takich informacji, jak wagi czy podstawowa jednostka losowania decydująca o pogrupowaniu danych. W kolejnym rozdziale opisano modele wielopoziomowe, które niejako automatycznie biorą pod uwagę pogrupowanie obserwacji we wskazanych jednostkach, rozpoznają hierarchiczną strukturę losowania, podając prawidłowe oceny błędów standardowych.

Korektę ze względu na pogrupowanie danych w przypadku szacowania średniej przedstawiono już powyżej dla programu statystycznego Stata (w wersji 10). Generalnie rzecz biorąc, korekta taka możliwa jest dla większości modeli statystycznych. Przykładowo: pakiet Stata 10 posiada 48 estymatorów działających z prefiksem **svy**, które biorą pod uwagę dowolnie złożony schemat losowania. Na przykład dla modelu regresji liniowej procedura **svy: regress** dostarczy ocen parametrów uwzględniających wszystkie aspekty losowania. Jednak można też wykorzystać prostsze modele, biorąc pod uwagę jedynie najważniejsze informacje dotyczące schematu losowania. W przypadku modeli regresji, poza tzw. „survey regression", którą szacuje wspomniana przed chwilą procedura, istnieją dwa najczęstsze sposoby korygowania błędów standardowych. Pierwszym z nich jest oszacowanie zwykłego modelu regresji liniowej ze skorygowanymi ocenami błędów standardowych przez tzw. „sandwich estimator". Tak oszacowane błędy standardowe nazywane są też „robust", „cluster-robust", „sandwich" lub „Huber-White". W programie Stata uzyskuje się je bardzo prosto, przez podanie opcji **cluster()** po komendzie **regress**, wskazując identyfikator grupy. Drugim sposobem jest oszacowanie modelu regresji z efektami losowymi przypisanymi podstawowej jednostce losowania. Są to tzw. modele wielopoziomowe („multilevel models") lub hierarchiczne modele liniowe („hierarchical linear models") (por. Rabe-Hesketh, Skrondal, 2008; Goldstein, 1999; Raudenbush, Bryk, 2002). Modele zostały rozwinięte specjalnie do analizy danych mających hierarchiczną strukturę, przede wszystkim danych edukacyjnych, i uzyskiwanie poprawnych ocen błędów standardowych jest jedynie jedną z cech stanowiących o ich rosnącej popularności. Modele te omawiamy nieco szerzej w części z przykładami analiz.

Poniżej porównano oceny uzyskane przez różne modele regresji oszacowane na tych samych danych z omawianej bazy. Zmienną zależną jest tu wynik egzaminu gimnazjalnego w części humanistycznej, a zmiennymi niezależnymi (objaśniającymi) są płeć oraz wynik sprawdzianu. Jednostką obserwacji jest uczeń. Zacznijmy od zwykłej regresji bez żadnej korekty błędów standardowych. Wykorzystano tylko dane dla rocznika 2004/2007. Komendy poprzedzono kropką. Dla przejrzystości pominięto znaczną część wyników estymacji standardowo pokazywanych przez program Stata.

```
. keep if rok==2007

(18207 observations deleted)


. regress hum plec spr

     Number of obs =    9054                            R-squared     =  0.6065

     ------------------------------------------------------------------------------
              hum |      Coef.   Std. Err.       t    P>|t|     [95% Conf. Interval]
     -------------+----------------------------------------------------------------
             plec |   2.349126   .1230122    19.10   0.000     2.107994    2.590258
              spr |   .9379975   .0082626   113.52   0.000     .9218008    .9541941
            _cons |   6.362624   .2250537    28.27   0.000     5.921468     6.80378
     ------------------------------------------------------------------------------
```

Jak widać, zarówno płeć ucznia (1 = dziewczynka, 0 = chłopiec), jak i wynik na sprawdzianie są dobrymi predyktorami wyników w części humanistycznej. Wg tego podstawowego modelu obie są istotne statystycznie (ich efekt jest różny od zera). Dziewczynki, biorąc pod uwagę ich wiedzę mierzoną wynikiem sprawdzianu, uzyskują wyższe wyniki na egzaminie. Jeden punkt na sprawdzianie podnosi przewidywany wynik na egzaminie o 0,94 punktu. Te dwie zmienne są zdolne wyjaśnić ok. 60% wariancji wyników egzaminu gimnazjalnego, co oznacza bardzo wysoką moc objaśniającą modelu.

Kolejna komenda wykonuje podobny model regresji, jednak błędy standardowe zostają skorygowane o pogrupowanie obserwacji wewnątrz szkół.

```
. regress hum plec spr, cluster(szkola)

     Linear regression                                     Number of obs =    9054
                                                           R-squared     =  0.6065

                               (Std. Err. adjusted for 200 clusters in szkola)
     ------------------------------------------------------------------------------
                  |               Robust
              hum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
     -------------+----------------------------------------------------------------
             plec |   2.349126   .1264684    18.57   0.000     2.099736    2.598516
              spr |   .9379975   .0103766    90.40   0.000     .9175353    .9584596
            _cons |   6.362624   .3401696    18.70   0.000     5.691824    7.033424
     ------------------------------------------------------------------------------
```

Jak widać, skorygowane o pogrupowanie obserwacji błędy standardowe są nieco większe, jednak oceny parametrów dla obu zmiennych są takie same. Obydwa modele nie biorą pod uwagę wag przypisanych uczniom, szkołom i klasom, przez co oceny parametrów nie odpowiadają tym w całej populacji. Możliwe jest wykonanie regresji z wagami analitycznymi na poziomie ucznia (przez podanie po nazwach zmiennych opcji **[aw=pw_ucznia]**). Jednak w pełni prawidłowym modelem jest „survey regression”, którą można wykonać przez dodanie prefiksu **svy**, zakładając, że schemat losowania został wcześniej zdefiniowany procedurą **svyset**, tak jak to opisano powyżej.

```
. svy: regress hum plec spr

     Survey: Linear regression

     Number of strata   =       100                  Number of obs      =       9054
     Number of PSUs     =       200                  Population size    = 423229.47
                                                     R-squared          =     0.6044

     ------------------------------------------------------------------------------
                  |             Linearized
              hum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
     -------------+----------------------------------------------------------------
             plec |   2.342444   .1295406    18.08   0.000     2.085439    2.599449
              spr |   .9317179   .0102377    91.01   0.000     .9114067    .9520292
            _cons |   6.628612   .2962125    22.38   0.000     6.040935    7.216289
     ------------------------------------------------------------------------------
```

Tym razem współczynniki przy obu zmiennych niezależnych zmieniły się, choć nieznacznie, co wynika z „autoważenia” próby, dzięki dobraniu optymalnego schematu losowania. Błędy standardowe i statystyki t-studenta są bardzo podobne do uzyskanych przez regresję z korektą błędów standardowych. Nie znaczy to jednak, że podobnie niewielkie różnice będą występować zawsze. Nawet dla tych samych danych mogą być one znaczne, o ile badać będziemy podgrupy uczniów, a w równaniu pojawią się zmienne, które są silnie skorelowane wewnątrz szkół lub zmienne z charakterystykami szkół.

Jak już wspominaliśmy, opracowano specjalne modele regresyjne biorące pod uwagę hierarchiczną strukturę danych edukacyjnych, a więc to, że uczniowie przypisani są klasom, klasy szkołom itp. W programie Stata istnieją trzy procedury, które mogą być wykorzystane do oszacowania modeli wielopoziomowych – **xtreg**, **xtmixed** oraz napisana przez użytkowników procedura **gllamm** stanowiąca w istocie osobną platformę o niemal nieograniczonych możliwościach, służącą do estymacji nawet najbardziej złożonych modeli (por. Rabe-Hesketh, Skrondal, 2008). Zacznijmy od procedury **xtmixed**, stanowiącej oficjalną procedurę do modeli wielopoziomowych z tzw. mieszanymi efektami (losowymi i stałymi). Poniższe dwa przykłady pokazują jej zastosowanie dla modelu uwzględniającego 2 poziomy analizy: ucznia oraz szkoły, a także modelu 3-poziomowego, uwzględniające poziom ucznia, klasy i szkoły.

```
. xtmixed hum plec spr || szkola:

     Mixed-effects REML regression                   Number of obs      =      9054
     Group variable: szkola                          Number of groups   =       200

                                                     Obs per group: min =        21
                                                                    avg =      45.3
                                                                    max =        60


                                                     Wald chi2(2)       =  13708.48
     Log restricted-likelihood =  -28467.97          Prob > chi2        =    0.0000

     ------------------------------------------------------------------------------
              hum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
     -------------+----------------------------------------------------------------
             plec |   2.275387   .1177176    19.33   0.000     2.044664    2.506109
              spr |   .9283971   .0082533   112.49   0.000     .9122209    .9445734
            _cons |   6.636808   .2612053    25.41   0.000     6.124855     7.14876
     ------------------------------------------------------------------------------

     ------------------------------------------------------------------------------
       Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
     -----------------------------+------------------------------------------------
     szkola: Identity             |
                        sd(_cons) |   1.912615    .114407      1.701027    2.150522
     -----------------------------+------------------------------------------------
                     sd(Residual) |   5.497337   .0413204      5.416944    5.578923
     ------------------------------------------------------------------------------
```

```
. xtmixed hum plec spr || szkola: || klasa:

      Mixed-effects REML regression                   Number of obs      =      9054

      -----------------------------------------------------------
                      |   No. of       Observations per Group
       Group Variable |   Groups    Minimum    Average    Maximum
      ----------------+------------------------------------------
               szkola |      200         21       45.3         60
                klasa |      393         15       23.0         32
      -----------------------------------------------------------

                                                      Wald chi2(2)       =  13333.00
      Log restricted-likelihood = -28432.364          Prob > chi2        =    0.0000

      ------------------------------------------------------------------------------
               hum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
      -------------+----------------------------------------------------------------
              plec |   2.269424   .1171248    19.38   0.000     2.039864    2.498985
               spr |   .9228812   .0083253   110.85   0.000     .9065639    .9391985
             _cons |   6.772981   .2624587    25.81   0.000     6.258571     7.28739
      ------------------------------------------------------------------------------

      ------------------------------------------------------------------------------
        Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
      -----------------------------+------------------------------------------------
      szkola: Identity             |
                        sd(_cons)  |   1.705729   .1353897      1.459981    1.992842
      -----------------------------+------------------------------------------------
      klasa: Identity              |
                        sd(_cons)  |    1.22929   .1178032      1.018787    1.483288
      -----------------------------+------------------------------------------------
                     sd(Residual)  |   5.429201   .0412716       5.34891    5.510697
```

Modele te dają nieco inne oceny parametrów niż modele zwykłej regresji. Różnią się także oceny błędów standardowych. W tym przypadku różnice są jednak niewielkie. Modele wielopoziomowe mają jednak tę zaletę, że umożliwiają oszacowanie wariancji w zmiennej zależnej na poziomie szkół oraz na poziomie uczniów. W tabelach wyników, w części dotyczącej efektów losowych („Random-effects Parameters”) widoczne są oszacowania odchyleń standardowych efektów szkół oraz klas. Dla pierwszego modelu tylko z efektami szkół możemy obliczyć wariancję efektów szkół jako 1,912615^2 = 3,66. W modelu drugim możemy obliczyć wariancje na poziomie szkół oraz klas, które wynoszą odpowiednio 2,91 oraz 1,51. Oceny te możemy uzyskać bezpośrednio dodając opcję **variance** do komendy **xtmixed**. W ten sposób możemy prowadzić analizę wariancji, podobnie jak w modelach ANOVA, jednak wykorzystując modele z efektami losowymi oraz wprowadzając dowolne zestawy zmiennych niezależnych, obserwując, jaki wpływ mają one na oceny wariancji i sprawdzając istotność zmian odpowiednimi testami statystycznymi, np. „likelihood ratio-test”, dostępnym w Stata jako **lrtest**.

Warto zauważyć, że w powyższych modelach wielopoziomowych nie uwzględniono wag. Nie jest to możliwe w procedurze **xtmixed** w wersji 10 pakietu Stata. Umożliwia to jednak procedura **gllamm**, nienależąca do oficjalnego zestawu procedur Stata, jednak posiadająca szeroką dokumentację i wspierana ogromnym autorytetem jej autorów (por. www.gllamm.org). Procedura ta, jako jedna z niewielu dostępnych w pakietach statystycznych, umożliwia wykorzystanie wag na każdym poziomie analizy, co w modelach wielopoziomowych może mieć spore znaczenie. Poniżej podano przykład estymacji dwóch modeli, identycznych do modeli oszacowanych powyżej procedurą **xtmixed**, jednak tutaj wykorzystano wagi i procedurę **gllamm**. Zmienne z wagami dla każdego poziomu należy oznaczyć tą nazwą z inną liczbą na końcu, oznaczającą poziom analizy. W **gllamm** najniższy poziom oznaczany jest 1, kolejny 2 itd. W naszym przypadku musimy więc stworzyć wagi *pw1* dla ucznia oraz *pw2* dla szkoły dla modelu 2-poziomowego; oraz *pw1* dla ucznia, *pw2* dla klasy i *pw3* dla szkoły – dla modelu 3-poziomowego. Odpowiedni kod i wyniki przedstawiono poniżej. Ponownie komendy poprzedzono kropką.

```
. gen pw1=pw_ucznia
. gen pw2=pw_szkola
. gllamm hum plec spr, i(szkola) pw(pw)

     number of level 1 units = 9054
     number of level 2 units = 200

     log likelihood = -34810178

     Robust standard errors
     ------------------------------------------------------------------------------
              hum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
     -------------+----------------------------------------------------------------
             plec |   2.328893   .1187569    19.61   0.000     2.096134    2.561652
              spr |   .9175517   .0113502    80.84   0.000     .8953057    .9397977
            _cons |   7.121886   .3098646    22.98   0.000     6.514562    7.729209
     ------------------------------------------------------------------------------
     Variance at level 1
     ------------------------------------------------------------------------------
      28.666008 (.80791142)

     Variances and covariances of random effects
     ------------------------------------------------------------------------------
     ***level 2 (szkola)
         var(1): 1.1775714 (.07434594)
     ------------------------------------------------------------------------------


. drop pw2
. gen pw2=pw_klasa
. gen pw3=pw_szkola
. gllamm hum plec spr, i(klasa szkola) pw(pw) adapt

     number of level 1 units = 9054
     number of level 2 units = 393
     number of level 3 units = 200

     log likelihood = -63735689

     Robust standard errors
     ------------------------------------------------------------------------------
              hum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
     -------------+----------------------------------------------------------------
             plec |    2.24418    .119857    18.72   0.000     2.009265    2.479095
              spr |   .9027216   .0122023    73.98   0.000     .8788056    .9266377
            _cons |   7.153772   .3486415    20.52   0.000     6.470447    7.837097
     ------------------------------------------------------------------------------
     Variance at level 1
     ------------------------------------------------------------------------------
      27.689731 (.63869302)

     Variances and covariances of random effects
     ------------------------------------------------------------------------------
     ***level 2 (klasa)
         var(1): 1.9816961 (.15894279)

     ***level 3 (szkola)
         var(1): 1.9534778 (.04733188)
     ------------------------------------------------------------------------------
```

C. Modelowanie metodą EWD

Powyższe przykłady regresji stanowią w istocie modele edukacyjnej wartości dodanej, które wykorzystywane są w Polsce do oceny efektywności pracy gimnazjów (por. www.ewd.edu.pl). Zaletą tych modeli jest wzięcie pod uwagę wyników ucznia ze sprawdzianu, określających jego potencjał przed rozpoczęciem nauki w gimnazjum. Dzięki temu wszelkie efekty zmiennych w modelu można traktować jako oceny wpływu tych czynników na **przyrost** wiedzy ucznia podczas nauki w gimnazjum. Stanowi to ogromną różnicę w porównaniu z modelem regresji, gdzie wyniki egzaminu pojawiają się jedynie po lewej stronie równania, czyli jako zmienna zależna. Wtedy wszystkie efekty dotyczą **poziomu** wiedzy ucznia, który jednak zależy przede wszystkim od jego ukrytych zdolności oraz pochodzenia społecznego. Inaczej mówiąc, modele EWD dają możliwość określenia czynników wpływających na rozwój uczniów. To takie czynniki są interesujące z punktu widzenia polityki oświatowej. Na poziom wyników polityka szkoły ma znacznie mniejszy wpływ, bowiem tylko nieliczne szkoły są w stanie wyrównać wyniki uczniów o różnym pochodzeniu społecznym. W większości przypadków nie jest to jednak możliwe, a celem szkoły powinno być jak najefektywniejsze podnoszenie umiejętności uczniów z poziomu, na jakim się oni znajdują. Efekty te można oceniać za pomocą modeli EWD.

W Polsce opracowano i udokumentowano metodę EWD dla gimnazjów (patrz Dolata (red.), 2007; Jakubowski, 2008a; www.ewd.edu.pl). Warianty tej metody wykorzystano w przykładach analizy w kolejnej części podręcznika. Przedstawiono tam modele do oceny efektywności wydatków na gimnazja na przyrost wiedzy uczniów. Są to tylko przykłady analizy udostępnianej bazy, a prace badawcze dotyczące tej samej tematyki i danych przedstawiono w artykułach (por. Jakubowski, 2007, 2008b, 2008d). W pracach tych szczegółowo opisano kontekst, metodologię, zbiory danych oraz interpretację wyników[11]. Ewaluacja metodą EWD może dotyczyć dowolnych aspektów polityki edukacyjnej, nie tylko aspektów finansowych, o ile możliwe są one do skwantyfikowania. Na przykład Jakubowski (2008d) podaje także przykładową analizę wpływu wielkości klasy na postępy uczniów. W polskim kontekście nasuwają się liczne pomysły na podobne badania dotyczące np. wpływu wykształcenia nauczycieli, segregacji uczniów między szkołami i klasami, wpływu różnic w realizowanych programach i metodach nauczania itp. Ważnych kwestii do badania jest sporo, warto więc spojrzeć, jak można wykorzystać modele EWD do różnych celów. W kolejnych wersjach bazy EWD-CKE mamy nadzieję umieszczać szczegółowe wskaźniki, które mogą takim badaniom posłużyć.

11 Prace te dostępne są na stronie www.wne.uw.edu.pl/mjakubowski. Mogą zostać także przesłane przez email: mjakubowski@uw.edu.pl.

5. Przykłady analiz

A. Szacowanie i porównanie średnich wyników między grupami uczniów

Podstawową analizą, jakiej zazwyczaj dokonujemy w każdym badaniu, jest porównanie średnich między grupami uczniów. Poniżej pokazano, w jaki sposób w programie Stata można testować różnice między wynikami egzaminów względem płci oraz miejsca zamieszkania ucznia. Ponownie wykorzystamy procedurę **mean** z prefiksem **svy**. Zakładamy, że wcześniej zdefiniowaliśmy schemat losowania procedurą **svyset**, tak jak poprzednio. Zacznijmy od porównania wyników chłopców i dziewcząt na egzaminie gimnazjalnym i sprawdzianie. Znów ograniczymy się do danych z 2007 roku.

```
. svy: mean hum mat spr, over(plec)

     Survey: Mean estimation

     Number of strata =     100          Number of obs    =    9052
     Number of PSUs   =     200          Population size  =  423134
                                         Design df        =     100

         mezczyzna: plec = mezczyzna
           kobieta: plec = kobieta

     --------------------------------------------------------------
                  |             Linearized
             Over |       Mean   Std. Err.     [95% Conf. Interval]
     -------------+------------------------------------------------
     hum          |
        mezczyzna |   30.14038   .1793127      29.78463    30.49613
          kobieta |    34.0935    .178908      33.73855    34.44845
     -------------+------------------------------------------------
     mat          |
        mezczyzna |   26.07372   .2326769      25.61209    26.53534
          kobieta |   24.83963   .2028633      24.43716    25.24211
     -------------+------------------------------------------------
     spr          |
        mezczyzna |   25.23463   .1819482      24.87365    25.59561
          kobieta |   26.96396    .141394      26.68343    27.24448
     --------------------------------------------------------------
```

Jak widać, na wszystkich egzaminach występują istotne statystycznie różnice między wynikami chłopców i dziewcząt, bowiem przedziały ufności dla chłopców i dla dziewcząt nie pokrywają się. Największe ilościowo różnice występują w części humanistycznej, najmniejsze w części matematyczno-przyrodniczej. Trzeba jednak przypomnieć, że porównanie wielkości tych różnic może się odbywać jedynie przy założeniu, że wyniki są wyrażone na tej samej skali, co nie jest w przypadku polskich egzaminów prawdą. Faktem jest jednak, że dziewczynki lepiej wypadają w części humanistycznej, a nieco gorzej w części matematyczno-przyrodniczej, co zgodne jest z wynikami setek innych badań, w tym badań międzynarodowych, np. PISA lub TIMSS. Spójrzmy, jak wyglądają różnice między uczniami z różnych typów miejscowości.

```
. svy: mean hum mat spr, over(mw)

     Number of strata =     100          Number of obs    =    9052
     Number of PSUs   =     200          Population size  =  423134
                                         Design df        =     100

              wieś: mw = wieś
         _subpop_2: mw = miasto do 20 tys
         _subpop_3: mw = miasto 20-100 tys
         _subpop_4: mw = miasto powyżej 100 tys

     --------------------------------------------------------------
                  |             Linearized
             Over |       Mean   Std. Err.     [95% Conf. Interval]
     -------------+------------------------------------------------
     hum          |
             wieś |   30.91926   .2208091      30.48118    31.35734
        _subpop_2 |   31.02591   .4875985      30.05853    31.99329
        _subpop_3 |    33.6693   .4852418      32.70659    34.63201
        _subpop_4 |    34.4958   .4934636      33.51678    35.47482
     -------------+------------------------------------------------
     mat          |
             wieś |   24.26453   .2897071      23.68976     24.8393
        _subpop_2 |   24.59337   .5661965      23.47005    25.71669
        _subpop_3 |   26.79478   .5628669      25.67807    27.91149
        _subpop_4 |   27.52425   .5767861      26.37993    28.66858
     -------------+------------------------------------------------
     spr          |
             wieś |   25.03666   .1913655      24.65699    25.41632
        _subpop_2 |   25.37399   .3863157      24.60756    26.14043
        _subpop_3 |   27.36251   .3754201      26.61769    28.10733
        _subpop_4 |   27.98501   .3121758      27.36566    28.60436
     --------------------------------------------------------------
```

Na wszystkich egzaminach widzimy tę samą zależność: im większa miejscowość, tym wyższe wyniki uczniów. Jednak w wielu przypadkach przedziały ufności zazębiają się, więc nie możemy w pełni uprawniony sposób stwierdzić, że różnice te są istotne statystycznie. Przykładowo: różnica między wsią a miastami do 20 tys. jest na wszystkich egzaminach nieznaczna i przedziały ufności posiadają wspólne wartości, co sugeruje, że różnica ta może w rzeczywistości nie występować. Oczywiście jest to tylko przykład, bowiem mamy dokładne dane dla całej populacji potwierdzające, że różnice te w istocie istnieją.

Chociaż procedura **mean** pozwala oszacować średnie w podgrupach wraz z przedziałami ufności i porównywać je między sobą[12], to nie podaje ona przedziału ufności dla samej różnicy. Inaczej mówiąc, nie pozwala odpowiedzieć na pytanie, czy różnica między dwiema grupami jest inna niż różnica między dwiema innymi grupami. Taki test można jednak łatwo przeprowadzić, posługując się modelem regresji. Załóżmy, że chcemy sprawdzić, czy różnica w wynikach chłopców i dziewczynek w części humanistycznej (nazwijmy to „efektem płci") jest taka sama na wsi i w dużych miastach. Ograniczamy więc próbę do wsi i miast z ponad 100 tys. mieszkańców, czyli dla wartości zmiennej *mw* równych 0 oraz 3. Szacujemy regresję ze zmienną oznaczającą płeć oraz efektem interakcji między płcią a identyfikatorem dużych miast.

12 Przedziały ufności są tu prawidłowe, jeśli grupy są niezależne statystycznie i porównujemy jedynie dwie średnie między sobą. Dla zależnych grup przedziały ufności powinno skonstruować się nieco inaczej, a dla wielokrotnych porównań niezbędne jest zastosowanie poprawek podających prawdziwy poziom istotności.

```
. gen duze=0 if mw==0
. replace duze=1 if mw==3
. gen inter=plec*duze
. svy, subpop(if mw==0 | mw==3): regress hum plec duze inter

      Survey: Linear regression

      Number of strata   =        84                 Number of obs      =      7514
      Number of PSUs     =       168                 Population size    = 350896.06
                                                     Subpop. no. of obs =      5386
                                                     Subpop. size       = 255048.38

      ------------------------------------------------------------------------------
                   |             Linearized
               hum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
      -------------+----------------------------------------------------------------
              plec |   4.517735   .2963802    15.24   0.000      3.92835    5.107119
              duze |   4.315289   .5669257     7.61   0.000     3.187895    5.442683
             inter |  -1.685091   .5423163    -3.11   0.003    -2.763547   -.606636
             _cons |   28.65138   .2828999   101.28   0.000      28.0888    29.21396
      ------------------------------------------------------------------------------
      Note: 16 strata omitted because they contain no subpopulation members.
```

Wyniki regresji należy odczytać następująco. Dziewczynki uzyskują na wsi wyniki o 4,52 punktu wyższe (współczynnik przy zmiennej płeć). Wszyscy uczniowie dużych miast uzyskują przeciętnie 4,32 punktu więcej niż uczniowie gimnazjów wiejskich. Różnica między wynikami dziewcząt i chłopców jest w dużych miastach mniejsza o -1,69 punktu niż podobna różnica na wsi. Mniejszy efekt płci w dużych miastach potwierdzają testy statystyczne (statystyka t ponad 3, przedział ufności niezawierający zera). Różnicę w wynikach chłopców i dziewcząt w miastach można obliczyć jako 4,517735 - 1,685091 = 2,832644.

Te same obliczenia można uzyskać, szacując procedurą **mean** różnice w wynikach między chłopcami a dziewczętami na wsi (wynosi ona 33,16912 - 28,65138 = 4,51774) i w miastach (wynosi ona 35,79931 - 32,96667 = 2,83264). Odejmując te różnice dostaniemy ocenę różnic w efekcie płci między wsią i miastem, dokładnie taką samą, jak z równania regresji powyżej, jednak bez odpowiedniego przedziału ufności. Widać więc, że model regresji można też łatwo wykorzystać do podobnych testów. Należy jedynie pamiętać, aby zawsze uwzględnić w równaniu identyfikatory grup (tutaj zmienne zerojedynkowe oznaczające duże miasta oraz dziewczęta), a dopiero potem efekt interakcji (tutaj zmienna oznaczająca dziewczęta w dużych miastach). W ten sposób kontrolujemy ogólne różnice w wynikach między grupami (różnicę między wsią a dużym miastem dla wszystkich uczniów oraz różnice między chłopcami i dziewczętami niezależnie od miejsca zamieszkania). Należy też pamiętać, aby podawać podpróbę, dla której liczymy w opcji prefiksu **svy**, wedle składni **svy, subpop(if podproba==1)**. Tylko wtedy uzyskamy prawidłowe oceny błędów standardowych.

B. Analiza wariancji

Analizę wariancji z uwzględnieniem schematu losowania (tak, aby wnioskować o parametrach dla całej populacji) można przeprowadzić na dwa sposoby. Klasyczny model ANOVA można uzyskać, przeprowadzając analizę regresji liniowej podobnie jak powyżej[13]. W ten sposób można także

[13] W programie Stata standardowe komendy do analizy ANOVA mają nazwę: **anova**, **oneway**, **loneway**. Pozwalają one na uzyskanie wszelkiego rodzaju statystyk, porównań między grupami, testów rzetelności itp. Jednak wszystkie dopuszczają jedynie wagi analityczne, a nie prawdopodobieństwa. Komendy te nie biorą też pod uwagę pełnego schematu losowania, nie powinno się więc ich stosować z opisywanym tu danymi.

przeprowadzić analizę ANCOVA, gdzie obok zmiennych kategorialnych włączamy do analizy zmienną ciągłą. W istocie model regresji jest uogólnieniem modelu ANOVA i tak może być stosowany. W programie Stata dodatkową korzyścią z korzystania z modelu regresji liniowej jest możliwość uwzględnienia schematu losowania, czego nie da się zrobić w standardowej procedurze ANOVA. Innym sposobem analizy jest model ANOVA z efektami losowymi („mixed ANOVA"). Jest to właściwy sposób analizy dla danych edukacyjnych, pochodzących z losowej próby szkół i mających strukturę hierarchiczną. W ten sposób uzyskujemy prawidłowe oceny wariancji w populacji. Przykłady analizy metodą regresji oraz metodą regresji z efektami losowymi przedstawiono poniżej. W obu przypadkach interesuje nas, jaki procent wariancji wyjaśniany jest przez przynależność ucznia do szkoły.

Zacznijmy od modelu ANOVA dla całej populacji. Nie musimy tu oczywiście brać pod uwagę schematu losowania, stąd możemy skorzystać ze standardowej komendy do analizy ANOVA z dużą liczbą grup, czyli **loneway**. Wyniki oszacowań zaprezentowano poniżej.

```
. loneway hum szkola

      One-way Analysis of Variance for hum: cz. humanistyczna

                                             Number of obs =    423460
                                                 R-squared =    0.1132

    Source                SS         df      MS            F     Prob > F
-------------------------------------------------------------------------
Between szkola        4133829.9   5035    821.01884     10.61     0.0000
Within szkola          32391836 418424    77.413906
-----------------------------------------------------------
Total                  36525666 423459    86.255496

         Intraclass       Asy.
         correlation      S.E.       [95% Conf. Interval]
         -----------------------------------------------
            0.10253     0.00236       0.09791     0.10715

         Estimated SD of szkola effect          2.973897
         Estimated SD within szkola             8.798517
         Est. reliability of a szkola mean       0.90571
              (evaluated at n=84.08)
```

Tabela zawiera typowe statystyki i testy dla analizy ANOVA, a także tzw. „intraclass correlation", który można interpretować jako procent wariancji wyjaśniony przez przynależność do szkoły. Dla pełnej próby 423 460 uczniów i 5 036 gimnazjów odchylenie standardowe efektów szkół wynosi 2,973897. Podnosząc je do kwadratu i dzieląc przez wariancję wyników z egzaminu w części humanistycznej uzyskujemy 2,973897^2/9,287384^2 = 0,10253. Taką samą liczbę podaje program w tabeli wyników, dodając też błąd standardowy i przedział ufności. Podsumowując, w całej populacji przynależność ucznia do szkoły wyjaśnia 10% wyniku ucznia. Podobną informację uzyskujemy ze statystyki R-kwadrat, równej 0,1132.

Dla pełnej populacji oszacowaliśmy podobne statystyki posługując się modelem z efektami losowymi. Do tego celu można wykorzystać znaną nam już procedurę **xtmixed**. Składnię i wyniki podano poniżej.

```
. xtmixed hum || szkola:

      Mixed-effects REML regression                  Number of obs      =    423460
      Group variable: szkola                         Number of groups   =      5036

                                                     Obs per group: min =        15
                                                                    avg =      84.1
                                                                    max =       359

      ------------------------------------------------------------------------------
               hum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
      -------------+----------------------------------------------------------------
             _cons |   31.87524   .0454358   701.54   0.000     31.78619     31.9643
      ------------------------------------------------------------------------------

      ------------------------------------------------------------------------------
        Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
      -----------------------------+------------------------------------------------
      szkola: Identity             |
                        sd(_cons)  |   3.011483   .0344088      2.944793    3.079684
      -----------------------------+------------------------------------------------
                    sd(Residual)   |   8.798731   .0096183        8.7799    8.817603
      ------------------------------------------------------------------------------
```

Współczynnik „intraclass correlation" możemy obliczyć, dzieląc wariancję na poziomie szkół, którą uzyskujemy, podnosząc do kwadratu odchylenie standardowe efektów szkół (podane jako **sd(_cons)**), przez sumę wariancji na poziomie szkół oraz uczniów (odchylenie standardowe na poziomie uczniów podano jako **sd(Residual)**). Interesujący nas współczynnik wynosi więc 3,011483^2/(3,011483^2+8,798731^2) = 0,105. Wynik jest niemal identyczny z uzyskanym klasycznym modelem ANOVA.

Przejdźmy do analizy bazy danych. Tutaj powinniśmy uwzględnić schemat losowania. Zacznijmy od regresji odpowiadającej modelowi ANOVA, lecz uwzględniającej pełny schemat losowania i wagi probabilistyczne. Do regresji wprowadzamy zestaw zerojedynkowych zmiennych, przyjmujących jeden dla każdej szkoły. Jest to równoważne modelowi ANOVA z efektami stałymi. W programie Stata można do tego celu wykorzystać prefiks **xi**:, którego zastosowanie spowoduje, że każda zmienna poprzedzona znakiem **i**. zostanie automatycznie przełożona na zestaw zmiennych zerojedynkowych i tak wprowadzona do modelu. Ponownie prefiks **svy**: powołuje się na wcześniej zdefiniowany schemat losowania komendą **svyset**. Dla przejrzystości pominięto większość ocen efektów pojedynczych szkół.

```
. xi: svy: reg hum i.szkola

      Survey: Linear regression

      Number of strata   =       100                    Number of obs      =       9054
      Number of PSUs     =       200                    Population size    = 423229.47
                                                        R-squared          =     0.1322
      ------------------------------------------------------------------------------
                   |             Linearized
               hum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
      -------------+----------------------------------------------------------------
        _Iszkola_2 |  -3.704343      .13731   -26.98   0.000    -3.976762   -3.431924
        _Iszkola_3 |   5.696939    .2423771    23.50   0.000      5.21607    6.177808

      (……)

      _Iszkola_199 |   .9323046    .1625939     5.73   0.000      .609723    1.254886
      _Iszkola_200 |   3.346939    .0456068    73.39   0.000     3.256456    3.437421
             _cons |   29.65306    .0456068   650.19   0.000     29.56258    29.74354
      ------------------------------------------------------------------------------
```

Statystyka R-kwadrat sugeruje, że efekty szkół wyjaśniają 13% wariancji wyników w części humanistycznej. Jest to wynik bliski uzyskanemu dla całej populacji.

Model ANOVA z efektami losowymi uzyskujemy ponownie, wykorzystując znaną nam już procedurę **xtmixed**.

```
. xtmixed hum || szkola:

      Mixed-effects REML regression                     Number of obs      =       9054
      Group variable: szkola                            Number of groups   =        200
                                                        Obs per group: min =         21
                                                                       avg =       45.3
                                                                       max =         60

      ------------------------------------------------------------------------------
               hum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
      -------------+----------------------------------------------------------------
             _cons |    31.8888    .2399921   132.87   0.000     31.41842    32.35917
      ------------------------------------------------------------------------------

      ------------------------------------------------------------------------------
        Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
      -----------------------------+------------------------------------------------
      szkola: Identity             |
                         sd(_cons) |   3.130655   .1840738      2.789889    3.513043
      -----------------------------+------------------------------------------------
                      sd(Residual) |   8.710995   .0654578       8.58364     8.84024
      ------------------------------------------------------------------------------
```

Te oceny można porównać z ocenami dla całej populacji, gdzie odchylenie standardowe efektów szkół wynosi 3,0, a odchylenie standardowe na poziomie ucznia 8,8. Widać więc, że oceny z próby są bliskie tym w całej populacji. Procedura **xtmixed** nie pozwala jednak zastosować wag probabilistycznych. Stąd ponownie musimy odwołać się do procedury **gllamm**. Skorzystamy jedynie z końcowych wag na poziomie ucznia. Procedura **gllamm** standardowo podaje wariancję efektów losowych. Czytelnicy mogą sprawdzić, że uzyskane oceny wariancji są bliskie tym w całej populacji. Współczynnik „intraclass correlation" wynosi 0,11 i jest niemal identyczny jak współczynnik oszacowany dla całej populacji (0,10). Udostępniana baza danych pozwala więc szacować wariancję wyników w populacji, także w podgrupach uczniów. Trzeba jednak pamiętać, że próba została wylosowana tak, aby zapewnić reprezentatywność na poziomie uczniów i wszelkie analizy na poziomie szkół należy traktować z większą ostrożnością.

```
. gen pw1=pw_ucznia
. gllamm hum, i(szkola) pw(pw) adapt nip(16)

      gllamm model

      Robust standard errors
      ------------------------------------------------------------------------------
               hum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
      -------------+----------------------------------------------------------------
             _cons |   31.65512   .1863572   169.86   0.000     31.28986    32.02037
      ------------------------------------------------------------------------------

      Variance at level 1
      ------------------------------------------------------------------------------
        73.318621 (1.6129868)

      Variances and covariances of random effects
      ------------------------------------------------------------------------------
      ***level 2 (szkola)

          var(1): 9.5713017 (.42814353)
```

C. Przykład analizy metodą EWD: ewaluacja efektywności wydatków na gimnazja

Poniżej prezentujemy przykład analizy metodą edukacyjnej wartości dodanej przy wykorzystaniu danych dla kilku roczników. Przykłady te opierają się na badaniach opisanych dokładniej w innych pracach, gdzie wykorzystano dane z całej populacji (Jakubowski, 2007, 2008b). W szczególności poniższy tekst stanowi modyfikację dyskusji zawartej w: Jakubowski, 2008d, gdzie znaleźć można też inne przykłady zastosowania metod EWD do oceny polityki edukacyjnej. Poniżej przedstawiono podobne analizy, jednak w oparciu o omawianą bazę danych.

Wykorzystanie na poziomie indywidualnym danych o wynikach egzaminu kończącego szkołę podstawową powoduje, że modele EWD należy interpretować jako wyjaśniające nie tyle **poziom** wiedzy, co jej **przyrost** w trakcie trzyletniej nauki uczniów w gimnazjum. Stanowi to główną zaletę modeli EWD, jaką jest, przynajmniej częściowe, wyłączenie wpływu na poziom osiągnięć czynników niezależnych od szkoły, także niemierzalnych, takich jak wrodzone zdolności (por. Raudenbush, Bryk, 2002; Goldstein, 1999).

Zaproponowane podejście opiera się na dominującej na świecie metodzie EWD, jeśli chodzi o ocenę efektywności pracy szkół lub nauczycieli. Jest to hierarchiczny model liniowy (inne stosowane nazwy to: model mieszany, z efektami losowymi, wielopoziomowy, z losową stałą lub współczynnikiem), gdzie dane indywidualne uczniów (w tym wyniki sprawdzianu) stanowią poziom pierwszy analizy i są podstawą do oszacowania efektów losowych klas (poziom drugi) oraz szkół (poziom trzeci). Podejście to zapobiega błędom związanym z agregacją danych lub pominięciem powiązania obserwacji wewnątrz grup. Omówienie, przykłady zastosowań i różnice w interpretacji wyników dla innych modeli, przede wszystkim modelu z efektami stałymi, omówiono w przywoływanej już pracy: Jakubowski, 2008d.

Postać ogólną trzypoziomowego modelu EWD można zapisać poniższym równaniem:

$$y_{iks} = \pi_0 + \mathbf{X_{iks}}\boldsymbol{\beta} + u_s + v_{ks} + \varepsilon_{iks}$$

gdzie y_{iks}, to wynik *i*-tego ucznia w klasie *k* i szkole *s*, a wektor $\mathbf{X}_{\mathbf{iks}}$ zawiera interesujące nas charakterystyki ucznia, klasy i szkoły. W naszym przypadku *y* oznacza wynik egzaminu gimnazjalnego w części humanistycznej przeniesiony na skalę wyników z 2007 roku. Wektor X zawiera wynik sprawdzianu danego ucznia na skali z 2004 roku (także kwadrat sprawdzianu, umożliwiając dopasowanie nieliniowej zależności między wynikami egzaminów). Inne uwzględnione w modelach wydatków na gimnazja zmienne można odczytać z tabeli zawierającej wyniki regresji (patrz poniżej). Część losowa równania zawiera reszty na poziomie ucznia ε_{iks} oraz efekt losowy klasy v_{ks} i szkoły u_s. Przyjmujemy typowe założenia o niezależności efektów losowych oraz o normalności i skończonej wariancji ich rozkładów $\varepsilon_{iks} \sim N(0,\sigma_\varepsilon^2)$, $v_{ks} \sim N(0,\sigma_v^2)$, $u_s \sim N(0,\sigma_u^2)$.

Model opisany powyższym równaniem nazywany jest modelem wielopoziomowym z losowymi stałymi, w tym przypadku na poziomie szkoły i klasy. Inną grupę stanowią modele EWD z losowym nachyleniem wyników egzaminu na niższym poziomie, w naszym przypadku z losowym nachyleniem współczynnika określającego relację między wynikami sprawdzianu i egzaminu gimnazjalnego. O ile szkoły różnią się pod względem tej relacji, o tyle model z losowym nachyleniem pozwala określić odrębne efekty dla uczniów słabszych oraz zdolniejszych. Bardziej „płaskie" nachylenie sprawdzianu oznacza większą zdolność szkoły do wyrównywania szans edukacyjnych: szkoła ta radzi sobie lepiej z uczniami słabszymi niż podobne szkoły a gorzej z uczniami lepszymi. Nie oznacza to jednak, że taka szkoła nie może być ogólnie szkołą bardziej (lub mniej) efektywną niż inne gimnazja, a jedynie to, że jej efektywność jest zróżnicowana i zależy od poziomu uczniów. Aby analizować tego typu zależności, zakłada się niezerową kowariancję między losową stałą oraz nachyleniem. W praktyce oznacza to, że dopuszczamy możliwość np. negatywnej korelacji, a więc tego, że szkoły z wyższymi wynikami uczniów cechuje także słabsza (bardziej płaska) relacja między wynikami na wejściu i wyjściu, a więc cechuje je też wyższa zdolność do wyrównywania wyników uczniów. Są to jednak tylko przykłady, które mogą zostać potwierdzone lub obalone w konfrontacji z rzeczywistymi danymi. W rzeczywistości takie relacje mogą nie zachodzić lub też mogą być odwrotne (np. szkoły z wyższą przeciętną EWD są równocześnie szkołami zwiększającymi zróżnicowanie wyników uczniów, a więc sprzyjającymi uczniom zdolniejszym). Współczynnik nachylenia można próbować objaśniać innymi zmiennymi, testując hipotezy o ich „wyrównawczym" wpływie. W naszym przypadku będziemy badać efekt interakcji między wydatkami a nachyleniem wyników sprawdzianu, a więc testować hipotezę o wyrównującym wpływie wydatków na osiągnięcia uczniów.

Przy szacowaniu modeli z losowymi współczynnikami celowe jest „wycentrowanie" zmiennej objaśniającej tak, aby uzyskane oceny parametrów mogły być interpretowane w interesujący z praktycznego punktu widzenia sposób. W naszym przypadku oryginalna zmienna zawierająca wyniki sprawdzianu ma wartość minimalną 0, przez co oszacowane efekty nie mają bezpośredniej interpretacji. Stąd w modelu z losowym nachyleniem wyniki sprawdzianu zostały przekształcone tak, aby wartość 0 odpowiadała dolnemu decylowi (10 percentylowi) wyników sprawdzianu uczniów danego gimnazjum (tzw. „group centering"). O ile wycentrowanie wokół średniej dla próby (tzw. „grand mean centering") daje przy pewnych założeniach model ekwiwalentny modelowi bez przekształceń zmiennej objaśniającej, o tyle już centrowanie wokół statystyki dla grupy zmienia interpretację modelu (de Leeuw, 2005). Model ten ma teraz ciekawą interpretację, mierząc wpływ interesujących nas czynników na sytuację uczniów o relatywnie niskim poziomie wiedzy na progu gimnazjum (w stosunku do poziomu wiedzy uczniów danej szkoły).

Model z losowym nachyleniem zapiszemy zestawem równań, osobno definiując poziom ucznia oraz stałą i nachylenie dla szkoły[14]:

poziom indywidualny: $y_{iks} = \pi_0 + \mathbf{D_{iks}}\boldsymbol{\beta} + \pi_1 spr^*_{iks} + \varepsilon_{iks}$

poziom szkoły: $\pi_0 = \gamma_{00} + \mathbf{G}_0\boldsymbol{\gamma}_{\mathbf{01}} + u_0$

$$\pi_1 = \gamma_{10} + \mathbf{G}_1\boldsymbol{\gamma}_{\mathbf{11}} + u_1$$

gdzie na poziomie indywidualnym y_{iks} oraz ε_{iks} zdefiniowano jak powyżej, D_{iks} to wektor zerojedynkowych charakterystyk ucznia (płeć, dysleksja, laureat), a spr^*_{iks} to przekształcony w opisany powyżej sposób wynik sprawdzianu (wycentrowany wokół dolnego decyla). Równania na poziomie szkoły objaśniają przeciętny wynik egzaminu (parametr π_0) oraz nachylenie indywidualnych wyników sprawdzianu (parametr π_1) za pomocą zestawu zmiennych na poziomie szkoły (zawartych w wektorach $\mathbf{G_0}$ oraz $\mathbf{G_1}$), w tym wydatków na ucznia podczas jego nauki w gimnazjum[15]. Przez u_0 oraz u_1 oznaczono efekty losowe na poziomie szkoły. Zakładamy, że $\varepsilon_{isg} \sim N(0,\sigma^2_\varepsilon)$, $u_0 \sim N(0,\sigma^2_{u0})$, , $u_1 \sim N(0,\sigma^2_{u1})$, oraz $\mathrm{cov}(u_0,u_1) \neq 0$.

Wszystkie powyższe modele oszacowano metodą największej wiarygodności za pomocą procedury **xtmixed** w programie Stata (por. Rabe-Hesketh, Skrondal, 2008) łącznie dla trzech roczników uczniów. Przez to wariancja wykorzystana do oszacowania interesujących nas efektów dotyczy zarówno zróżnicowania wewnątrz szkół na przestrzeni lat, jak i zróżnicowania między szkołami w każdym roku. Tak więc model ten korzysta zarówno z wariancji wyników w każdej szkole i każdym roku („within-variance"), jak i między szkołami („between-variance"). W modelu nie wykorzystano wag odzwierciedlających prawdopodobieństwo wylosowania szkoły i ucznia. Docelowe analizy powinny zostać przeprowadzone z ważeniem obserwacji.

Poniżej przedstawiono rezultaty uzyskane modelami EWD opisanymi powyżej. Naszym celem jest zaprezentowanie możliwości ewaluacji metodą EWD, a nie dyskusja dotycząca polityki edukacyjnej, przez co w opisie wyników skupimy się na ich poprawnym odczytaniu, a nie na interpretacji. Zainteresowanych czytelników odsyłamy do oryginalnych artykułów, gdzie przeprowadzono szerszą dyskusję w kontekście efektów decentralizacji w Polsce (por. Jakubowski, 2007, 2008b).

Zacznijmy od modelu wydatków z losowym efektem szkół i klas. Poniżej przedstawiono kod oraz wyniki estymacji procedurą **xtmixed**. Wykorzystano kilka zmiennych, których w bazie nie ma, lecz łatwo je utworzyć. Uwzględniono kwadrat wyników sprawdzianu w skali z 2004 roku, aby dopasować nieliniową zależność między wynikami sprawdzianu i egzaminu gimnazjalnego. Obliczono także odchylenie standardowe wyników sprawdzianu uczniów każdej klasy (zmienna *sd_spr*), które stanowi wskaźnik zróżnicowania poziomu umiejętności uczniów w klasie przed rozpoczęciem nauki w gimnazjum. W bazie co prawda istnieją podobne zmienne, jednak dotyczą one całego gimnazjum a nie klas. Podobnie obliczono średni wynik sprawdzianu uczniów w klasie (zmienna *sr_spr*), który mierzy przeciętny poziom uczniów na progu gimnazjum. Oprócz wyników egzaminów w modelu

14 Pominięto poziom klasy, aby ułatwić trudne obliczenia numeryczne. Oczywiście możliwe jest też modelowanie losowych efektów na poziomie klas, jednak w takim przypadku użytkownicy często napotkają problemy natury obliczeniowej i oszacowanie parametrów modelu może być niemożliwe.

15 W oryginalnej pracy model ten dotyczył poziomu ucznia oraz gminy, ponieważ dane o wydatkach na szkoły dostępne są na tym poziomie (por. Jakubowski, 2008b, 2008d). W bazie nie ma jednak zbyt wiele szkół z tej samej gminy, stąd poziom gminy i szkoły można uznać za tożsamy.

uwzględniono takie zmienne indywidualne, jak płeć ucznia (1 = dziewczynki), dysleksję w szkole podstawowej i na egzaminie gimnazjalnym (1 = uczeń pisał egzamin jako dyslektyk), status laureata olimpiady z przedmiotów humanistycznych (1 = laureat). Dodatkowo, dzięki prefiksowi **xi**: przed komendą **xtmixed**, podając jako zmienną *i.mw*, dodano zmienne zerojedynkowe oznaczające wielkość miejscowości, w której znajduje się szkoła (wg kategorii stosowanych przez CKE; wieś przyjęto jako punkt odniesienia dla 3 pozostałych kategorii). Jest to o tyle ważne, że gminy wiejskie i miejskie posiadają inną kwotę bazową na ucznia w algorytmie naliczania subwencji oświatowej. Jak już wspomniano, dodano średnią i odchylenie standardowe wyników sprawdzianu w klasie, ale także średni wynik sprawdzianu w szkole. Ostatnią zmienną jest wskaźnik przeciętnych nakładów na ucznia gimnazjum w trakcie jego 3-letniej nauki. Wskaźnik ten jest wyrażony w złotówkach z 2005 roku (po uwzględnieniu inflacji).

```
. gen kwadr=spr_2004^2
. egen sd_spr=sd(spr_2004), by(klasa)
. egen sr_spr=mean(spr_2004), by(klasa)

. xi: xtmixed hum_2007 spr_2004 kwadr plec dys_sp dys_gim laur_hum i.mw sr_gim_spr
sr_spr sd_spr naklady_gim || szkola: || klasa:

     Mixed-effects REML regression                    Number of obs      =     27261

     -------------------------------------------------------------
                     |   No. of       Observations per Group
      Group Variable |   Groups    Minimum    Average    Maximum
     ----------------+--------------------------------------------
              szkola |      200         63      136.3        174
               klasa |     1181         15       23.1         34
     -------------------------------------------------------------

     ---------------------------------------------------------------------------------
         hum_2007 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
     -------------+-------------------------------------------------------------------
```

Procedura **xtmixed** na początku podaje przydatną tabelę opisującą liczbę grup i obserwacji wewnątrz nich. Jak widać, mamy 200 szkół i 1 181 klas, co stanowi pełną bazę oddaną użytkownikom. Minimalna liczba obserwacji w klasach to 15, a maksymalna to 34 – co odpowiada przyjętym założeniom. Przeciętna wielkość oddziału to 23,1 ucznia, co jest bliskie średniej dla Polski. Z 200 szkół najmniej liczna posiada 63 obserwacje (zapewne po jednej klasie z 3 lat), a największa 174 (po 2 klasy z każdego roku). Średnia liczba obserwacji na szkołę to 134 uczniów. Tabela ta jest o tyle przydatna, że pozwala szybko ocenić liczbę obserwacji w grupach. Jeśli grupy są niewielkie lub jest ich mało, to najczęściej modelowanie efektów na danym poziomie, jako mających rozkład losowy bliski normalnemu, nie ma sensu. Wtedy należy albo pominąć dany poziom, albo też analizować dane za pomocą zwykłego modelu regresji liniowej, ew. z efektami stałymi. W naszym przypadku zarówno liczba grup, jak i obserwacji wewnątrz nich, jest zadowalająca.

Tabela podaje oszacowane współczynniki dla każdej zmiennej, ich błąd standardowy, statystykę z, która w tym przypadku zastępuje statystkę t podawaną w zwykłej regresji, „p-value", oraz przedział ufności dla współczynnika. Jak widać, wyniki sprawdzianu mają bardzo dużą moc objaśniającą wyników egzaminu gimnazjalnego w części humanistycznej, co potwierdza zasadność modelu EWD. Odpowiedni test na łączną istotność wyników sprawdzianu i ich kwadratu potwierdza, że z bardzo wysokim prawdopodobieństwem są one różne od zera (test ten wykonuje komenda **test spr_2004 kwadr** wykonana po opisanej powyżej komendzie **xtmixed**). Każdy dodatkowy punkt uzyskany przez ucznia na sprawdzianie podnosi przewidywany wynik egzaminu gimnazjalnego o ponad 1 punkt, choć dla uczniów z wyższymi wynikami efekt jest mniejszy ze względu na ujemny znak przy kwadracie. Efekty dla płci i dysleksji są bardzo podobne do uzyskanych w innych badaniach dla całej populacji. Zgodnie z intuicją, laureaci olimpiad uzyskują ponadprzeciętny przyrost umiejętności mierzony egzaminami[16]. Podobnie (zgodnie z naszą wcześniejszą wiedzą) uczniowie z terenów miejskich uzyskują wyższy przyrost wiedzy, choć istotne statystycznie i zauważalne z praktycznego punktu widzenia różnice występują jedynie między wsią a największymi miastami (współczynnik dla *_Imw_3* jest istotny statycznie i oznacza różnice w EWD między wsią a największymi miastami, dla których w bazie zmienna *mw* przyjmuje wartość 3).

Ciekawe oceny efektów uzyskano dla zmiennych skonstruowanych z wyników sprawdzianu dla szkoły oraz klasy. Średni wynik uczniów gimnazjum ze sprawdzianu negatywnie wpływa na przyrost wiedzy, jednak już średni wynik w klasie wpływa pozytywnie. Można postawić kilka hipotez wyjaśniających takie efekty, jednak wymagają one głębszej analizy, na którą nie ma tu miejsca. Przy interpretacji należy zachować tu wysoką ostrożność, bowiem średnie wyniki uczniów na poziomie szkół lub klas są z pewnością silnie skorelowane z kontekstem społecznym i ekonomicznym, w jakim działa szkoła. Co więcej, efekty te mogą być błędnie oszacowane ze względu na błąd pomiaru wyników sprawdzianu (por. Jakubowski, 2008c). Łatwiejsza byłaby interpretacja wpływu zróżnicowania uczniów w klasie na przyrost wiedzy mierzony modelem EWD. Jednak współczynnik przy zmiennej *sd_spr* jest nieistotny i nie może być interpretowany.

Ocena parametru dla kluczowej zmiennej określającej poziom wydatków w przeliczeniu na ucznia jest bardzo precyzyjna. Błąd standardowy wynosi 0,0000897, co pozwala wykryć nawet najmniejsze, istotne z praktycznego punktu widzenia, efekty. W istocie oszacowany współczynnik dla nakładów na gimnazja jest pozytywny i istotny statystycznie, jednak jego wartość jest bardzo niska. Efekt wydatków jest praktycznie bliski zeru. Wydatki mierzone są w tysiącach złotych, a więc efekt równy 0,0001989 oznacza, że wzrost wydatków o 1 000 PLN może powiększyć przyrost wiedzy uczniów w gimnazjum o niecałe 0,2 punktu. Jest to liczba niewielka, właściwie pomijalna, bowiem z pewnością zwiększenie wydatków o 5 000 PLN (więcej niż średnio wydają na naukę ucznia gminy) nie jest możliwe, a tylko takie nakłady mogłyby podnieść wyniki uczniów o 1 punkt. Mając na uwadze powyższe oszacowania, nie jest to z pewnością pomysł wart rozważenia. Wyniki pokazują zatem, że wydatki nie mają wpływu na rozwój wiedzy uczniów. W badaniach dla całej populacji uzyskano bardzo podobne wyniki (por. Jakubowski, 2007, 2008b). Nie znaczy to jednak, że zwiększenie lub zmniejszenie nakładów nigdy nie może wpłynąć na osiągnięcia uczniów. Znaczy to jedynie tyle, że przeciętnie w kraju wydatki gmin na gimnazja nie są związane z jakością nauczania (postępami uczniów), a więc przeciętnie wydatki są nieefektywne. Z pewnością istnieją pojedyncze gminy, które umiejętnie alokują fundusze, podnosząc jakość nauczania, jednak wyniki pokazują, że w skali kraju tak się nie dzieje.

Kolejne rezultaty dotyczą modelu EWD z losowym nachyleniem wyników sprawdzianu. Tym razem efektywność wydatków oceniana jest przez ich wpływ na przyrost wiedzy uczniów relatywnie słabych, z wynikiem ze sprawdzianu ok. 10 percentyla wyników w danej szkole. Nową zmienną w modelu jest przekształcony wynik sprawdzianu wycentrowany wokół wyników 10 percentyla uczniów danej szkoły (zmienna *x_spr*) oraz efekt interakcji między wynikami sprawdzianu a nakładami na gimnazja. Testujemy tu dwie hipotezy. Po pierwsze, czy wpływ wydatków jest widoczny dla uczniów relatywnie słabych. Po drugie, czy wydatki mają charakter „wyrównujący" osiągnięcia uczniów w gimnazjum. Pierwszą hipotezę testujemy poprzez ocenę współczynnika przy zmiennej nakłady. Jeśli jest ona istotna statystycznie, to oznacza, że wydatki mają wpływ na przyrost umiejętności uczniów względnie słabych. Drugą hipotezę testujemy poprzez ocenę współczynnika interakcji między wynikami sprawdzianu a nakładami. Jeśli ten współczynnik byłby ujemny, to oznaczałoby to, że im wyższe wydatki, tym mniej

[16] Można argumentować, że oceniając efektywność wydatków na gimnazja nie powinno się „wyłączać" wpływu na EWD wyników olimpijczyków. Omawiana analiza ma jednak charakter przykładowy, a liczba olimpijczyków jest tak znikoma, że na pewno nie wpływa to na uzyskane wyniki.

strome nachylenie wyników sprawdzianu. To mogłoby potwierdzać tezę, że wydatki wyrównują wyniki uczniów na egzaminie gimnazjalnym lub mówiąc inaczej – powodują, że zależność wyniku egzaminu gimnazjalnego od wyniku sprawdzianu jest tym słabsza, im wyższe nakłady na ucznia.

Komendy zaprezentowane poniżej tworzą zmienną zawierającą 10 percentyl wyników sprawdzianu w każdej szkole, który służy do stworzenia zmiennej zawierającej wycentrowane wokół 10 percentyla wyniki sprawdzianu. Nakłady na gimnazja podzielono przez 1 000, tak więc wyrażają one teraz efekt zmiany wydatków na ucznia o 1 000 PLN. Następnie stworzono zmienną opisującą efekt interakcji między nakładami a nachyleniem sprawdzianu. W komendzie **xtmixed**, po podaniu zmiennej identyfikującej szkoły z dwukropkiem, podano scentrowany wynik sprawdzianu, wskazując tym samym zmienną o losowym nachyleniu wewnątrz grup definiowanych przez zmienną *szkola*. Jako opcję podano **cov(unstr)**, co powoduje, że estymowana jest macierz wariancji-kowariancji między efektami losowymi, a przez to mogą być one ze sobą skorelowane. Jest to niezbędne, gdyż procedura **xtmixed** standardowo przyjmuje, że efekty losowe nie mogą być ze sobą skorelowane, co w tym przypadku nie ma uzasadnienia.

```
. egen p10=pctile(spr_2004), by(szkola) p(10)
. gen x_spr=spr_2004-p10
. replace naklady1000=naklady_gim/1000
. gen inter=x_spr*naklady_gim
```

```
. xi: xtmixed hum_2007 x_spr kwadr plec dys_sp dys_gim laur_hum i.mw sr_gim_spr sr_spr
naklady_gim inter || szkola: x_spr, cov(unstr)

     Mixed-effects REML regression                   Number of obs      =     27261
     Group variable: szkola                          Number of groups   =       200
                                                     Obs per group: min =        63
                                                                    avg =     136.3
                                                                    max =       174
     ------------------------------------------------------------------------------
         hum_2007 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
     -------------+----------------------------------------------------------------
            x_spr |   1.178946   .0353529    33.35   0.000     1.109656    1.248237
            kwadr |  -.0065069   .0005219   -12.47   0.000    -.0075297    -.005484
             plec |   2.516759    .072463    34.73   0.000     2.374734    2.658784
           dys_sp |  -2.315846   .1919365   -12.07   0.000    -2.692035   -1.939658
          dys_gim |   2.541267   .1681358    15.11   0.000     2.211727    2.870807
         laur_hum |     8.7711   1.972933     4.45   0.000     4.904223    12.63798
           _Imw_1 |    1.46436   .5469117     2.68   0.007     .3924326    2.536287
           _Imw_2 |    2.00859   .5716042     3.51   0.000     .8882664    3.128914
           _Imw_3 |   3.164718   .6624224     4.78   0.000     1.866394    4.463042
       sr_gim_spr |  -.3956694   .0378446   -10.46   0.000    -.4698434   -.3214955
           sr_spr |   .1750563   .0181333     9.65   0.000     .1395157     .210597
      naklady1000 |   .0006521   .0961528     0.01   0.995    -.1878038     .189108
            inter |   .0034642   .0053713     0.64   0.519    -.0070634    .0139919
            _cons |   28.44299   1.034103    27.50   0.000     26.41619     30.4698
     ------------------------------------------------------------------------------
       Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
     -----------------------------+------------------------------------------------
     szkola: Unstructured         |
                       sd(x_spr)  |    .072807   .0073986      .0596588    .0888528
                       sd(_cons)  |   4.593215   .2628604      4.105861    5.138417
              corr(x_spr,_cons)   |  -.6369134    .071318     -.7565719   -.4759827
     -----------------------------+------------------------------------------------
                   sd(Residual)   |   5.857637   .0253005      5.808258    5.907435
     ------------------------------------------------------------------------------
```

Odpowiednie testy potwierdzają, że gimnazja różnią się pod względem relacji między wynikami sprawdzianu a wynikami części humanistycznej egzaminu gimnazjalnego[17]. W dolnej części tabeli podano oszacowania odchyleń standardowych nachylenia sprawdzianu **sd(x_spr)** oraz efektów szkół. Podano także ocenę korelacji między nachyleniem sprawdzianu w szkole a stałą dla szkoły. Jest ona ujemna, co oznacza, że przeciętnie wyższe nachylenie wyników sprawdzianu w szkole związane jest z niższym przyrostem wiedzy dla grupy uczniów słabych, co jest zgodne z intuicją. Oceny parametrów dla zmiennych *naklady_gim1000* oraz *inter* są nieistotne statystycznie. Pokazuje to, że wydatki na gimnazja nie mają wpływu na osiągnięcia słabszych uczniów, ale także nie przyczyniają się do zmniejszania zróżnicowania wyników w części humanistycznej. Czytelnik może te wyniki porównać z oszacowaniami dla całej populacji i nieco innego zestawu zmiennych, przedstawionymi w: Jakubowski, 2008b, 2008d. Wnioski z obydwu analiz są podobne.

[17] Metodą największej wiarygodności oszacowano dwa identyczne modele różniące się jedynie wprowadzeniem w jednym z nich losowego efektu nachylenia sprawdzianu w każdej szkole. Następnie przeprowadzono „likelihood ratio-test", który pozwolił odrzucić hipotezę, że dwa modele mają podobną moc objaśniającą. W programie Stata testy tego typu można wykonać bardzo prosto, zapamiętując wyniki estymacji procedurą **estimates store**, a następnie podstawiając je do komendy **lrtest**.

Literatura

- Dolata R. (2007). Edukacyjna wartość dodana jako metoda oceny efektywności nauczania na podstawie egzaminów zewnętrznych. Centralna Komisja Egzaminacyjna.
- Goldstein H. (1999). Multilevel Statistical Models (poprawione wydanie II dostępne na: www.ats.ucla.edu/stat/examples/msm_goldstein/goldstein.pdf).
- Jakubowski M. (2007). Czy wydatki na gimnazja są efektywne? Gospodarka Narodowa 11/12/2007.
- Jakubowski M. (2008a). Implementing value-added models of school assessment. European University Institute RSCAS 2008/06 working paper.
- Jakubowski M. (2008b). Decentralization and teaching quality (wersja robocza dostępna na: www.wne.uw.edu.pl/mjakubowski).
- Jakubowski M. (2008c). Wpływ rzetelności pomiaru umiejętności uczniów na edukacyjną wartość dodaną. Opole: Opublikowane w materiałach pokonferencyjnych XIV Krajowej Konferencji Diagnostyki Edukacyjnej.
- Jakubowski M. (2008d). Zastosowanie modeli EWD do ewaluacji polityki edukacyjnej. Opole: Opublikowane w materiałach pokonferencyjnych XIV Krajowej Konferencji Diagnostyki Edukacyjnej.
- de Leeuw J. (2005). Centering in Multilevel Models. W: Everitt B., Howell D. (red.). Encyclopedia of Statistics in Behavioral Science. Wiley & Sons.
- McCaffrey D., Koretz D., Lockwood J. R., Hamilton L. (2004). Evaluating Value-Added Models for Teacher Accountability. RAND Corporation. MG-158.
- OECD (2008). Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools. Paris: OECD.
- Rabe-Hesketh S., Skrondal A. (2008). Multilevel and Longitudinal Modeling using Stata. Wydanie II, College Station. TX: Stata Press.
- Raudenbush S.W., Bryk A. (2002). Hierarchical Linear Models. Wydanie II, Sage Publications.