# Analiza danych z międzynarodowych badań edukacyjnych w IEA IDB Analyzer

Mateusz Kleczaj (m.kleczaj@ibe.edu.pl)

## Spis treści

# 1 Czym jest IDB Analyzer?

IDB Analyzer to bezpłatne narzędzie opracowane przez Międzynarodowe Stowarzyszenie Mierzenia Osiągnięć Szkolnych, IEA, służące do pracy z danymi z międzynarodowych badań edukacyjnych organizowanych przez IEA i OECD (Organizację Współpracy Gospodarczej i Rozwoju). Oferuje intuicyjny interfejs, który ułatwia pracę z danymi bez konieczności zaawansowanej znajomości programowania.

**Główne funkcjonalności narzędzia obejmują:**

- **Generowanie kodu analitycznego**: Automatycznie tworzy skrypty w językach SPSS, SAS lub R, uwzględniające metodologię badań IEA i OECD, w tym wagi statystyczne, losowy dobór próby oraz wartości prawdopodobne (*plausible values*, PVs)
- **Integracja danych**: Umożliwia łączenie różnych typów plików danych (np. dotyczących uczniów, nauczycieli, szkół czy rodziców) z wielu krajów w jeden spójny zbiór danych
- **Przygotowanie danych do analizy**: Pozwala na wybór określonych zmiennych i tworzenie dostosowanych zestawów danych, gotowych do dalszego przetwarzania statystycznego
- **Analizy statystyczne**: Umożliwia obliczanie statystyk takich jak: średnie, procenty, percentyle, korelacje, współczynniki regresji oraz odsetki uczniów osiągających określone poziomy umiejętności (benchmarki)
- **Konwersja formatów danych**: Umożliwia przekształcanie baz danych z formatu `.sav` (SPSS) do formatu `.RData` (R), co ułatwia pracę w środowisku R

**!** Ważne informacje o działaniu programu

IEA IDB Analyzer nie wykonuje analiz samodzielnie – generuje gotowy kod, który należy uruchomić w wybranym środowisku statystycznym (SPSS, SAS lub R). Wersja 5.0 IEA IDB Analyzer wymaga oprogramowania R (w wersji 4.2.0 lub nowszej), SPSS lub SAS.

Dodatkowe zasoby

- Narzędzie jest dostępne do pobrania ze strony IEA Data and Tools
- Szczegółowa dokumentacja, w tym instrukcja obsługi, znajduje się w interfejsie aplikacji pod przyciskiem „Help”

## 1.1 Obsługiwane badania

IEA IDB Analyzer obsługuje dane z szeregu międzynarodowych badań edukacyjnych realizowanych przez IEA, OECD, UNESCO oraz inne organizacje. W zależności od projektu możliwe jest zarówno łączenie plików danych (*merge*), jak i prowadzenie analiz statystycznych (*analysis*). Obsługiwane są trzy środowiska: SPSS, SAS oraz R.

| **Badanie / Organizacja** | **Typ obsługi** | **SPSS** | **SAS** | **R** |
|---|---|---|---|---|
| TIMSS / PIRLS / ICILS / ICCS (IEA) | Łączenie baz i analiza | • | • | • |
| TALIS / PIAAC (OECD) | Łączenie baz i analiza | • | • | • |
| PISA (OECD) | Analiza | • | • | • |

## 1.2 Wymagania systemowe, instalacja i uruchamianie programu

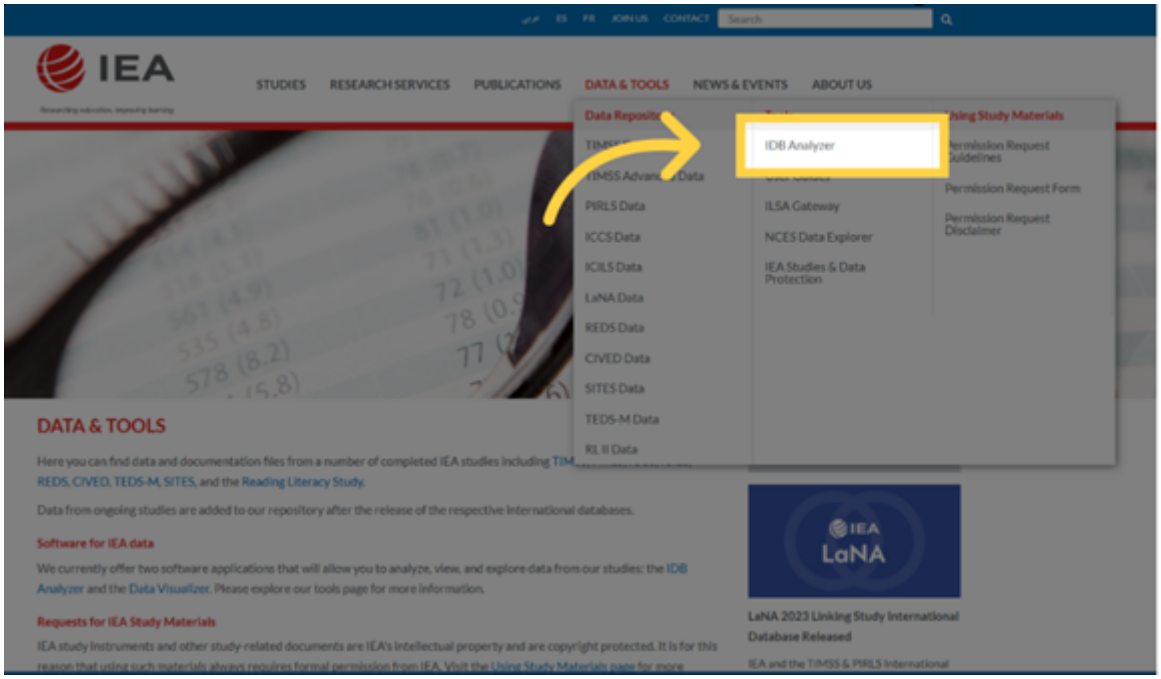
**i** Wymagania systemowe

Korzystanie z programu wymaga:

- **Systemu operacyjnego**: Windows 10 lub 11
  - Użytkownicy macOS mogą korzystać z programu wyłącznie przez maszynę wirtualną z zainstalowanym systemem Windows (np. Parallels Desktop lub VirtualBox)
- **Zainstalowanego wybranego środowiska analitycznego**:
  - IBM SPSS Statistics
  - SAS
  - lub R (w wersji 4.2.2 lub nowszej) wraz z RStudio
- **Dostępu do baz danych**: pobrane na dysk w formacie `.sav`, `.sas7bdat` lub `.RData` z Repozytorium Danych i Narzędzi IEA lub strony OECD

### 1.2.1 Instalacja programu

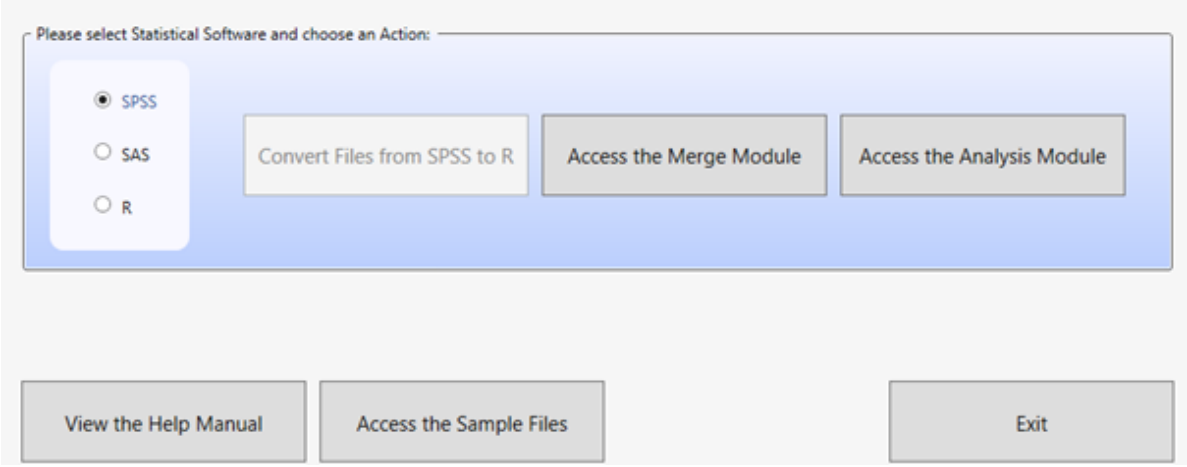
1. Pobierz najnowszą wersję instalatora ze strony: https://www.iea.nl/data-tools/tools




2. Zapisz plik instalacyjny w dowolnym folderze i uruchom instalator jako administrator
3. Przejdź przez proces instalacji, pozostawiając domyślne ustawienia
4. Po zakończeniu instalacji uruchom program: Menu Start → IEA → IEA IDB Analyzer

### 1.2.2 Uruchamianie aplikacji i dostępne moduły

Po uruchomieniu programu wyświetli się ekran główny z wyborem środowiska statystycznego: SPSS, SAS lub R.




W zależności od wybranego środowiska dostępne są trzy moduły:

- **Convert Module** (moduł konwersji) - dostępny po zaznaczeniu środowiska R, umożliwia konwersję plików `.sav` do formatu `.RData`
- **Merge Module** (moduł łączenia) - służy do łączenia danych z różnych plików odpowiadających zazwyczaj różnym narzędziom/bazom danych z danego badania (np. uczniowie, nauczyciele, szkoły) lub baz z kilku krajów w ramach danego badania
- **Analysis Module** (moduł analizy) - pozwala na przygotowanie skryptów do analizy danych (w SPSS, SAS lub R)

> Zalecana kolejność pracy
>
> Zwykle pracę rozpoczyna się od **Merge Module**, a następnie przechodzi do **Analysis Module**. W przypadku pracy w R, może być konieczne wcześniejsze użycie **Convert Module** w celu przekonwertowania plików danych na format obsługiwany przez RStudio.

Dodatkowo na ekranie startowym dostępne są przyciski:

- **View the Help Manual** - otwiera instrukcję obsługi
- **Access the Sample Files** - umożliwia pobranie przykładowych danych
- **Exit** - zamyka program

# 2 Moduły programu

Program IEA IDB Analyzer składa się z trzech modułów, które wspólnie umożliwiają pełny cykl przygotowania i analizy danych z międzynarodowych badań edukacyjnych: od konwersji plików, przez ich łączenie, aż po generowanie kodu analitycznego.

## 2.1 Moduł konwersji (Convert Module)

Convert Module służy do konwersji plików danych w formacie `.sav` (SPSS) do formatu `.RData`, który jest wymagany do dalszej analizy w środowisku R. Jest to pierwszy krok, jeśli planujesz przeprowadzać analizy w R a bazy danych, które posiadasz są w formacie `.sav`.

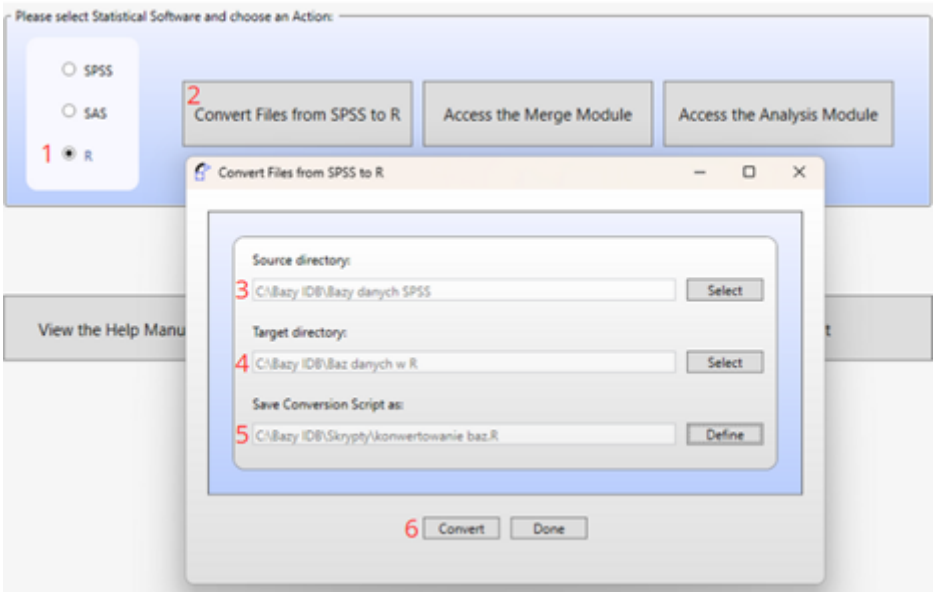
**Funkcje modułu:**

- Automatyczne wykrywanie wszystkich plików `.sav` we wskazanym folderze źródłowym
- Generowanie skryptu do konwersji danych SPSS do R
- Konwersja zmiennych z zachowaniem etykiet nazw i wartości
- Zapisanie plików wynikowych w folderze docelowym oraz otwarcie gotowego skryptu w RStudio lub innym edytorze

⚠ Wymagania i zalecenia

- Wszystkie zmienne muszą mieć przypisane etykiety (label), w przeciwnym razie mogą wystąpić błędy
- Zalecane jest trzymanie plików źródłowych i wynikowych w osobnych folderach
- Konieczne jest zainstalowanie R oraz RStudio

**Kroki użytkowania:**




1. Uruchom program i wybierz środowisko „R”
2. Kliknij przycisk „Convert Files from SPSS to R”
3. Wskaż folder źródłowy z plikami `.sav`
4. Wskaż folder docelowy, gdzie mają trafić pliki `.RData`
5. Zdefiniuj nazwę skryptu konwersji
6. Naciśnij „Convert” i uruchom wygenerowany skrypt w RStudio

## 2.2 Moduł łączenia (Merge Module)

Merge Module służy do łączenia danych z wielu krajów oraz różnych poziomów i źródeł (np. uczniowie, nauczyciele, szkoły), dzięki czemu powstaje jeden spójny zbiór gotowy do analizy. Program automatycznie generuje kod w wybranym języku (SPSS, SAS lub R), który umożliwia utworzenie zintegrowanego pliku danych zgodnie z konfiguracją użytkownika.

**Główne funkcje modułu:**

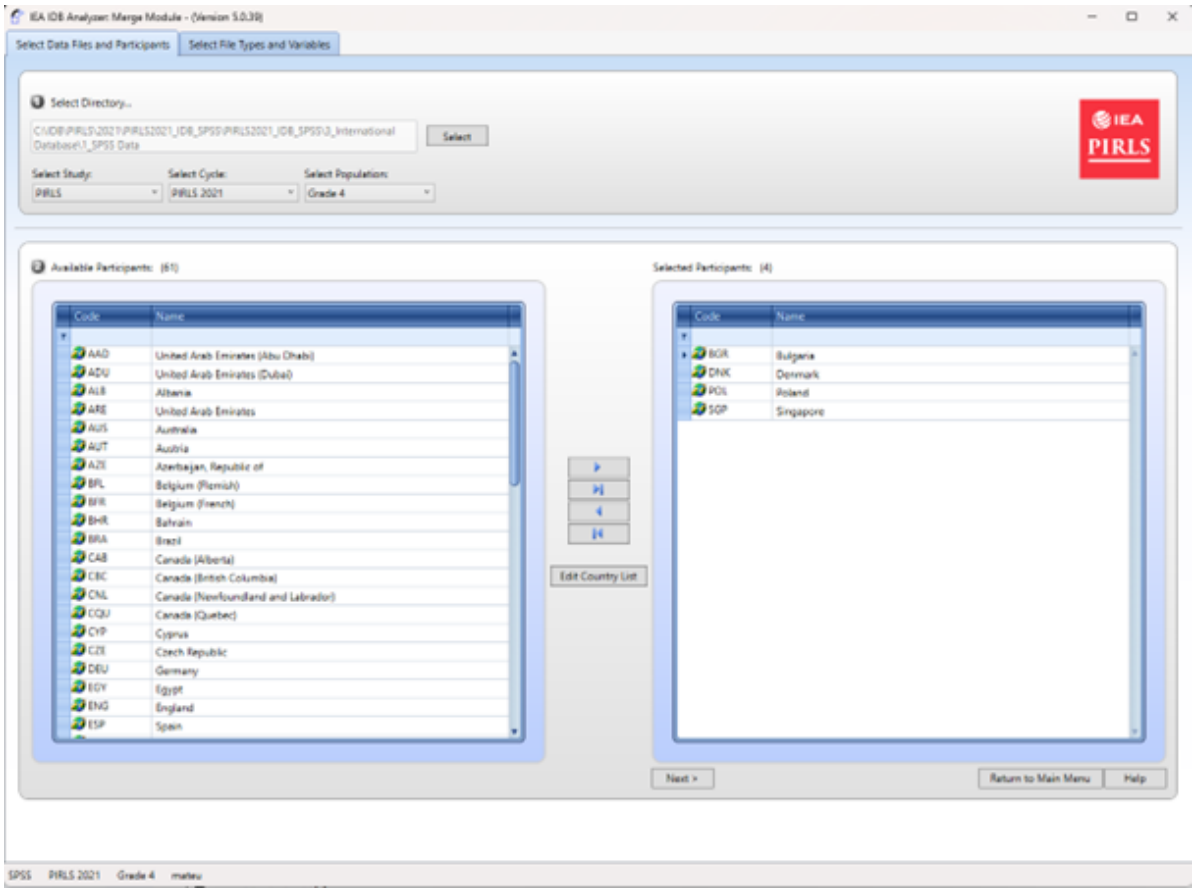
- Łączenie danych z wielu krajów biorących udział w badaniu
- Integracja plików pochodzących od różnych typów respondentów (np. uczniowie, szkoły, nauczyciele)
- Wybór zmiennych do analizy - możliwość ograniczenia liczby zmiennych w tworzonym zbiorze
- Edycja listy krajów (zmiana etykiet, dodanie lub usunięcie pozycji)

⚠ Ograniczenia i uwagi

- Moduł obsługuje tylko jedną edycję badania jednocześnie (np. PIRLS 2016). Aby połączyć dane z różnych cykli (np. PIRLS 2011 i 2016), należy wcześniej przygotować zbiory poza programem (często instrukcje, jak to zrobić, są w raportach technicznych danych badań)
- W środowisku R Merge Module wykorzystuje funkcję `left_join()` z pakietu `dplyr`. Aby zachować zgodność, zaleca się użycie tej samej metody przy łączeniu danych poza programem
- Program rozpoznaje pliki na podstawie ich nazw, zgodnych z ustalonymi wzorcami. Dlatego nie należy zapisywać pliku wynikowego w tym samym folderze, co pliki źródłowe - może to prowadzić do błędów
- Pliki muszą zawierać etykiety zmiennych i wartości (szczególnie istotne w środowisku R), a sama struktura plików musi być spójna między krajami

**Kroki użytkowania:**

1. **Uruchom program** i wybierz środowisko statystyczne (SPSS, SAS lub R). Przejdź do zakładki Merge Module
2. **Wybierz folder z danymi**: Wskaż folder zawierający pobrane wcześniej pliki danych z wybranego badania. Wszystkie pliki muszą znajdować się w tym samym folderze. Program automatycznie rozpozna badanie, cykl i populację, wyświetlając listę dostępnych krajów
3. **Wybierz kraje do analizy**: Z listy Available Participants wybierz kraje, klikając je i przenosząc do panelu Selected Participants za pomocą strzałki (→) lub podwójnego kliknięcia. Aby wybrać wiele krajów, przytrzymaj klawisz Ctrl. Użyj strzałki (→|) do zaznaczenia wszystkich krajów



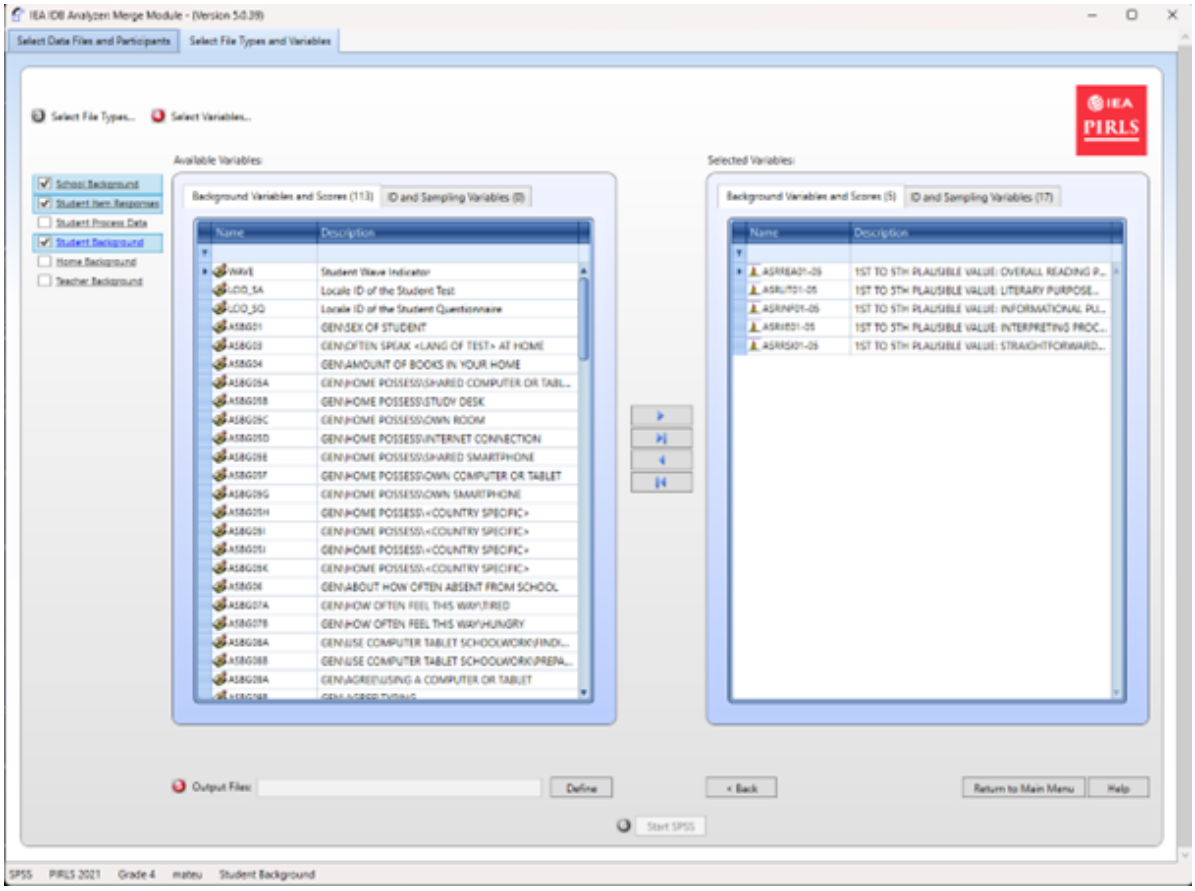

4. **(Opcjonalnie)** Kliknij Edit Country List, aby edytować etykiety krajów poprzez dodanie lub usunięcie pozycji (każda musi zawierać 3-literowy kod, kod ISO i pełną nazwę)

5. **Wybierz typy plików danych**: Kliknij Next, aby przejść do zakładki Select File Types and Variables. Zaznacz pola obok typów danych, które chcesz uwzględnić, np. School Context, Student Achievement, Student Context, Student Home, Teacher Context

> **!** Ważne
>
> Aby połączyć dane różnych typów, np. uczniów z danymi rodziców, szkół lub nauczycieli, wybierz odpowiednie typy plików. Upewnij się, że dane są łączone zgodnie z metodologią badania, ponieważ niektóre typy danych (np. nauczycielskie) wymagają analizy w odniesieniu do uczniów. Sprawdź dokumentację techniczną badania (np. w instrukcji Help), aby ustalić, jak poprawnie łączyć dane i interpretować wyniki.

6. **Wybierz zmienne**: Z listy Available Variables wybierz zmienne do analizy, przenosząc je do panelu Selected Variables za pomocą strzałki (→). Użyj strzałki (→|) do

zaznaczenia wszystkich zmiennych. Część zmiennych (identyfikacyjne i analityczne) przenosi się automatycznie. Zmienne identyfikacyjne i samplingowe są domyślnie wybrane




7. **Określ lokalizację zapisu**: W polu Output Files kliknij Define, aby wskazać folder i nazwę dla pliku wynikowego oraz skryptu (np. `*.R`, `*.SPS`, `*.SAS`). Nazwa pliku nie może zawierać znaków specjalnych

8. **Wygeneruj i uruchom skrypt**: Kliknij Start SPSS/SAS/R, aby utworzyć skrypt. Uruchom go w wybranym środowisku statystycznym (w R kliknij Source lub naciśnij Ctrl+Alt+R; w SPSS wybierz Run > All; w SAS kliknij Run lub Submit)

## 2.3 Moduł analizy (Analysis Module)

Analysis Module to kluczowy komponent programu IEA IDB Analyzer, który umożliwia generowanie gotowych skryptów analitycznych w środowiskach SPSS, SAS lub R. Moduł został zaprojektowany do analizy danych z międzynarodowych badań edukacyjnych, uwzględniając ich złożony schemat doboru próby, w tym wagi próbkowania, wagi replikacyjne oraz wartości prawdopodobne (*plausible values*, PV).

**i** Metodologia analizy

Program automatycznie wykorzystuje metodę Jackknife Repeated Replication (JRR) lub Balanced Repeated Replication (BRR), dzięki czemu poprawnie oblicza odchylenia standardowe analizowanych statystyk, w pełni uwzględniając strukturę losowania próby. Użytkownik może wybrać typ analizy, zmienne, a także określić, czy uwzględniać braki danych w zmiennych grupujących (domyślnie są one wykluczane, ale można je uwzględnić jako kategorie raportowania). W zależności od badania i wybranej analizy program sam dobiera odpowiednie metody. Wyniki są generowane w formie składni, którą należy uruchomić w wybranym oprogramowaniu statystycznym.

Poniżej przedstawiono dostępne typy analiz wraz z ich obsługą w różnych środowiskach:

| **Typ analizy** | **SPSS** | **SAS** | **R** |
|---|---|---|---|
| Rozkłady częstości (Percentages only) | • | • | • |
| Rozkłady częstości i średnie (z t-testami) (Percentages and Means) | • | • | • |
| Rozkłady umiejętności (Benchmarks) | • | • | • |
| Percentyle (Percentiles) | • | • | • |
| Korelacje (Correlations Pearson/Spearman) | • | • | • |
| Regresja liniowa (Linear Regression) | • | • | • |
| Regresja logistyczna (dychotomiczna) (Logistic Regression) | • | • | × |

### 2.3.1 Procentowe rozkłady częstości (Percentages only)

Ten typ analizy służy do obliczania procentowych rozkładów zmiennych kategorycznych z uwzględnieniem błędów standardowych. Analiza może być przeprowadzona z uwzględnieniem jednej lub więcej zmiennych grupujących, takich jak np. płeć czy kraj. Domyślnie jako pierwsza zmienna grupująca stosowana jest IDCNTRY, co umożliwia porównanie wyników między krajami. Opcja Separate Tables by generuje osobne tabele dla każdej analizowanej zmiennej.

> **i** Przykład zastosowania
>
> Możliwe jest określenie odsetka chłopców i dziewcząt wśród uczniów klasy czwartej w każdym z krajów uczestniczących w badaniu PIRLS.

### 2.3.2 Procentowe rozkłady i średnie (Percentages & Means)

Moduł oblicza procentowe rozkłady oraz średnie dla zmiennych ciągłych lub kategorycznych, wraz z odchyleniami standardowymi i błędami standardowymi. Generuje także statystyki t-testów do porównania średnich i procentów między grupami. Użytkownik wybiera zmienne grupujące (np. IDCNTRY, płeć) oraz zmienne analizowane (np. wyniki testów). Aby uwzględnić wyniki osiągnięć, w menu Plausible Value Option należy wybrać opcję Use PVs i określić zestaw wartości prawdopodobnych.

> **i** Przykład zastosowania
>
> Można porównać średnie wyniki z matematyki między chłopcami a dziewczętami w różnych krajach i sprawdzić istotność statystyczną różnic.

### 2.3.3 Poziomy umiejętności (Benchmarks)

Moduł Benchmarks oblicza odsetek respondentów osiągających określone poziomy biegłości (np. międzynarodowe benchmarki) lub punkty odcięcia zdefiniowane przez użytkownika. Analiza może być przeprowadzona w dwóch trybach:

- **Kumulatywny**: Odsetek respondentów na lub powyżej danego progu
- **Dyskretny**: Odsetek respondentów w każdym z określonych przedziałach

> **i** Przykład zastosowania
>
> Można zbadać, jaki procent uczniów w każdym kraju osiągnął określony poziom biegłości w czytaniu według skali PIRLS.

### 2.3.4 Percentyle (Percentiles)

Moduł Percentiles oblicza wartości punktowe (np. 25., 50., 75. percentyl), które dzielą rozkład zmiennych ciągłych (np. wyniki testów) na określone części. Analiza jest przeprowadzana w podgrupach zdefiniowanych przez zmienne grupujące (np. IDCNTRY, płeć). Użytkownik wpisuje percentyle (oddzielone spacją) w polu Percentiles.

> **i** Przykład zastosowania
>
> Można sprawdzić, jaki wynik z matematyki należało uzyskać, aby znaleźć się wśród 10% najlepszych uczniów (jaki wynik odpowiada 90. percentylowi w danym kraju).

### 2.3.5 Korelacje (Correlations)

Moduł korelacji umożliwia obliczanie współczynników korelacji Pearsona dla zmiennych ilościowych oraz korelacji rang Spearmana dla zmiennych porządkowych. Analiza może być przeprowadzona w podgrupach, np. w obrębie krajów. Umożliwia analizę związku między takimi zmiennymi, jak nastawienie do czytania a częstotliwość czytania. Możliwe jest również obliczanie korelacji między wynikami testu a dowolną skalą zawartą w badaniu.

> **i** Przykład zastosowania
>
> Możliwe jest zbadanie, czy istnieje związek między czasem poświęcanym na naukę a wynikami.

### 2.3.6 Regresja liniowa (Linear Regression)

Moduł regresji liniowej umożliwia budowę modeli statystycznych przewidujących wartość zmiennej zależnej na podstawie jednego lub wielu predyktorów. Analiza może uwzględniać zmienne ciągłe (np. liczba godzin nauki), i kategoryczne (np. płeć czy typ szkoły), które program automatycznie koduje (np. metodą dummy coding lub effect coding w menu Contrast). Moduł umożliwia także uwzględnienie wartości prawdopodobnych (*Plausible values*, PVs) jako zmiennych zależnych lub niezależnych.

> **i** Przykład zastosowania
>
> Użytkownik może zbudować model pokazujący wpływ płci, statusu społeczno-ekonomicznego oraz liczby godzin nauki na wyniki z matematyki. Wyniki przedstawiają szacowaną zmianę wartości zmiennej zależnej przy zmianie każdego z predyktorów, przy założeniu niezmienności pozostałych czynników.

### 2.3.7 Regresja logistyczna (Logistic Regression)

Moduł Logistic Regression umożliwia modelowanie prawdopodobieństwa wystąpienia zdarzenia binarnego (np. tak/nie) na podstawie zmiennych niezależnych (ciągłych lub kategorycznych). Generuje on składnię do analizy w programach SPSS lub SAS (moduł ten nie jest dostępny dla R). Model może uwzględniać zmienne niezależne o charakterze ciągłym (np. SES) lub kategorycznym (np. płeć).

> **i** Dodatkowe funkcje w SAS
>
> W przypadku programu SAS dostępna jest również opcja regresji wielomianowej, umożliwiająca analizę zmiennych zależnych posiadających więcej niż dwie kategorie.

Wyniki analizy obejmują współczynniki regresji, błędy standardowe oraz ilorazy szans (*odds ratios*), które wskazują, w jaki sposób zmienne niezależne wpływają na prawdopodobieństwo wystąpienia zdarzenia.

> **i** Przykład zastosowania
>
> Można zbadać, jak płeć ucznia i status socjoekonomiczny (SES) wpływają na szansę osiągnięcia poziomu „zaawansowanego" w czytaniu (np. w badaniu PIRLS).

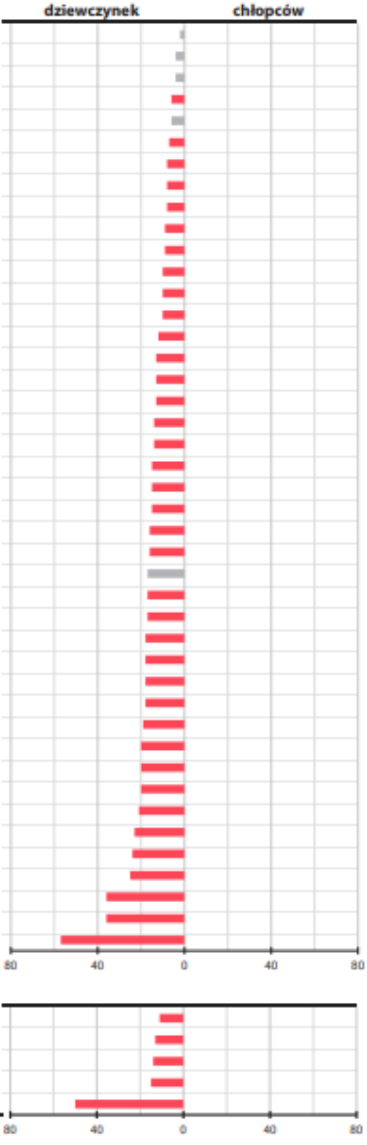
# 3 Przykładowa analiza z wykorzystaniem programu IDB Analyzer

Poniższy przykład opiera się na badaniu PIRLS 2021 i przedstawia analizę odpowiadającą na pytanie: **„Jaki jest średni wynik z czytania uczniów czwartej klasy — chłopców i dziewczynek?"** Analiza zostanie przeprowadzona za pomocą programu IEA IDB Analyzer, z wykorzystaniem oficjalnych danych ze strony IEA.

> **i** Kontekst analizy
>
> Wyniki analizy odpowiadają danym przedstawionym w Tabeli 5.5 krajowego raportu PIRLS 2021 (s. 65) i zostaną zreplikowane w tym przykładzie.

## Tabela 5.5. Średni wynik uczniów w zakresie czytania w podziale na płeć w krajach 1 i 3 fali badania

| Kraj | Dziewczynki | | Chłopcy | | Różnica | Różnica pomiędzy płciami |
|---|---|---|---|---|---|---|
| | Procent uczniów (%) | Średnia | Procent uczniów (%) | Średnia | | Wyższy wynik dziewczynek / Wyższy wynik chłopców |
| Hiszpania | 47 (0,9) | 522 (2,6) | 53 (0,9) | 520 (2,5) | 2 (2,6) | |
| Czechy | 49 (0,9) | 541 (2,8) | 51 (0,9) | 538 (2,7) | 4 (3,0) | |
| [3] Izrael ∞ | 50 (1,1) | 512 (2,8) | 50 (1,1) | 508 (2,6) | 4 (3,0) | |
| [2] Portugalia | 48 (0,7) | 523 (2,3) | 52 (0,7) | 517 (2,7) | 6 (2,0) | |
| Malta | 46 (3,4) | 518 (3,6) | 54 (3,4) | 512 (3,2) | 6 (4,1) | |
| [2] Włochy | 49 (0,6) | 541 (2,4) | 51 (0,6) | 534 (2,4) | 7 (2,0) | |
| Belgia (flamandzka) | 49 (0,8) | 515 (2,6) | 51 (0,8) | 507 (2,8) | 8 (2,8) | |
| [2 †] Hongkong (Chiny) | 51 (1,0) | 577 (2,8) | 49 (1,0) | 569 (3,3) | 8 (2,8) | |
| [†] Słowacja | 52 (0,9) | 533 (2,9) | 48 (0,9) | 525 (3,2) | 8 (2,8) | |
| Cypr | 51 (0,7) | 515 (3,2) | 49 (0,7) | 506 (3,1) | 9 (2,7) | |
| [3] Serbia | 49 (0,8) | 518 (3,4) | 51 (0,8) | 509 (3,2) | 9 (3,5) | |
| Makao (Chiny) | 50 (0,7) | 540 (1,5) | 50 (0,7) | 531 (1,9) | 10 (2,2) | |
| Anglia ∞ | 51 (0,9) | 562 (3,1) | 49 (0,9) | 553 (3,1) | 10 (3,7) | |
| [2] Belgia (francuskojęzyczna) | 49 (0,8) | 499 (3,2) | 51 (0,8) | 489 (2,9) | 10 (3,2) | |
| [2 †] Dania | 52 (0,6) | 545 (2,5) | 48 (0,6) | 533 (2,8) | 12 (3,0) | |
| [ǂ] Holandia | 50 (0,8) | 534 (2,9) | 50 (0,8) | 521 (2,8) | 13 (2,6) | |
| Tajwan | 48 (0,5) | 551 (2,5) | 52 (0,5) | 537 (2,4) | 13 (2,3) | |
| Rosja | 49 (0,7) | 574 (3,4) | 51 (0,7) | 561 (4,5) | 13 (3,7) | |
| Francja | 50 (0,7) | 521 (3,0) | 50 (0,7) | 507 (2,7) | 14 (2,6) | |
| Austria | 49 (0,9) | 537 (2,6) | 51 (0,9) | 523 (2,6) | 14 (2,7) | |
| [2] Szwecja | 50 (0,9) | 551 (2,5) | 50 (0,9) | 536 (2,3) | 15 (2,3) | |
| Bułgaria | 48 (0,9) | 548 (3,0) | 52 (0,9) | 533 (4,0) | 15 (3,9) | |
| Niemcy | 49 (0,8) | 532 (2,5) | 51 (0,8) | 516 (2,5) | 15 (2,6) | |
| [2] Egipt | 49 (1,5) | 386 (5,7) | 51 (1,5) | 370 (6,4) | 16 (5,6) | |
| Norwegia (klasa 5) | 49 (0,7) | 547 (2,3) | 51 (0,7) | 531 (2,4) | 16 (2,4) | |
| Iran ∞ | 46 (2,3) | 422 (7,5) | 54 (2,3) | 405 (5,9) | 17 (9,1) | |
| [2] Turcja | 49 (0,6) | 505 (3,8) | 51 (0,6) | 488 (3,6) | 17 (2,8) | |
| Australia ∞ | 50 (0,7) | 549 (2,5) | 50 (0,7) | 532 (2,8) | 17 (3,0) | |
| Finlandia | 50 (0,8) | 558 (2,7) | 50 (0,8) | 541 (2,7) | 18 (2,7) | |
| [3] Singapur | 49 (0,6) | 596 (3,0) | 51 (0,6) | 578 (3,7) | 18 (2,7) | |
| Azerbejdżan | 47 (0,8) | 450 (4,1) | 53 (0,8) | 432 (4,0) | 18 (3,7) | |
| Słowenia | 49 (0,7) | 529 (2,1) | 51 (0,7) | 511 (2,3) | 18 (2,3) | |
| [†] Nowa Zelandia | 49 (0,7) | 531 (2,9) | 51 (0,7) | 512 (2,7) | 19 (3,2) | |
| [3] Czamogóra | 48 (0,6) | 497 (2,0) | 52 (0,6) | 478 (2,2) | 20 (2,6) | |
| **Polska** | **47 (1,0)** | **560 (2,5)** | **53 (1,0)** | **540 (2,7)** | **20 (2,9)** | |
| [2] Albania | 49 (1,0) | 523 (3,5) | 51 (1,0) | 503 (3,4) | 20 (3,2) | |
| [2] Kosowo | 51 (0,9) | 431 (3,1) | 49 (0,9) | 410 (3,8) | 21 (3,1) | |
| [2 †] Brazylia ∞ | 49 (1,1) | 431 (6,0) | 51 (1,1) | 408 (6,1) | 23 (6,0) | |
| Uzbekistan | 48 (0,9) | 449 (3,1) | 52 (0,9) | 425 (3,5) | 24 (3,4) | |
| Macedonia Północna | 51 (1,0) | 454 (5,8) | 49 (1,0) | 429 (6,0) | 25 (5,2) | |
| Oman | 50 (0,6) | 447 (4,2) | 50 (0,6) | 412 (4,1) | 36 (3,8) | |
| Jordania | 51 (2,6) | 398 (6,8) | 49 (2,6) | 362 (7,9) | 36 (10,3) | |
| Południowa Afryka ∞ | 49 (0,6) | 317 (4,4) | 51 (0,6) | 260 (5,0) | 57 (3,6) | |
| **Średnia międzynarodowa** | **49 (0,2)** | **509 (0,5)** | **51 (0,2)** | **493 (0,6)** | | |
| **Dodatkowe miasta i regiony** | | | | | | |
| Rosja (Moskwa) | 49 (0,7) | 604 (2,2) | 51 (0,7) | 593 (2,5) | 11 (2,1) | |
| [2] Kanada (Kolumbia Brytyjska) | 49 (1,0) | 542 (3,5) | 51 (1,0) | 529 (4,3) | 13 (3,3) | |
| [2] Kanada (Nowa Funlandia i Labrador) | 50 (1,2) | 530 (3,1) | 50 (1,2) | 516 (4,3) | 14 (4,0) | |
| [3 ǂ] Kanada (Alberta) | 49 (1,5) | 546 (4,1) | 51 (1,5) | 531 (4,2) | 15 (4,3) | |
| Południowa Afryka (klasa 6) ∞ | 52 (0,7) | 408 (4,5) | 48 (0,7) | 359 (5,2) | 50 (3,9) | |




∞ Kraje, które przeprowadziły badanie rok później, niż zakładano (z uczniami klasy czwartej kolejnego rocznika).

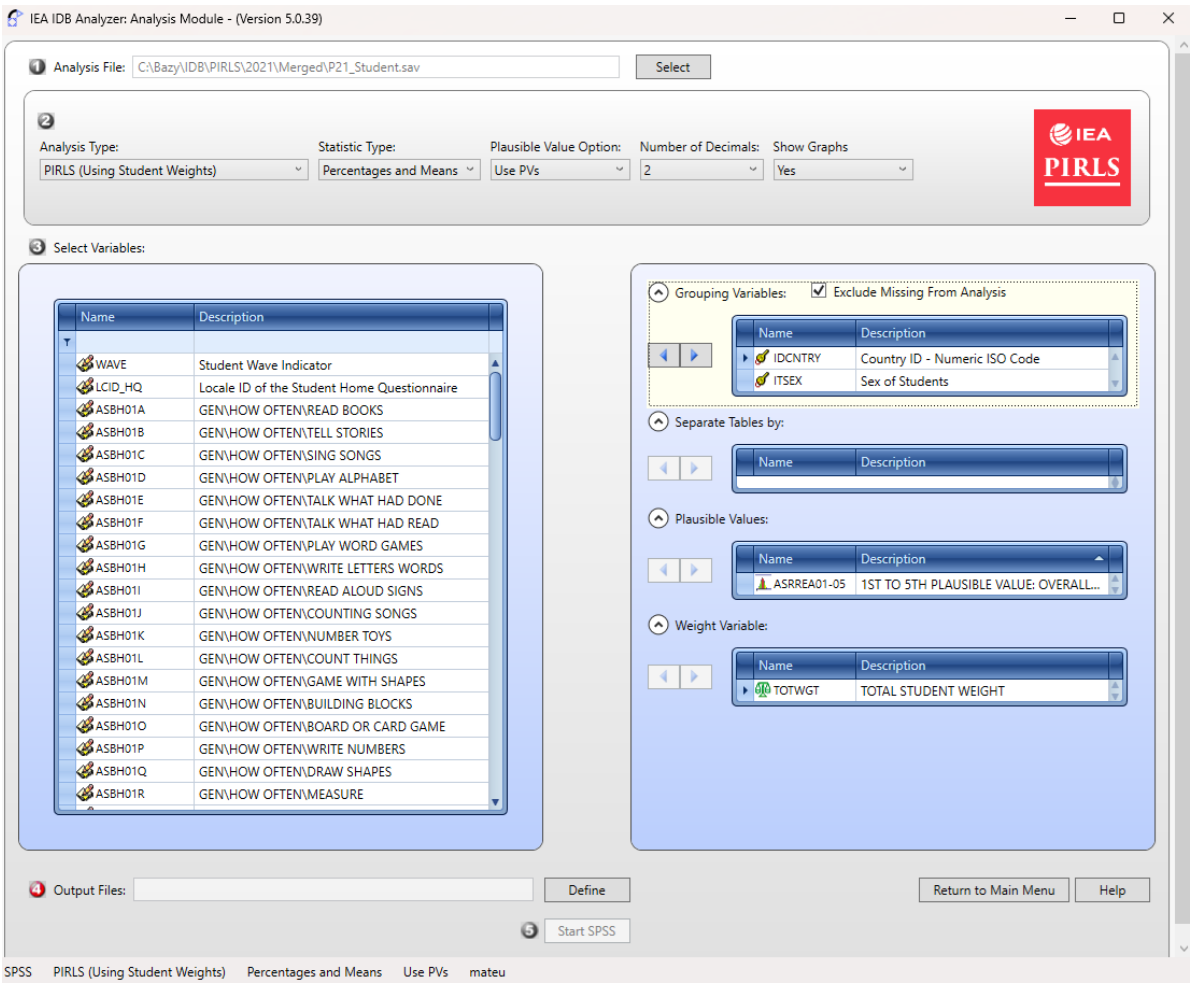
Kraje zaprezentowane w porządku rosnącym ze względu na różnicę pomiędzy płciami.
W nawiasie ( ) podano błąd standardowy. Ze względu na zaokrąglenia niektóre wyniki mogą wydawać się niespójne.
Oznaczenia dotyczące definicji populacji (1, 2, 3) opisano w tabeli 3.1. Oznaczenia dotyczące poziomu realizacji próby (†, ǂ, ǂǂ) opisano w tabeli 3.2.

Różnica istotna statystycznie
Różnica nieistotna statystycznie

## 3.1 Krok 1: Pobranie i przygotowanie danych

1. Pobierz dane z badania PIRLS 2021 ze strony IEA Study Data Repository
2. **Scal dane za pomocą Merge Module** w programie IEA IDB Analyzer:
   - Uruchom program i wybierz Merge Module
   - Wskaż odpowiednie pliki danych źródłowych (Student Background oraz Student Item Response) zgodnie z instrukcjami z rozdziału 2.2
   - Zaznacz kraje, które chcesz uwzględnić w analizie (w przykładzie wybrano Bułgarię, Danię, Irlandię oraz Polskę)
   - Wygeneruj i uruchom kod, który utworzy scalony plik danych

## 3.2 Krok 2: Uruchomienie IDB Analyzer - moduł analityczny



1. Otwórz moduł **Analysis Module** w programie IEA IDB Analyzer
2. Wybierz scalony plik danych utworzony w Kroku 1 jako Analysis File, klikając przycisk Select
3. Wybierz **PIRLS (Using Student Weights)** jako typ analizy (Analysis Type)
4. Wybierz **Percentages & Means** jako Statistic Type
5. Wybierz opcję **Use PVs** w menu Plausible Value Option

> Ustawienia formatowania
>
> - Domyślna wartość w menu Number of Decimals wynosi 2. Zmiana tej wartości wpływa jedynie na liczbę widocznych miejsc dziesiętnych w plikach wyjściowych
> - Domyślna wartość w menu Show Graphs to Yes. Wybranie opcji Yes spowoduje wygenerowanie wykresów w pliku wyjściowym

6. **Określ zmienną płeć (ITSEX)** jako drugą zmienną grupującą (pierwszą domyślnie jest IDCNTRY), klikając najpierw pole Grouping Variables, aby je aktywować. Następnie wybierz zmienną ITSEX z listy zmiennych w lewym panelu i przenieś ją do pola Grouping Variables, klikając strzałkę w prawo (→)

> Ważne
>
> Program automatycznie zaznacza opcję **Exclude Missing From Analysis**, która wyklucza przypadki z brakującymi danymi w zmiennych grupujących. Ta opcja powinna być zaznaczona dla tej analizy.

7. Pole **Separate Tables by** powinno pozostać puste (dotyczy analiz z wieloma zmiennymi grupującymi lub zmiennymi zależnymi innymi niż wynik osiągnięć)
8. **Określ wyniki testów do analizy**, klikając najpierw pole Plausible Values, aby je aktywować. Następnie wybierz zestaw wartości prawdopodobnych (ASRREA01-05) z listy dostępnych zmiennych w lewym panelu i przenieś je do pola Plausible Values po prawej stronie, klikając strzałkę w prawo (→)
9. Zmienna wagowa jest wybierana automatycznie przez program; dla danych kontekstowych uczniów domyślnie wybierana jest TOTWGT
10. **Określ nazwę dla plików wynikowych** oraz folder, w którym będą przechowywane, klikając przycisk Define w polu Output Files. IEA IDB Analyzer utworzy również skrypt R (`.R`), składnię SPSS (`.SPS`) lub składnię SAS (`*.SAS`) o tej samej nazwie i w tym samym folderze, zawierające kod niezbędny do przeprowadzenia analizy

**i** Przykład nazewnictwa

Skrypt i pliki wyjściowe mogą być nazwane WynikiSex i zapisane w folderze `C:\PIRLS2021\Analizy`

11. **Kliknij przycisk Start R** (lub Start SPSS/SAS), aby utworzyć skrypt R (lub składnię SPSS/SAS) i otworzyć go do wykonania. Program wyświetli ostrzeżenie, jeśli plik w określonym folderze ma zostać nadpisany
12. **Uruchom skrypt**: Skrypt R można wykonać, klikając przycisk Source lub naciskając Ctrl+Alt+R na klawiaturze. W SPSS otwórz menu Run i wybierz opcję All. W SAS kliknij przycisk Run (lub wybierz Submit w menu Run)

## 3.3 Interpretacja wyników

**!** Pliki wynikowe

Program IEA IDB Analyzer generuje i zapisuje wyniki w folderze określonym w kroku 10. Dla statystyki Percentages & Means z dodatkową zmienną grupującą (np. płeć, obok IDCNTRY), generowane są dwa dodatkowe pliki wyników:

- plik z sufiksem **„_sig”**, który raportuje istotność różnic między grupami analitycznymi - w tym przypadku między dziewczętami a chłopcami - dla każdego kraju
- plik z sufiksem **„_sig2”**, który raportuje istotność różnic między krajami w obrębie każdej grupy płci

Poniżej przedstawiono plik wynikowy “_sig” wygenerowany dla analizy wykonanej w R.

# Report

Analysis for ASRREA by IDCNTRY ITSEX

| Country ID - Numeric ISO Code | Sex of Students | N of Cases | Sum of TOTWGT | Sum of TOTWGT (s.e.) | Percent | Percent (s.e.) | ASRREA (Mean) | ASRREA (s.e.) | Confidence Interval (95) | Std.Dev. | Std.Dev. (s.e.) | Percent Missing | Number of Variance Strata |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bulgaria | Girl | 1976 | 28301 | 621.20 | 47.77 | 0.95 | 547.82 | 3.01 | 542 to 554 | 86.66 | 2.36 | 0.00 | 76 |
| Bulgaria | Boy | 2067 | 30947 | 691.63 | 52.23 | 0.95 | 532.64 | 3.99 | 525 to 540 | 88.58 | 3.34 | 0.00 | 76 |
| Denmark | Girl | 2479 | 30375 | 526.85 | 51.61 | 0.65 | 545.02 | 2.54 | 540 to 550 | 71.80 | 1.30 | 0.00 | 99 |
| Denmark | Boy | 2342 | 28477 | 522.28 | 48.39 | 0.65 | 532.58 | 2.81 | 527 to 538 | 73.96 | 2.03 | 0.00 | 99 |
| Ireland | Girl | 2303 | 33545 | 1154.95 | 48.90 | 1.01 | 582.99 | 3.27 | 577 to 589 | 77.87 | 1.99 | 0.00 | 71 |
| Ireland | Boy | 2360 | 35052 | 1199.75 | 51.10 | 1.01 | 571.91 | 2.81 | 566 to 577 | 75.13 | 1.48 | 0.00 | 70 |
| Poland | Girl | 2016 | 155977 | 3957.16 | 47.01 | 0.98 | 559.75 | 2.48 | 555 to 565 | 69.86 | 1.72 | 0.00 | 75 |
| Poland | Boy | 2163 | 175785 | 3952.94 | 52.99 | 0.98 | 539.69 | 2.71 | 534 to 545 | 72.31 | 2.15 | 0.00 | 75 |

W wynikowej tabeli możemy przeczytać, że dziewczęta w Polsce osiągnęły średni wynik 559,75 (błąd standardowy: 2,48), natomiast chłopcy 539,69 (błąd standardowy: 2,71). Różnica wynosi więc 20,06 punktu.

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| IDCNTRY | refgroup | compgroup | pct | pct_se | cpct | cpct_se | pctdiff | pctdiff_se | pctdiff_t | mnpv | mnpv_se | cmnpv | cmnpv_se | mnpvdiff | mnpvdiff_se | mnpvdiff_t |
| Bulgaria | Girl | Boy | 47,77 | 0,95 | 52,23 | 0,95 | 4,47 | 1,89 | 2,36 | 547,82 | 3,01 | 532,64 | 3,99 | -15,18 | 3,86 | -3,93 |
| Bulgaria | Boy | Girl | 52,23 | 0,95 | 47,77 | 0,95 | -4,47 | 1,89 | -2,36 | 532,64 | 3,99 | 547,82 | 3,01 | 15,18 | 3,86 | 3,93 |
| Denmark | Girl | Boy | 51,61 | 0,65 | 48,39 | 0,65 | -3,23 | 1,30 | -2,49 | 545,02 | 2,54 | 532,58 | 2,81 | -12,43 | 2,98 | -4,17 |
| Denmark | Boy | Girl | 48,39 | 0,65 | 51,61 | 0,65 | 3,23 | 1,30 | 2,49 | 532,58 | 2,81 | 545,02 | 2,54 | 12,43 | 2,98 | 4,17 |
| Ireland | Girl | Boy | 48,90 | 1,01 | 51,10 | 1,01 | 2,20 | 2,02 | 1,08 | 582,99 | 3,27 | 571,91 | 2,81 | -11,07 | 3,52 | -3,15 |
| Ireland | Boy | Girl | 51,10 | 1,01 | 48,90 | 1,01 | -2,20 | 2,02 | -1,08 | 571,91 | 2,81 | 582,99 | 3,27 | 11,07 | 3,52 | 3,15 |
| Poland | Girl | Boy | 47,01 | 0,98 | 52,99 | 0,98 | 5,97 | 1,96 | 3,04 | 559,75 | 2,48 | 539,69 | 2,71 | -20,06 | 2,95 | -6,80 |
| Poland | Boy | Girl | 52,99 | 0,98 | 47,01 | 0,98 | -5,97 | 1,96 | -3,04 | 539,69 | 2,71 | 559,75 | 2,48 | 20,06 | 2,95 | 6,80 |

**Istotność statystyczna różnic** różnic między płciami jest określana na podstawie pliku wyjściowego z sufiksem „_sig”. W kolumnie mnpvdiff raportowana jest różnica średnich wyników, a w kolumnie mnpvdiff_se - jej błąd standardowy. Iloraz tych wartości daje statystykę t (mnpvdiff_t). Dla poziomu istotności $\alpha = 0,05$, wartości t poniżej -1,96 lub powyżej +1,96 wskazują na istotne różnice statystyczne. W przypadku Polski wartość t wynosi -6,80, co oznacza, że różnica między płciami jest statystycznie istotna.