

PORTFOLIO ANALITYKA DANYCH

MARCIN DESZCZKA

adres e-mail: marcin.deszczka at [gmail.com](mailto:marcin.deszczka@gmail.com)

Linkedin: <https://www.linkedin.com/in/marcin-deszczka-407b113a9/>

Analiza idealnej liczby dzieci w rodzinie w opinii respondentów na podstawie badania ISSP 2022

Executive Summary:

- Najczęściej wskazywana idealna liczba dzieci w rodzinie to 2 (ponad 50% badanych).
- Różnice między płciami są minimalne. Płeć nie jest istotnym predyktorem deklarowanej idealnej liczby dzieci po uwzględnieniu wieku i kraju.
- Istnieją znaczące różnice między krajami, kraj o najwyższej średniej (Izrael) ma wynik o połowę wyższy niż kraje o najniższej średniej (Włochy, Czechy, Bułgaria). Kraj zamieszkania pozostaje najważniejszym czynnikiem, z wyraźnymi różnicami między krajami (np. Izrael +1,28 dziecka względem kraju bazowego).
- Wiek ma ograniczony wpływ na średnią, ale wpływa na zmienność odpowiedzi. Ma też minimalny, ale statystycznie istotny efekt. W praktyce nie jest to duża zmiana.
- Model płeć + wiek + kraj wyjaśnia **13,2% wariacji**, co w kontekście danych społecznych jest umiarkowanie silną wartością — większość różnic w deklarowanej idealnej liczbie dzieci wynika z krajowych różnic kulturowych.
- Pomimo, że większość państw przebadanych ma dzietność poniżej poziomu zastąpienia (2,1 dziecka na kobietę) nadal międzynarodowa wśród 32 państw średnia liczba idealnej liczby dzieci w rodzinie wynosi 2,47 dziecka.

1. Cel projektu:

Celem projektu jest analiza idealnej liczby dzieci w rodzinie deklarowanej przez respondentów.

2. Wykorzystywane dane:

W analizie wykorzystano dane z badania ISSP – *Family and Changing Gender Roles V* (2022).

3. Badanie ISSP przeprowadzono na **45762** osobach. Badania przeprowadzono w 32 krajach - Australii, Austrii, Bułgarii, Tajwanie, Chorwacji, Czechach, Danii, Finlandii, Francji, Niemczech, Grecji, Węgrzech, Islandii, Indii, Izraelu, Włochy, Japonii, Litwy, Holandia, Nowa Zelandia, Norwegii, Filipinów, Polski, Rosji, Słowacji, Słowenii, Republiki Południowej Afryki, Hiszpanii, Szwecji, Tajlandii i Stanach Zjednoczonych. ISSP należy do najbardziej uznanych międzynarodowych programów badań porównawczych.

4. Projekt bada:

- ☐ średnie w populacjach
- ☐ różnice między płciami
- ☐ różnice między krajami
- ☐ różnice między grupami demograficznymi
- ☐ istotność statystyczną.

5. **Analiza globalnych preferencji rodzinnych (ISSP 2022)** – przetwarzanie i wizualizacja danych 45762 respondentów z 32 krajów, uwzględniając wagi. Identyfikacja kluczowych trendów: średnia idealnej liczby dzieci, różnice między krajami i wiekiem. Analiza obejmuje przetwarzanie danych o dużej skali oraz zastosowanie metod statystyki opisowej i wizualizacji.

6. Dane:

- Liczba obserwacji: 45762 respondentów
- Liczba krajów: 32
- Typ badania: międzynarodowe badanie ankietowe
- Dane ważone (uwzględniono zmienną `WEIGHT_COM`)
- Jako narzędzie do analizy wybieramy język programowania Python.

7. Narzędzia analityczne:

Język Python, w tym biblioteki tego języka:

- ☐ `pandas` – przetwarzanie danych
- ☐ `numpy` – obliczenia numeryczne
- ☐ `matplotlib / seaborn` – wizualizacja
- ☐ `pycountry` – mapowanie kodów ISO na nazwy krajów
- ☐ `statsmodels` – testy statystyczne.

Język Python umożliwia nam szybką analizę, czyszczenie danych, mapowanie kodów (np. ISO dzięki bibliotece `pycountry`) oraz automatyczne tworzenie wizualizacji, co mogłoby być bardziej czasochłonne w arkuszu kalkulacyjnym. Python też pozwala na szybsze zastosowanie wag analitycznych. Zamiast serii prostych testów istotności (ANOVA/t-Student), zdecydowano się na wielozmiennowy model regresji liniowej (biblioteka `statsmodels`). Pozwala on na jednoczesną kontrolę wielu zmiennych (kraj, wiek, płeć) i ocenę ich relatywnego wpływu na wynik (czysty efekt zmiennej).

8. Zmienne wykorzystane w analizie:

Zmienna	Opis
"v19"	idealna liczba dzieci w rodzinie
"sex"	płeć respondenta
"country"	kraj (kod ISO)

"age"	wiek
"WEIGHT_COM"	waga analityczna

9. Czyszczenie Danych

Z analizy w przypadku zmiennej "v19" wykluczono odpowiedzi „nie wiem” / „trudno powiedzieć” (kod -9) oraz „ile jest to możliwe” (kod 95), ponieważ nie umożliwiają one prawidłowej estymacji średniej arytmetycznej. Dodatkowo w przypadku prezentacji odsetków procentowych grupujemy odpowiedzi więcej 6 dzieci i więcej jako jedną, aby nasza wizualizacja była bardziej czytelna.

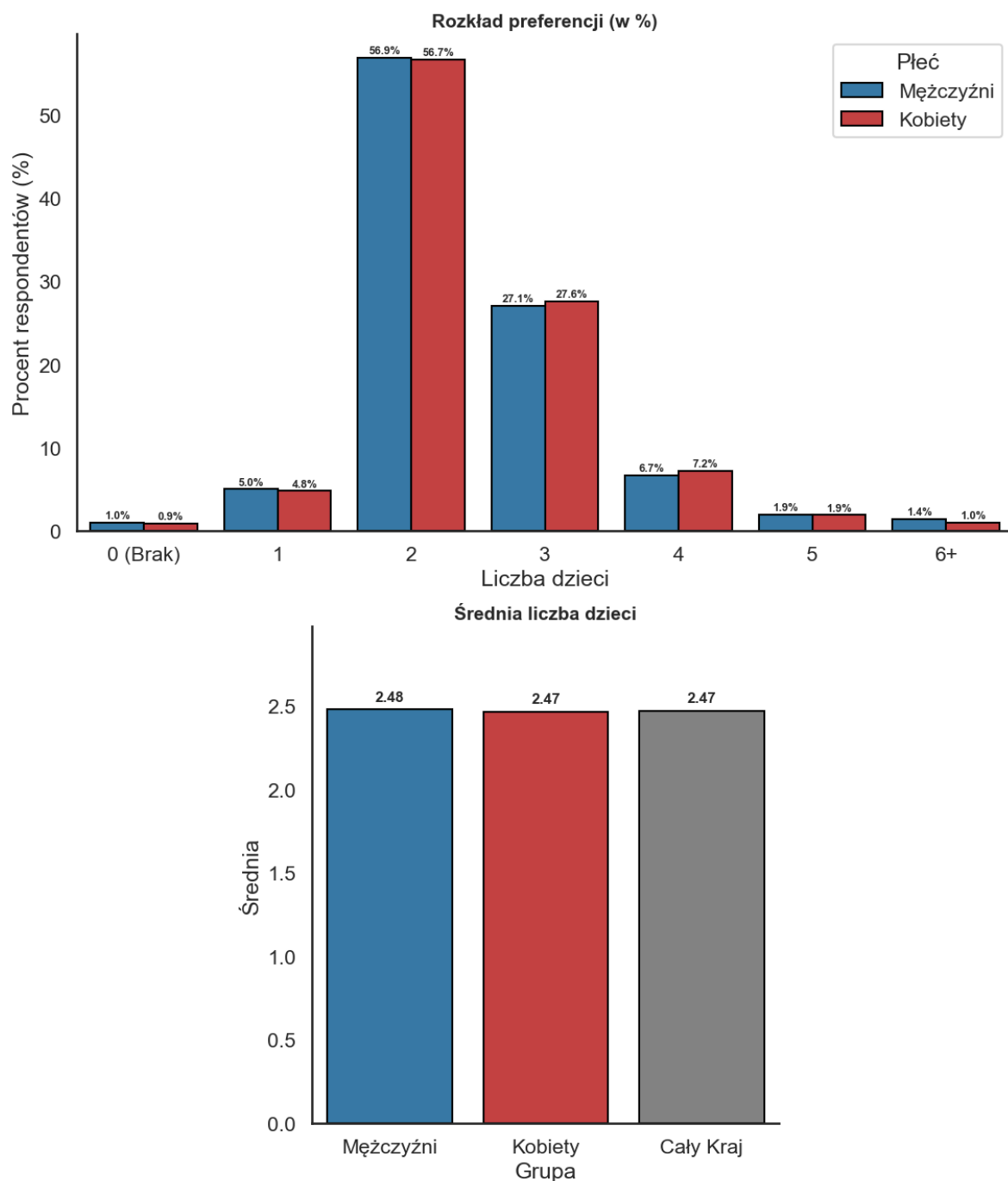
10. Waga analityczna

W analizie zastosowano zmienną WEIGHT_COM, która stanowi kombinację wag konstrukcyjnych (design weights) oraz wag post-stratyfikacyjnych. Waga ta koryguje zarówno nierówne prawdopodobieństwo doboru respondentów do próby, jak i ewentualne odchylenia struktury próby od struktury populacji (np. w zakresie wieku, płci czy regionu).

Zastosowanie WEIGHT_COM pozwala uzyskać estymacje reprezentatywne dla populacji krajowych, minimalizując ryzyko błędów systematycznych. Użycie wyłącznie wag konstrukcyjnych (DWEIGHT_COM) nie uwzględniałoby korekt post-stratyfikacyjnych, natomiast użycie wyłącznie PSWEIGHT pomijałoby aspekty związane z konstrukcją próby.

11. Idealna liczba dzieci wg płci i ogółu populacji przebadanej populacji

W pierwszym etapie oszacowano średnią idealną liczbę dzieci dla całej populacji oraz w podziale na płeć. Obliczono również rozkład procentowy odpowiedzi według płci. Analizę przeprowadzono z wykorzystaniem zmiennych „v19” i „sex” oraz wagi analitycznej WEIGHT_COM, zapewniającej reprezentatywność wyników. Wyniki przedstawiono na wykresie słupkowym (tab. 1).



Tab. 1. Idealna liczba w rodzinie wg całej reprezentatywnej przebadanej grupy i wg płci.

12. Jak widzimy w tab. 1., największą grupą osób są osoby, które uważają, że idealna liczba dzieci to dwójka. Jest to większość w przypadku obu płci – 56,7% dla kobiet i 56,9% dla mężczyzn. Kolejną najpopularniejszą odpowiedź to trójka dzieci – 27,1% dla mężczyzn i 27,6% dla kobiet. Trzecią najpopularniejszą odpowiedzią

jest czwórka dzieci – 6,7% w przypadku mężczyzn i 7,2% w przypadku kobiet. Średnia dla obu płci jest niemal identyczna – 2,48 i 2,47 dla kobiet. Dla całej przebadanej populacji – 2,47 dzieci.

13. Płeć i idealna liczba dzieci. Czy jest istotna statystycznie?

Dzięki wykorzystaniu regresji ważoną (WLS), można stwierdzić, że średnia idealna liczba dzieci w rodzinie dla mężczyzn wynosi 2,48, natomiast dla kobiet jest ona średnio o 0,01 niższa (dokładnie 0,013).

Współczynnik przy zmiennej płeć nie jest istotny statystycznie ($p = 0,198$ przy poziomie istotności $\alpha = 0,05$; $p > 0,05$), co oznacza, że nie ma podstaw do stwierdzenia istotnej różnicy między kobietami a mężczyznami w zakresie deklarowanej idealnej liczby dzieci.

Wartość współczynnika determinacji $R^2 = 0,00$ wskazuje, że płeć nie wyjaśnia zróżnicowania zmiennej „idealna liczba dzieci w rodzinie”.

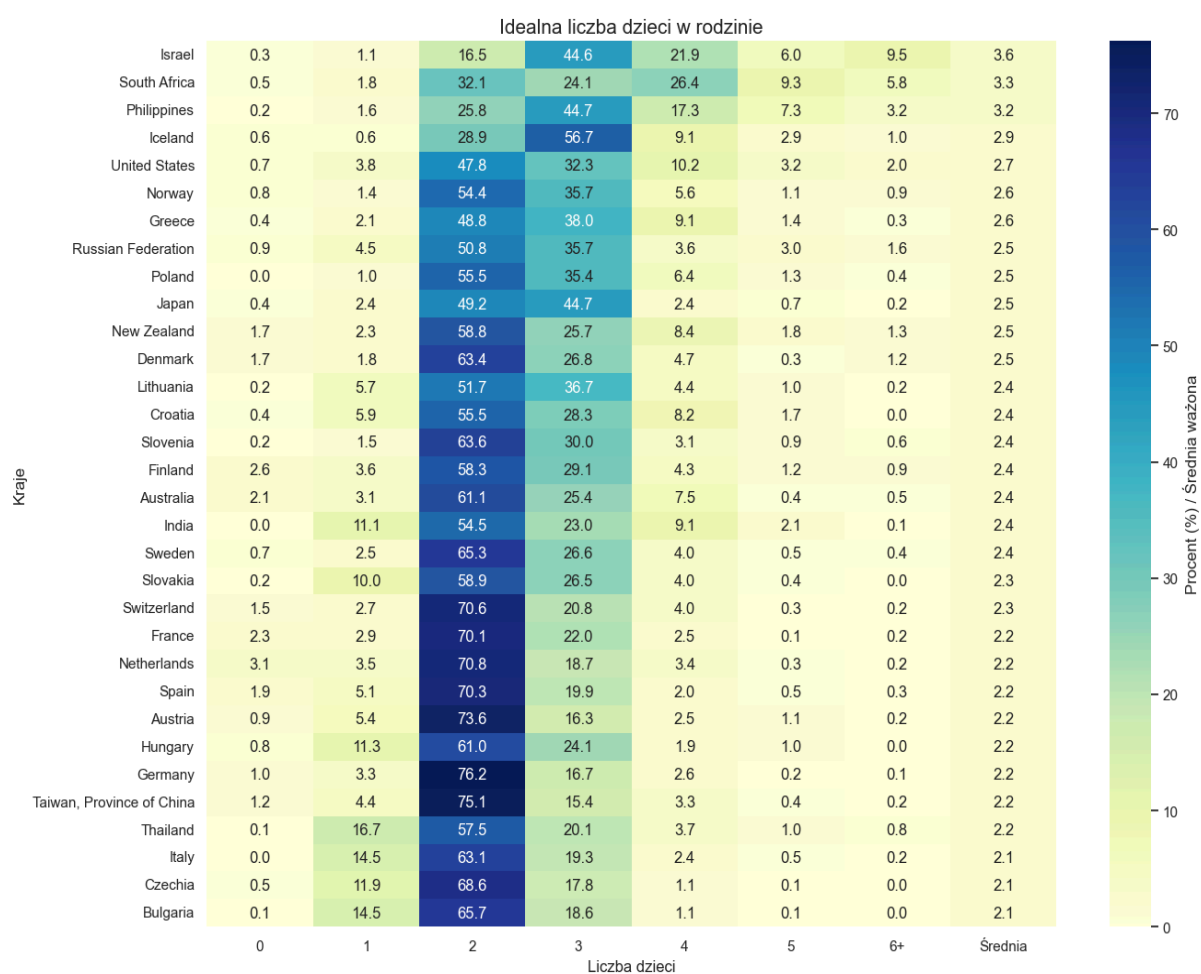
Test F ocenia istotność modelu jako całości. Ponieważ jego wartość jest nieistotna ($p = 0,198 > 0,05$), model regresji nie jest istotny statystycznie.

Płeć nie stanowi istotnego predyktora idealnej liczby dzieci w rodzinie.

14. Idealna liczba dzieci w rodzinie w podziale na kraje

W kolejnym etapie przeprowadzono analizę w podziale na kraje z wykorzystaniem zmiennej „country”. Wyniki zaprezentowano w formie mapy cieplnej (heatmapy), umożliwiającej porównanie rozkładów odpowiedzi oraz średnich wartości między krajami. Kraje zostały posortowane malejąco według średniej idealnej liczby dzieci. Wyniki przedstawiono w tab. 2. Ponieważ kraje są zakodowane

liczbowo jako kod ISO wykorzystujemy bibliotekę pycountry, która zamienia nam kody na nazwy krajów. W wyniku tych działań uzyskano czytelną wizualizację porównawczą, której przedstawiono w tab. 2. Najwyższą średnią "idealnej liczby dzieci" ma Izrael (3,6), dalej Republika Południowej Afryki (3,3) i Filipiny (3,2). Najniższą średnią – Włochy, Czechy i Bułgaria (po 2,1).



Tab. 2. Idealna liczba dzieci w rodzinie wg krajów.

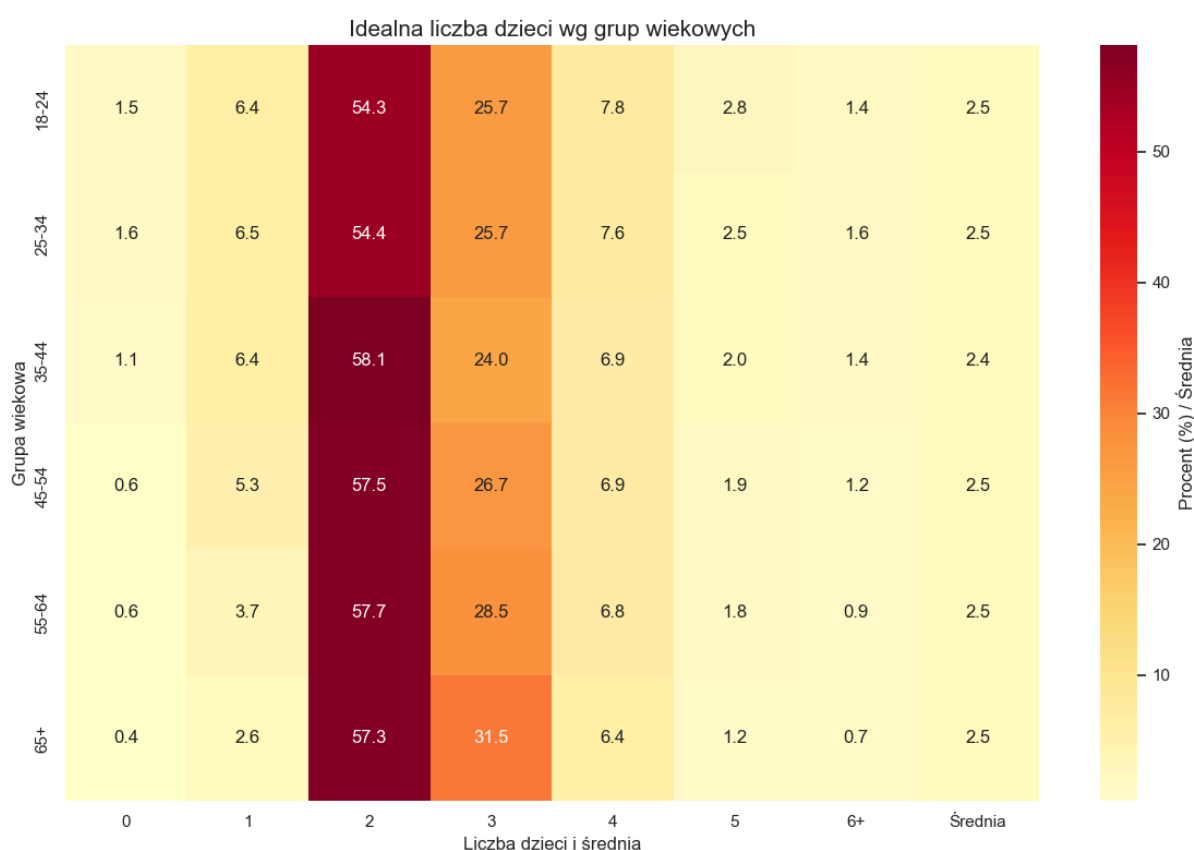
Kraj i idealna liczba dzieci w rodzinie. Czy zmienna jest istotna statystycznie?

Wyniki regresji ważonej wskazują, że kraj zamieszkania jest istotnym predyktorem deklarowanej idealnej liczby dzieci ($p < 0,001$). Model

wyjaśnia 12,7% wariancji zmiennej zależnej ($R^2 = 0,127$), co wskazuje na umiarkowaną siłę efektu. Oznacza to, że kraje w badaniu różnią się między sobą w odróżnieniu od płci z całej populacji badania.

15. Idealna liczba dzieci w rodzinie w podziale na wiek

Następnie przeanalizowano idealną liczbę dzieci w podziale na grupy wiekowe. W celu zwiększenia czytelności respondentów podzielono na następujące kategorie: 18–24, 25–34, 35–44, 45–54, 55–64 oraz 65+. Wyniki przedstawiono w formie mapy cieplnej (tab. 3).

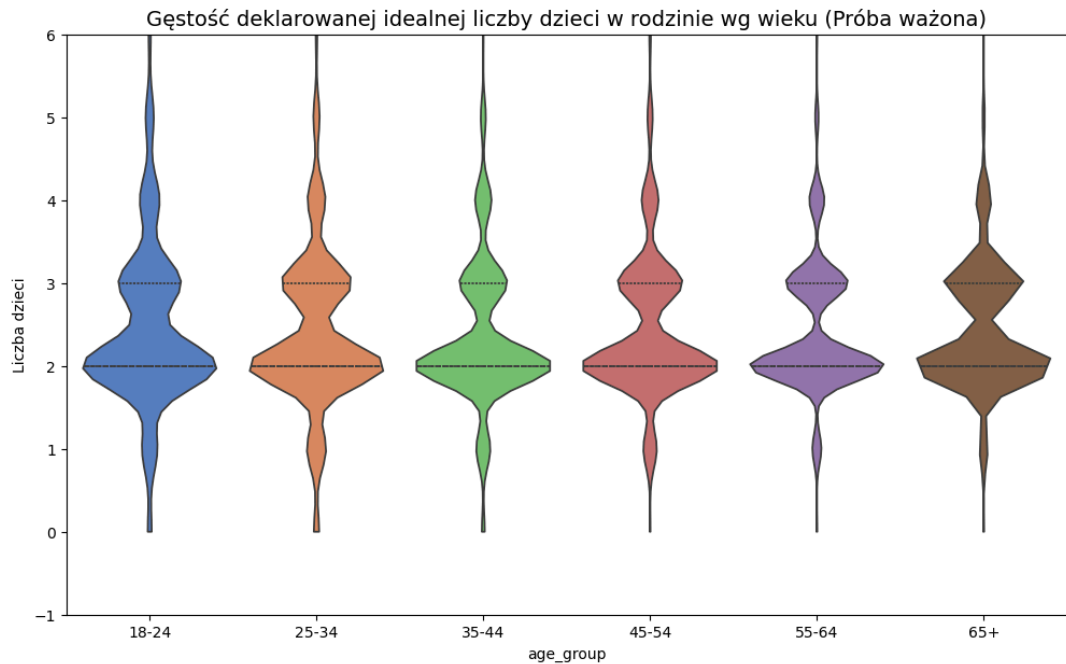


Tab. 3. Idealna liczba dzieci wg podziału respondentów wg wieku w formie mapy cieplnej.

16. Średnia idealna liczba dzieci we wszystkich grupach wiekowych wynosi około 2,5, z wyjątkiem grupy 35–44 lata, gdzie odnotowano nieznacznie niższą wartość (2,4). Różnice między grupami wiekowymi są minimalne,

jednak młodsze osoby częściej wskazują skrajne odpowiedzi (0–1 oraz 3–6+), podczas gdy w starszych grupach odpowiedzi są bardziej skoncentrowane wokół średniej.

Dla dodatkowej ilustracji rozkładu zastosowano wykres wiolinowy (tab. 4).



Tab. 4. Idealna liczba dzieci wg podziału respondentów wg wieku w wykresie wiolinowym

17. Wiek i idealna liczba dzieci. Czy jest istotna statystycznie?

Wiek ma formalnie istotny wpływ na zmienną, gdyż $p\text{-value}$ wynosi 0.0218 < 0.05 , **ale praktycznie znikomy wpływ** na deklarowaną idealną liczbę dzieci. Na każde 100 lat życia (hipotetycznie) efekt wynosiłby ok. 0,07 dziecka, co pokazuje, że w praktyce wiek jest bardzo słabym predyktorem.

18. Pełny model wielozmieniowy

Model wyjaśnia 13,2% wariacji deklarowanej idealnej liczby dzieci.

W porównaniu do pojedynczych modeli:

- płeć: $R^2 \approx 0$
- wiek: $R^2 \approx 0.0001$
- kraj: $R^2 \approx 0.127$

Kraj wyjaśnia większość tej wariancji, a wiek wnosi niewielki dodatkowy wkład do wyjaśnianej wariancji.

Wpływ zmiennych:

zmienna	coef	p-value	interpretacja
płeć	-0.015	0.116	nieistotna
wiek	+0.0041	0.000	istotna statystycznie, stały efekt
kraj	różne	0.000	istotnie statystyczna i bardzo silny efekt

Współczynniki krajowe:

Top 5 największych odchyłeń od bazy (względem średniej w przebadanej populacji):

kraj	odchylenie od bazy
Izrael	1.281
RPA	0.955
Filipiny	0.814
Islandia	0.466
USA	0.307

19. Wartość analityczna projektu :

- pracę na dużym zbiorze danych (45 tys. obserwacji),
- zastosowanie wag analitycznych,
- transformację i przygotowanie danych do analizy,
- wykorzystanie wizualizacji eksploracyjnych,
- interpretację wyników w kontekście społecznym i porównawczym.

20. Czego projekt nie robi i nie zamierza robić:

- projekt nie formułuje wniosków przyczynowo-skutkowych, ponieważ dane mają charakter obserwacyjny, a nie eksperymentalny,
- nie rekomenduje odpowiedniej polityki natalistycznej (prodemograficznej)
- nie mówi o tym ile respondent chce mieć dzieci, tylko o wyobrażonej idealnej dzieci w rodzinie wg badanego.

