

Opis wyników

Niniejszą analizę wykonano w Systemie Zautomatyzowanego Tworzenia Opisu Statystycznego - SZTOS (Hryniewicz, Milewska, 2023). Wizualizację wyników przeprowadzono z wykorzystaniem pakietu graficznego "ggplot2" (Wickham, 2016). Niniejsza analiza ma na celu opis zmiennych i wystąpień, a także wyliczenie jaki stanowiły one odsetek w analizowanej próbie. W celu weryfikacji czy te wystąpienia są tendencyjne lub przypadkowe wykonano test Chi Kwadrat dla proporcji (Pearson, 1900). W celu wyliczenia częstości i istotności występowania poszczególnych wartości zmiennych Płeć, Edukacja, Wiek wykonano serię 3 analiz częstości testem Chi Kwadrat. Braki danych wykluczono z analizy - obserwacje z brakami pomijano odrębnie dla każdej analizowanej zmiennej.

Analiza zmiennej *Płeć* wykazała, że liczebność jej poszczególnych poziomów kształtowała się następująco: Kobieta: $n = 173$ (57.10%); Mężczyzna: $n = 130$ (42.90%), co łącznie dało $N = 303$ wystąpień (100%). Zaobserwowane różnice w liczebnościach kategorii były istotne statystycznie $\chi^2(1) = 6.10$; $p = 0.014$. Zróznicowanie analizowanych wartości było słabe, miara siły efektu ϕ (Phi) wyniosła wartość $\phi = 0.14$. W rozkładzie zdecydowanie przeważała kategoria „Kobieta” (57.10%), najrzadziej zaś występowała kategoria „Mężczyzna” (42.90%).

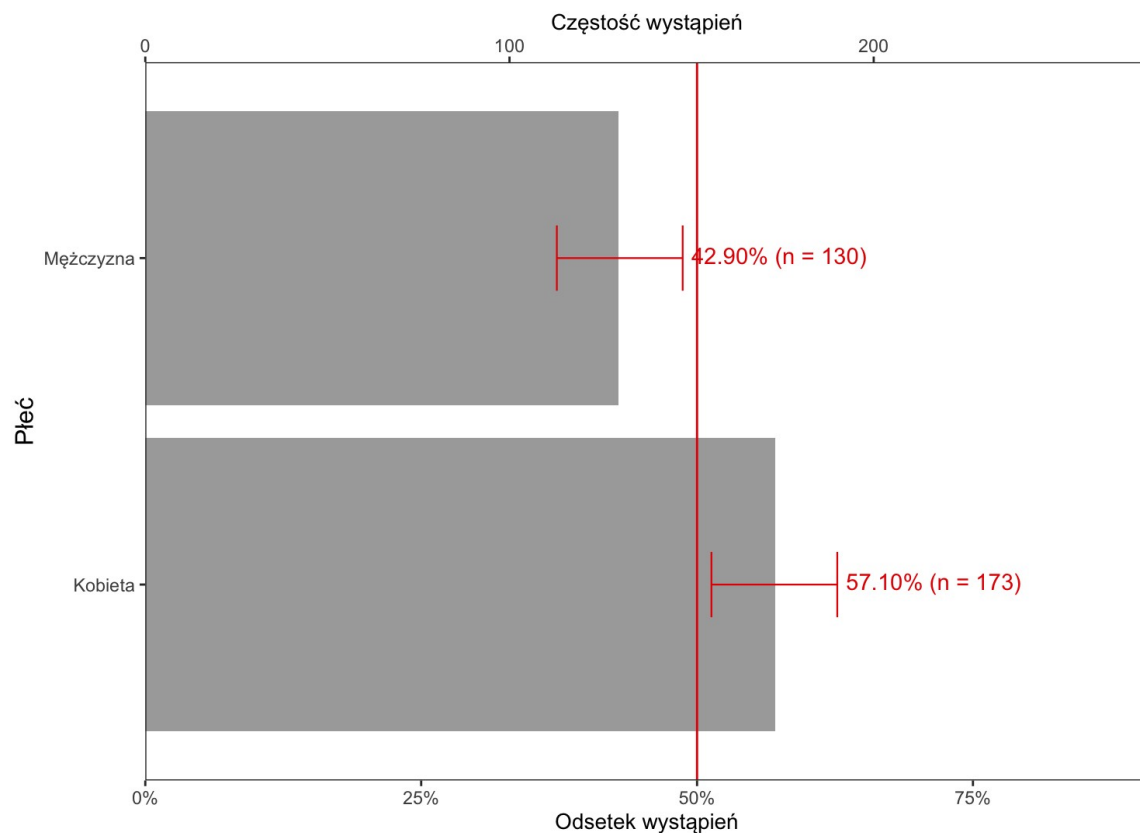
W przypadku zmiennej *Edukacja* stwierdzono, że liczebność jej poszczególnych kategorii kształtowała się następująco: Wyższe: $n = 171$ (56.44%); Średnie: $n = 130$ (42.90%); Podstawowe: $n = 2$ (0.66%), co łącznie dało $N = 303$ wystąpień (100%). Różnice te okazały się istotne statystycznie $\chi^2(2) = 153.88$; $p < 0.001$. Zróznicowanie analizowanych wartości było silne, miara siły efektu V Craméra wyniosła wartość $V = 0.50$. W rozkładzie zdecydowanie przeważała kategoria „Wyższe” (56.44%), najrzadziej zaś występowała kategoria „Podstawowe” (0.66%). Powyżej wartości oczekiwanej (wynoszącej $n = 101.00$) znalazły się kategorie: Średnie, Wyższe, natomiast poniżej niej znalazła się kategoria: Podstawowe.

Rozkład zmiennej *Wiek* pokazał, że liczebność jej poszczególnych wariantów kształtowała się następująco: Od 20 do 29 lat: $n = 225$ (74.26%); Od 30 do 39 lat: $n = 59$ (19.47%); Od 40 do 51 lat: $n = 19$ (6.27%), co łącznie dało $N = 303$ wystąpień (100%). Analiza testem chi-kwadrat wykazała, że różnice między kategoriami były istotne statystycznie $\chi^2(2) = 236.28$; $p < 0.001$.

Zróźnicowanie analizowanych wartości było silne, miara siły efektu V Craméra wyniosła wartość $V = 0.62$. W rozkładzie zdecydowanie przeważała kategoria „Od 20 do 29 lat” (74.26%), najrzadziej zaś występowała kategoria „Od 40 do 51 lat” (6.27%). Powyżej wartości oczekiwanej (wynoszącej $n = 101.00$) znalazła się kategoria: Od 20 do 29 lat, natomiast poniżej niej znalazły się kategorie: Od 40 do 51 lat, Od 30 do 39 lat. Wyniki analiz przedstawia seria rysunków od nr 1 do nr 3.

Wykres nr 1

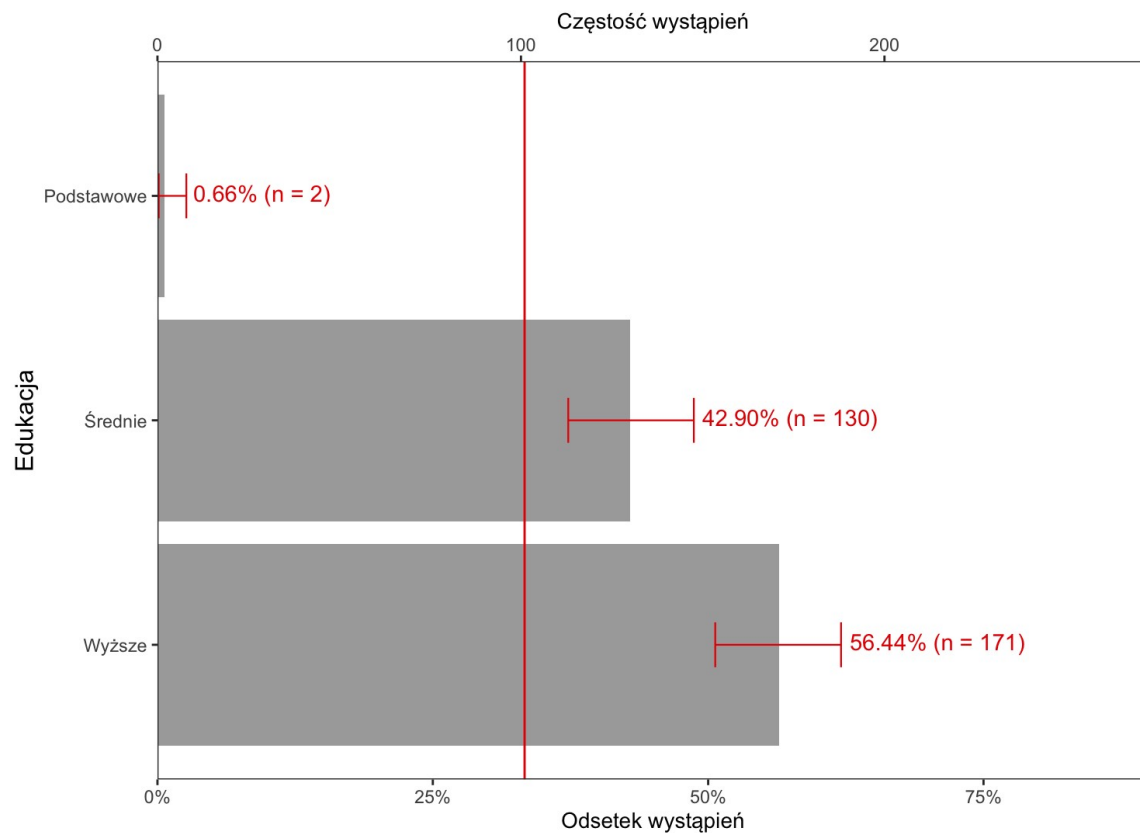
Częstość i odsetek występowania wartości zmiennej Płeć



Nota. Zróźnicowanie poziomów zmiennej Płeć było istotnie dalekie od wartości oczekiwanej wynoszącej 50.00% (czerwona linia). Wynik oszacowania testu Chi Kwadrat wyniósł: $\chi^2(1) = 6.10$; $p = 0.014$. Ciągła linia — oznacza istotne zróźnicowanie wartości zmiennej Płeć. Czerwone wąsy przy słupkach przedstawiają 95% przedział ufności dla odsetka wystąpień danej kategorii. Nakładanie się wąsów dwóch kategorii oznacza brak istotnych różnic między nimi, natomiast ich rozłączość świadczy o istotnych statystycznie różnicach.

Wykres nr 2

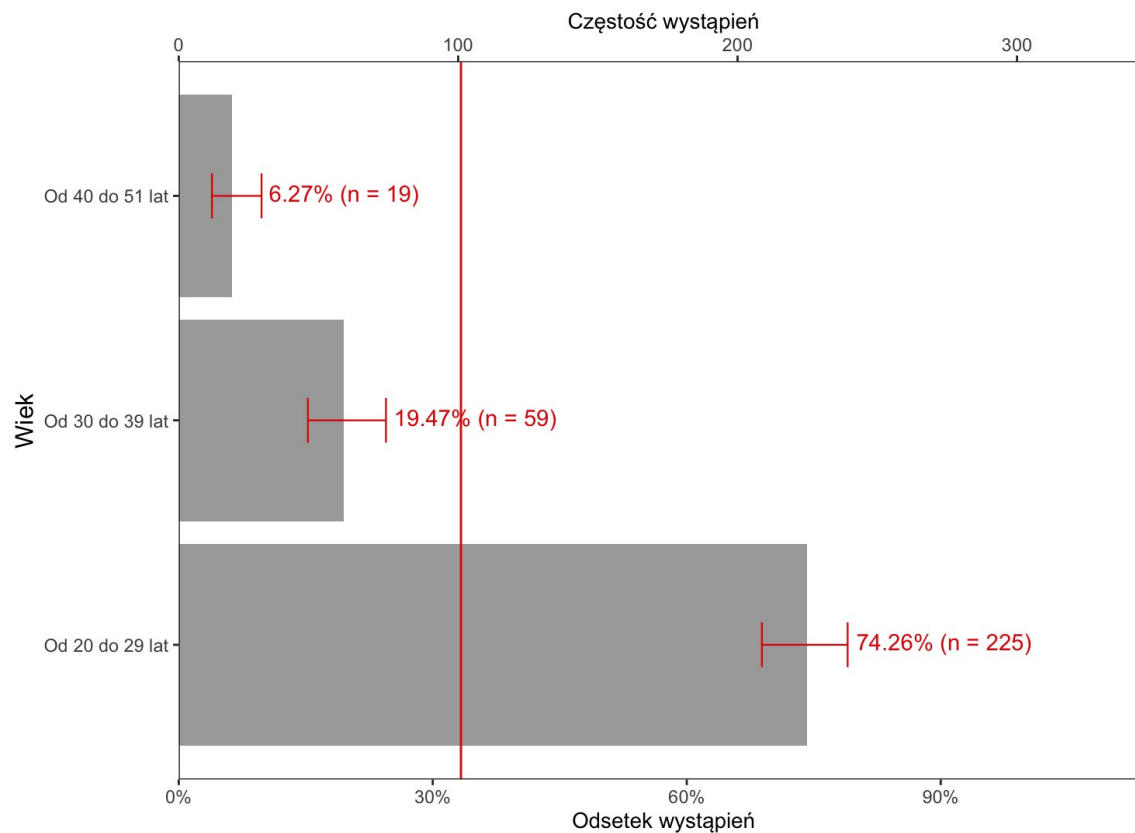
Częstość i odsetek występowania wartości zmiennej Edukacja



Nota. Zróżnicowanie poziomów zmiennej Edukacja było istotnie dalekie od wartości oczekiwanej wynoszącej 33.33% (czerwona linia). Wynik oszacowania testu Chi Kwadrat wyniósł: $\chi^2(2) = 153.88$; $p < 0.001$. Ciągła linia — oznacza istotne zróżnicowanie wartości zmiennej Edukacja. Czerwone wąsy przy słupkach przedstawiają 95% przedział ufności dla odsetka wystąpień danej kategorii. Nakładanie się wąsów dwóch kategorii oznacza brak istotnych różnic między nimi, natomiast ich rozłączość świadczy o istotnych statystycznie różnicach.

Wykres nr 3

Częstość i odsetek występowania wartości zmiennej Wiek



Nota. Zróżnicowanie poziomów zmiennej Wiek było istotnie dalekie od wartości oczekiwanej wynoszącej 33.33% (czerwona linia). Wynik oszacowania testu Chi Kwadrat wyniósł: $\chi^2(2) = 236.28$; $p < 0.001$. Ciągła linia — oznacza istotne zróżnicowanie wartości zmiennej Wiek. Czerwone wąsy przy słupkach przedstawiają 95% przedział ufności dla odsetka wystąpień danej kategorii. Nakładanie się wąsów dwóch kategorii oznacza brak istotnych różnic między nimi, natomiast ich rozłączość świadczy o istotnych statystycznie różnicach.

Podsumowanie wyników

Przeprowadzono serię trzech analiz częstości przy użyciu testu Chi Kwadrat dla proporcji, celem opisanego rozkładów zmiennych Płeć, Edukacja i Wiek oraz oceny, czy ich wystąpienia odbiegały od rozkładu losowego. Badana próba była nieco zdominowana przez kobiety, najczęściej wykształcenie uczestników było wyższe, a ich wiek w zdecydowanej większości był w przedziale od 20 do 29 lat.

Bibliografia

Hryniewicz, K., Milewska, A. (2023). SZTOS: System Zautomatyzowanego Tworzenia Opisu Statystycznego (Wersja SZTOS) [Oprogramowanie]. <https://sztos-it.com/>

Karl Pearson F.R.S. (1900) X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling , The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50:302, 157-175, DOI: 10.1080/14786440009463897

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis . Springer-Verlag New York. ISBN 978-3-319-24277-4