



# **Binarny model regresji na przykładzie determinant szczęścia w Szwajcarii**



Przedmiot: Regresja logistyczna z wykorzystaniem narzędzi SAS

Warszawa, 2023

## Spis treści

<b>1. Cel i zakres analizy.....</b>	<b>3</b>
1.1 Wstęp .....	3
1.2 Uzasadnienie wyboru zmiennych objaśniających.....	3
1.3 Hipotezy badawcze .....	5
<b>2. Prezentacja zbioru i przygotowanie danych do analizy .....</b>	<b>6</b>
2.1 Prezentacja zbioru danych .....	6
2.2 Przygotowanie danych do analizy .....	8
2.3 Rozkłady zmiennych objaśniających .....	9
2.4 Kodowanie zmiennych .....	16
<b>3. Model regresji logistycznej dla Szwajcarii.....</b>	<b>17</b>
3.1 Opis modelu .....	17
3.2 Ocena jakości modelu binarnego .....	18
3.3 Interpretacja parametrów modelu binarnego.....	21
3.4 Kodowanie odniesienia .....	22
<b>4. Model regresji logistycznej z efektami głównymi oraz interakcjami par zmiennych .....</b>	<b>24</b>
4.1 Proces regresji logistycznej z efektami głównymi przy wykorzystaniu metody stepwise .....	26
<b>5. Podsumowanie .....</b>	<b>34</b>
<b>6. Fragmenty Kodu SAS.....</b>	<b>35</b>
<b>7. Bibliografia.....</b>	<b>45</b>

# 1. Cel i zakres analizy

## 1.1 Wstęp

Celem projektu jest przeprowadzenie analizy regresji logistycznej w celu zidentyfikowania determinant szczęścia dla Szwajcarii na podstawie danych zebranych w rundzie dziesiątej Europejskiego Badania Społecznego (European Social Survey) dostępnych na stronie [www.europeansocialsurvey.org](http://www.europeansocialsurvey.org). ESS jest rozbudowanym projektem badawczym, którego zamiarem jest analiza postaw, opinii i zachowań społecznych w różnych krajach europejskich. Dane zgromadzone w ramach badania obejmują szeroki zakres tematyczny i odzwierciedlają różne aspekty życia społecznego. W projekcie skupiliśmy się na analizie tych czynników które mogą wpływać na poziom szczęścia respondentów. Przy wyborze kierowaliśmy się literaturą, szeregiem badań a także własnymi doświadczeniami empirycznymi.

W ramach projektu została przeprowadzona analiza regresji logistycznej - technika statystyczna wykorzystywana do modelowania zależności między zmienną zależną (w tym przypadku poziomem szczęścia) a zmiennymi niezależnymi (potencjalnymi determinantami). Analiza ta pozwoli na ocenę wpływu różnych czynników na prawdopodobieństwo wystąpienia szczęścia lub braku szczęścia wśród obywateli Szwajcarii.

Wyniki projektu mogą dostarczyć informacji na temat istotnych czynników, które determinują poziom szczęścia wśród mieszkańców oraz wskazać, które zmienne mają największy wpływ na badane zjawisko. Analiza danych pod tym kątem jest nie tylko ciekawa sama w sobie, ale również może okazać się użyteczna pod kątem decyzji politycznych a także być wartościowym źródłem wiedzy dla organizacji społecznych i innych zainteresowanych stron. Zebrane dane i wyniki analizy mogą być pomocne w podejmowaniu działań mających na celu poprawę jakości życia i dobrostanu społecznego mieszkańców Szwajcarii.

## 1.2 Uzasadnienie wyboru zmiennych objaśniających

Wybór zmiennych na potrzeby opracowania analizy został wsparty przeglądem literatury z zakresu psychologii oraz ekonomii uwzględniający czynniki moderujące poczucie szczęścia. Poniżej zostaje przedstawiona lista wybranych zmiennych wraz z krótkim opisem badań.

- sclmeet - Intensywność spotkań ze znajomymi/krewnymi

Według badań prowadzonych przez zespół naukowców z Włoch (Amati i inni, 2018), relacje z innymi ludźmi, takie jak znajomości czy przyjaźnie, są istotnym elementem odczuwanego szczęścia. Badanie wykazało, że osoby, które utrzymują bliskie relacje i często spotykają się ze znajomymi, wyrażają wyższy poziom zadowolenia z życia. To pokazuje, jak ważne jest budowanie i utrzymanie tzw. kapitału społecznego.

- sclact - Udział w aktywnościach społecznych w porównaniu z innymi osobami w tym samym wieku

Badania przeprowadzone na różnych grupach pokazują, że aktywność społeczna ma pozytywny wpływ na poczucie szczęścia:

- U osób starszych, które są aktywne społecznie, zauważalny jest wyższy poziom zadowolenia z życia, zwłaszcza u tych, którzy są ekstrawertykami (Oerlemans i inni, 2011).
- U osób z niepełnosprawnością fizyczną, aktywność społeczna stanowi strategię na zwiększenie poczucia szczęścia. Stąd wynika, że opiekunowie powinni oferować różnorodne formy aktywności społecznych, takie jak programy społecznościowe online, lokalne programy społeczne i wydarzenia społeczne (Kim i inni, 2021).

- health - Stan zdrowia

W literaturze naukowej znajduje się wiele badań na temat relacji między oceną własnego zdrowia a poczuciem szczęścia. Zgodnie z tymi badaniami, ocena własnego zdrowia ma istotny wpływ na zadowolenie z życia. W szczególności, starsi ludzie, którzy mają niską ocenę swojego zdrowia, często są mniej zadowoleni z życia (Dumitrache i inni, 2017).

- rlgblg - Przynależność do określonej religii lub wyznania

Z analizy wielu badań na temat związku między religijnością a poczuciem szczęścia wynika, że istnieje pewien związek między tymi dwiema zmiennymi (Rizvi i Hossain, 2016). Szczególnie interesujące są wyniki pokazujące, że najszczęśliwsi ludzie i najbardziej religijne społeczeństwa znajdują się na Bliskim Wschodzie.

- dscrrce - Dyskryminacja grupy respondenta: kolor skóry lub rasa

Dyskryminacja rasowa może mieć negatywny wpływ na zadowolenie z życia i poczucie szczęścia. W badaniu przeprowadzonym przez Williams i Williams-Morris w 2000 roku stwierdzono, że codzienne doświadczenia z rasizmem mogą prowadzić do chronicznego stresu, który z kolei negatywnie wpływa na zdrowie psychiczne i poczucie szczęścia.

- gndr - Płeć

Poczucie szczęścia może różnić się w zależności od płci. Na przykład, badania wykazały, że kobiety częściej niż mężczyźni czerpią szczęście i satysfakcję z życia z istotnych doświadczeń, podczas gdy mężczyźni częściej niż kobiety czerpią szczęście i satysfakcję z przyjemnych doświadczeń (Brakus, 2022).

- agea - Wiek

Psychologia i ekonomia sugerują, że zadowolenie z życia zmienia się w ciągu życia na zasadzie krzywej w kształcie litery U. Chociaż jest to trend obserwowany na poziomie populacji, to niekoniecznie odnosi się do każdej jednostki. Indywidualne doświadczenia i sytuacje mogą istotnie wpływać na zadowolenie z życia w różnych etapach życia (Blanchflower i Oswald, 2008).

- hincfel - Obecne odczucia dotyczące dochodów gospodarstwa domowego

Przegląd literatury sugeruje, że jest związek między bogactwem a poczuciem szczęścia na poziomie narodowym. Pomimo gospodarczego rozwoju w ostatnich dziesięcioleciach w większości rozwiniętych społeczeństwach, wzrost ogólnego poczucia szczęścia jest ograniczony. Ponadto, ludzie, którzy dążą do celów materialnych, są generalnie mniej szczęśliwi, chyba że są zamożni (Diener i Biswas-Diener, 2002).

- ipgdtim - Ważne, by dobrze się bawić

Można przypuszczać, że ludzie, którzy dążą do dobrej zabawy, są szczęśliwsi. Badanie na temat dążenia do przyjemności i komfortu (Huta i Richard, 2010) potwierdza tę hipotezę. Osoby, które aktywnie dążą do dobrej zabawy, a ich życie jest wypełnione pozytywnymi doświadczeniami, wykazują wyższy poziom szczęścia.

Dodatkowe zmienne użyte w analizie zostały dobrane kierując się intuicją oraz doświadczeniem autorów. Ich nazwy i opisy są widoczne w tabeli 1 w rozdziale 2.

### 1.3 Hipotezy badawcze

Główna hipoteza badawcza opiera się na założeniu, iż aktywność społeczna Szwajcarów w postaci intensywności spotkań ze znajomymi lub krewnymi, udziału w aktywnościach społecznych, liczby osób z którymi mogą rozmawiać o sprawach intymnych i osobistych, czy też istotność dobrej zabawy wpływa na ich poziom szczęścia – im wyższa aktywność, tym respondent szczęśliwszy. Kolejna hipoteza, która została zbadana to opinia twierdząca, iż im lepsza sytuacja polityczno-społeczna państwa tym szczęśliwszy jest jego obywatel. Do czynników charakteryzujących takową sytuację mogą należeć badany w projekcie wpływ satysfakcji z sytuacji ekonomicznej oraz ocena stanu usług zdrowotnych w kraju. Ponadto, na komfort życia w danym kraju może również wpływać poziom zaufania dla służb porządkowych (tj. policji), a także poczucie bezpieczeństwa podczas samotnych spacerów w okolicy po zmroku.

Powszechnie uważa się, że stopień wykształcenia oraz teoretycznie idące za tym zarobki pozytywnie wpływają na poziom szczęścia człowieka. Dlatego też w ramach analizy wzięto również zmienne określające najwyższy poziom osiągniętej edukacji, ukończoną edukację w latach oraz obecne odczucia dotyczące dochodów gospodarstwa domowego. Dodatkowo wzięto pod uwagę aspekty, które również

mogą bezpośrednio wpływać na poziom szczęścia takie jak częstość korzystania z internetu, ogólny stan zdrowia, czy też czy respondent przynależy do określonej religii lub wyznania.

## 2. Prezentacja zbioru i przygotowanie danych do analizy

### 2.1 Prezentacja zbioru danych

Na potrzeby niniejszej analizy posłużono się danymi dotyczącymi Szwajcarii pochodzącymi z dziesiątej rundy badania ESS z 2020 roku. Zbiór danych służący jako podstawa analizy obejmuje łącznie 19 zmiennych oraz 1 523 obserwacji. Zmienną objaśnianą jest zmienna HAPPY, która opisuje poziom szczęścia respondenta.

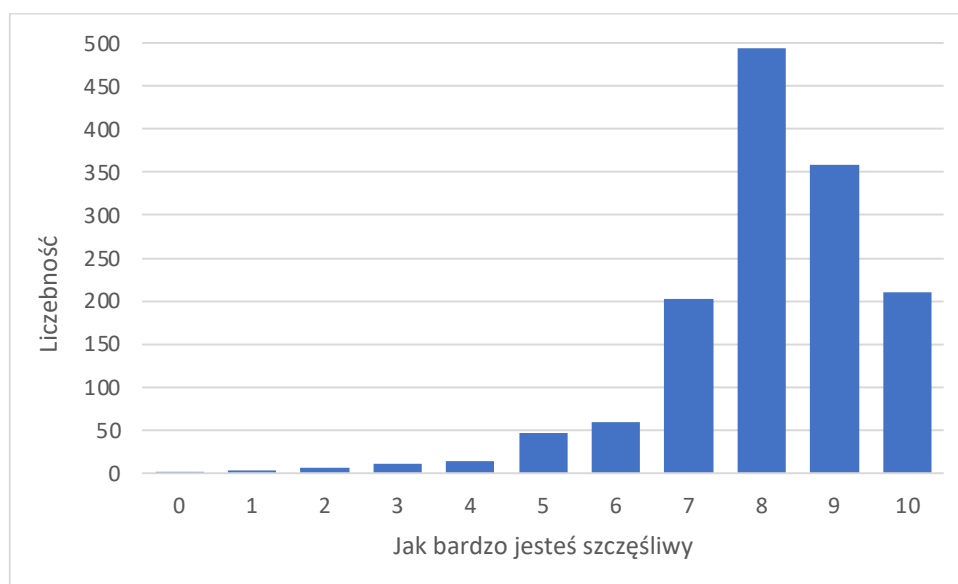
ZMIENNA	OPIS	WARTOŚĆ
happy	Czy ankietowany jest szczęśliwy	Zmienna porządkowa – 11 kategorii
netusoft	Częstość korzystania z internetu	Zmienna porządkowa – 11 kategorii
trstplc	Poziom ufności policji	Zmienna porządkowa – 11 kategorii
stfeco	Satysfakcja z sytuacji ekonomicznej w kraju	Zmienna porządkowa – 11 kategorii
stfhlth	Obecny stan usług zdrowotnych w kraju	Zmienna porządkowa – 7 kategorii
sclmeet	Intensywność spotkań ze znajomymi/krewnymi	Zmienna porządkowa – 7 kategorii
inprdsc	Liczba osób, z którymi respondent może rozmawiać o sprawach intymnych i osobistych	Zmienna porządkowa – 11 kategorii
sclact	Udział w aktywnościach społecznych w porównaniu z innymi osobami w tym samym wieku	Zmienna porządkowa – 5 kategorii
aesfdrk	Poczucie bezpieczeństwa podczas samotnych spacerów w okolicy po zmroku	Zmienna porządkowa – 4 kategorii
health	Stan zdrowia	Zmienna porządkowa – 5 kategorii
rlgblg	Przynależność do określonej religii lub wyznania	Zmienna binarna
dscrce	Dyskryminacja grupy respondenta: kolor skóry lub rasa	Zmienna binarna

gndr	Płeć	Zmienna binarna
agea	Wiek	Zmienna ilościowa
eised	Najwyższy poziom wykształcenia, ES - ISCED	Zmienna porządkowa – 7 kategorii
eduysr	Ukończone lata edukacji w pełnym wymiarze godzin	Zmienna ilościowa
hincfel	Obecne odczucia dotyczące dochodów gospodarstwa domowego	Zmienna porządkowa – 4 kategorii
livpnt	Czy rodzice wciąż żyją	Zmienna porządkowa – 4 kategorii
ipgdtim	Ważne, by dobrze się bawić	Zmienna porządkowa – 6 kategorii

*Tabela 1 Zmienne wykorzystane do analizy*

Źródło: Opracowanie własne na podstawie <https://www.europeansocialsurvey.org/data/>

Do przeprowadzenia analizy wybrano 14 zmiennych porządkowych, 3 zmienne binarne oraz 2 zmienne ilościowe. W celu poprawnego wykorzystania zbioru usunięto wszelkie braki danych oraz wartości zmiennych, które nie miały znaczenia w badaniu (np. brak odpowiedzi respondentów) – pozostało 1 407 obserwacji. Poniżej przedstawiono rozkład zmiennej happy, dla której średnia oraz mediana wynosi 8.



*Wykres 1 Rozkład zmiennej happy*

Źródło: Opracowanie własne

happy	Liczebność	Udział	Liczebność skumulowana	Procent skumulowany
-------	------------	--------	------------------------	---------------------

0	2	0,14%	2	0,14%
1	3	0,21%	5	0,36%
2	6	0,43%	11	0,78%
3	11	0,78%	22	1,56%
4	14	1,00%	36	2,56%
5	47	3,34%	83	5,90%
6	59	4,19%	142	10,09%
7	203	14,43%	345	24,52%
8	494	35,11%	839	59,63%
9	358	25,44%	1197	85,07%
10	210	14,93%	1407	100,00%

*Tabela 2 Częstość zmiennej happy*

*Źródło: Opracowanie własne*

## 2.2 Przygotowanie danych do analizy

Zmienna happy, w celu uproszczenia modelu oraz ze względu na licznosc niektórych jej kategorii, została podzielona na dwie następujące kategorie: Nieszczęśliwy (dla kategorii od 0 do 8) oraz Szczęśliwy (dla kategorii od 9 do 10). Przypisano im wartości kolejno 0 oraz 1. Podział ten również wynika z konieczności zbliżonych licznosci dla 0 i 1. Dokonano również podziału pozostałych zmiennych, co zbiorczo zostało przedstawione w tabeli poniżej.

ZMIENNA	NAZWA	KATEGORIE
happy	szczescie	0-8 -> Nieszczęśliwy/-a (0); 9-10 -> Szczęśliwy/-a (1)
netusoft	internet_czestosc	1 -> Nigdy; 2, 3 -> Czasami; 4, 5 -> Często
trstplc	policja_zaufanie	0-3 -> Brak zaufania; 4-7 -> Przeciętne zaufanie; 8-10 -> Wysokie zaufanie
stfecoc	ekonomia_satysfakcja	0-3 -> Niska satysfakcja; 4-7 -> Przeciętna satysfakcja; 8-10 -> Wysoka satysfakcja
stfhlth	uslugi_zdowotne_stan	0-3 -> Zły stan; 4-7 -> Przeciętny stan; 8-10 -> Bardzo dobry stan

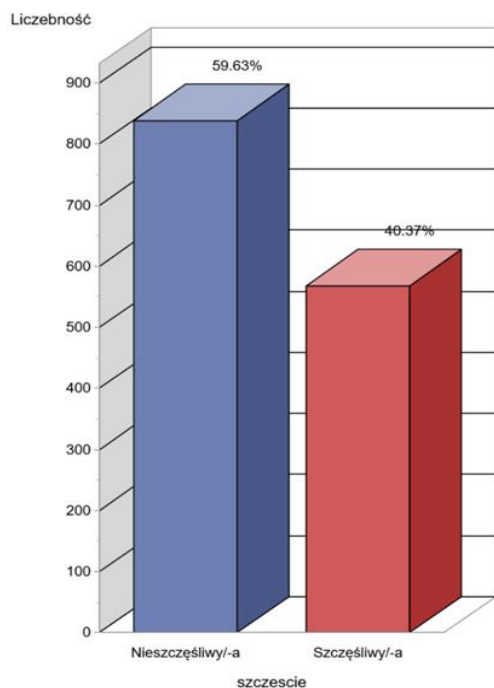


sclmeet	spotkania	1, 2 -> Rzadko; 3, 4 -> Czasami; 5, 6, 7 -> Często
inprdsc	osoby_rozmowy	0 -> Zero osób; 1-3 -> Od 1 do 3 osób; 4, 5 -> Od 4 do 9 osób; 6 -> 10 lub więcej osób
sclact	akt_spol	1,2 -> Mniejszy (niż innych osób); 3 -> Taki sam (jak innych osób); 4, 5 -> Większy (niż innych osób)
aesfdrk	spacer_zmrok	1, 2 -> Wysokie (poczucie bezpieczeństwa); 3, 4 -> Niskie (poczucie bezpieczeństwa)
health	zdrowie	1-2 -> Dobry stan zdrowia; 3 -> Przeciętny stan zdrowia; 4,5 -> Zły stan zdrowia
rlgblg	religia	1 -> Tak, przynależę; 2 -> Nie przynależę
dscrrce	dyskryminacja	1 -> Tak, jestem/byłem/-am dyskryminowany/-a; 0 -> Nie jestem/byłem/-am dyskryminowany/-a
gndr	plec	1 -> Mężczyzna; 2 -> Kobieta
agea	wiek	<= 38 - Młody wiek; 39 - 60 -> Średni wiek; >= 60 -> Stary wiek
eisced	edukacja	1-4 -> Niskie wykształcenie (szkoła średnia i poniżej); 5 -> Średnie wykształcenie (zaawansowane zawodowe, podyplomowe); 6, 7 -> Wyższe wykształcenie (licencjat, magister)
eduyrs	edukacja_lata	<= 12 - Wykształcenie podstawowe; > 12 - Wykształcenie wyższe
hincfel	dochody_odczucie	1 -> Wygodne życie (przy obecnych dochodach); 2 -> Radzę sobie (przy obecnych dochodach); 3, 4 -> Ciężkie życie (przy obecnych dochodach)
livpnt	rodzice	1 -> Tak, oboje; 2, 3 -> Jedno żyje; 4 -> Nie żyją
ipgdtim	dobra_zabawa	1, 2 -> Istotne; 3, 4 -> Średnio istotne; 5, 6 -> Nieistotne

*Tabela 3 Kategoryzacja zmiennych w analizie*

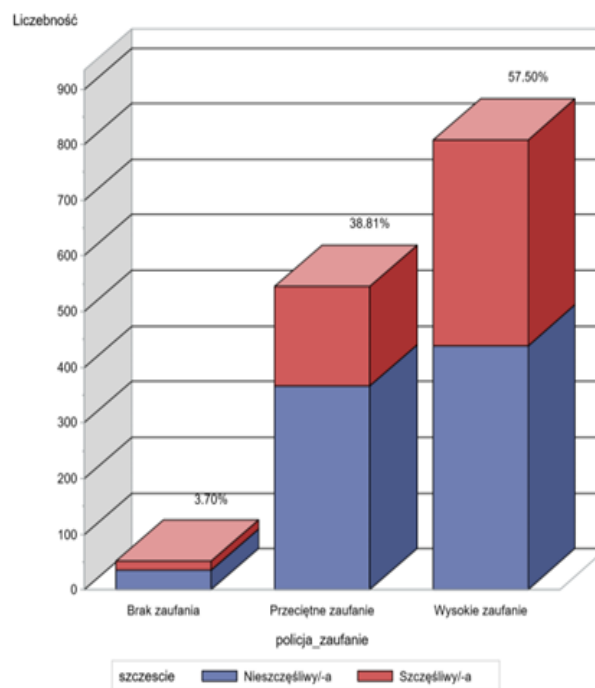
*Źródło: Opracowanie własne*

## 2.3 Rozkłady zmiennych objaśniających



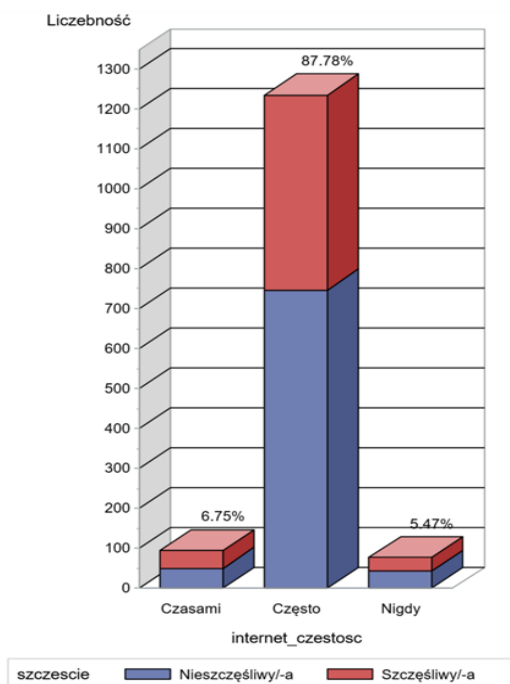
Wykres 2 Zmienna celu

Źródło: Opracowanie własne przy użyciu narzędzi SAS



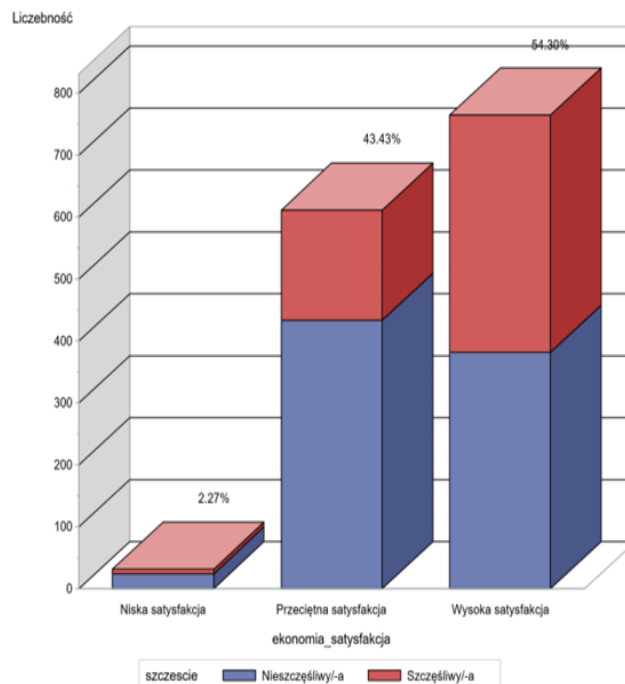
Wykres 4 Zmienna policja\_zaufanie

Źródło: Opracowanie własne przy użyciu narzędzi SAS



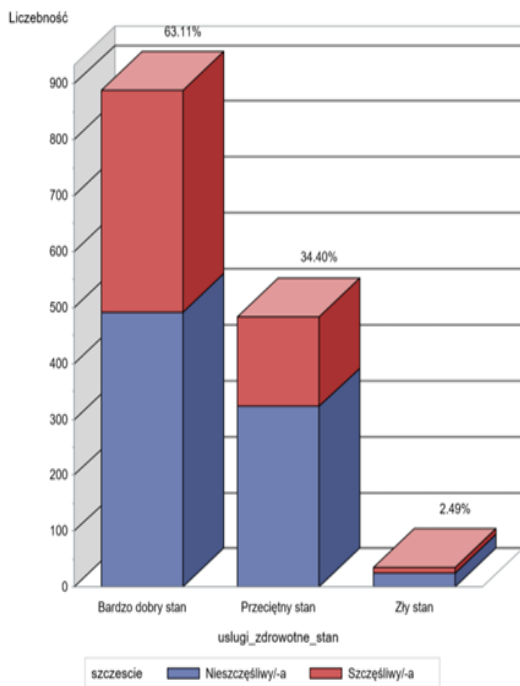
Wykres 3 Zmienna internet\_czestosc

Źródło: Opracowanie własne przy użyciu narzędzi SAS



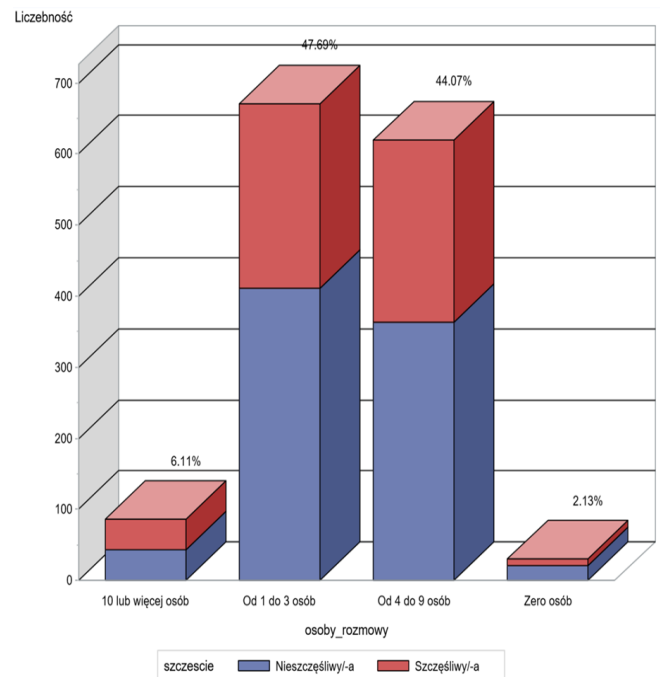
Wykres 5 Zmienna ekonomia\_satysfakcja

Źródło: Opracowanie własne przy użyciu narzędzi SAS



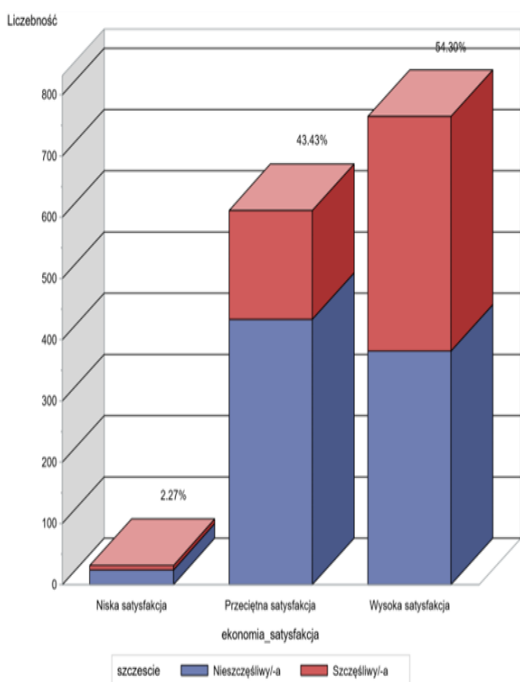
Wykres 6 Zmienna usługi\_zdrowotne\_stan

Źródło: Opracowanie własne przy użyciu narzędzi SAS



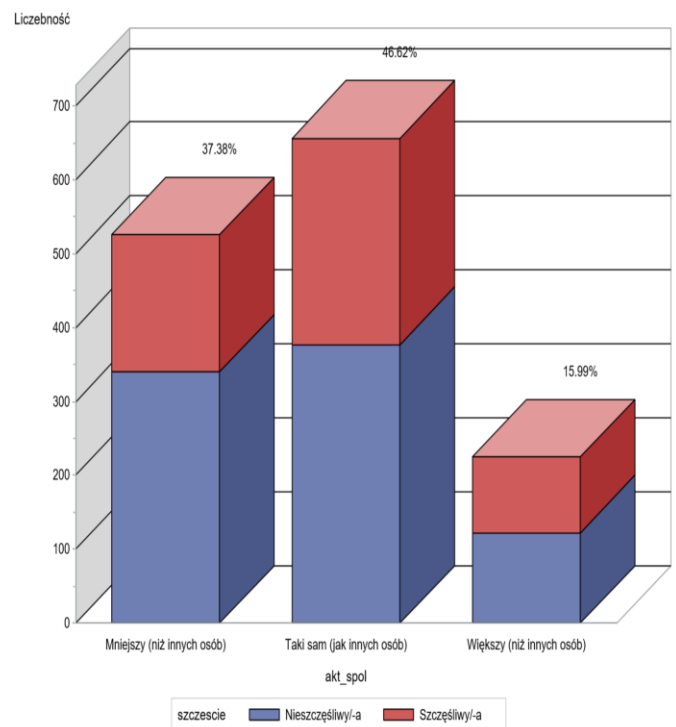
Wykres 8 Zmienna osoby\_rozmowy

Źródło: Opracowanie własne przy użyciu narzędzi SAS



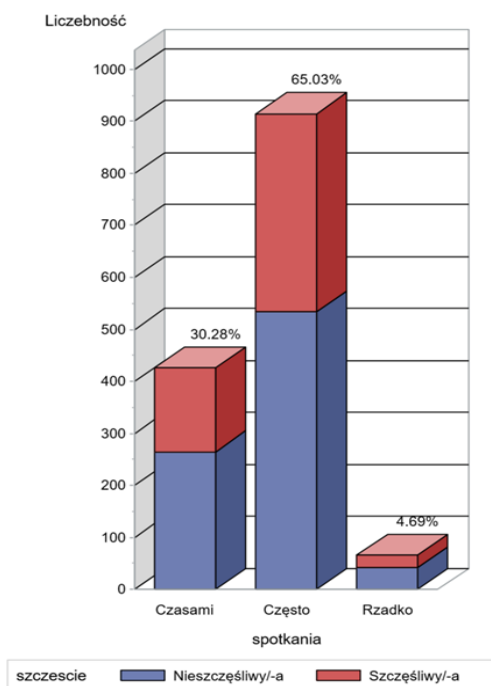
Wykres 7 Zmienna ekonomia\_satisfakcja

Źródło: Opracowanie własne przy użyciu narzędzi SAS



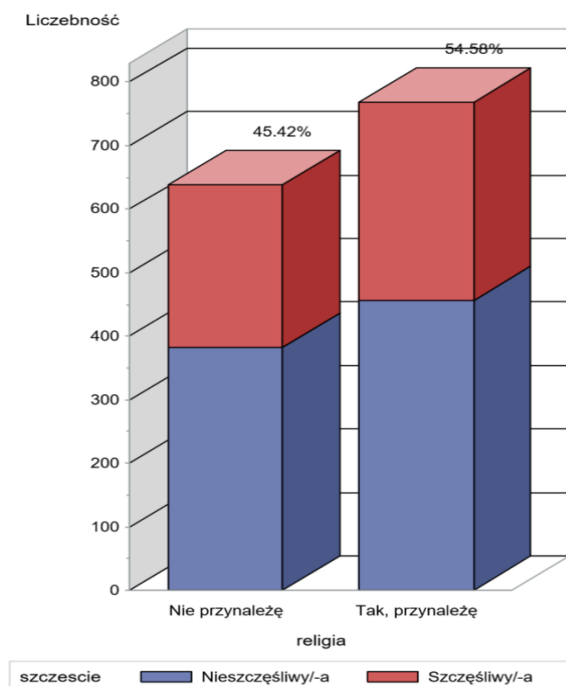
Wykres 9 Zmienna akt\_spol

Źródło: Opracowanie własne przy użyciu narzędzi SAS



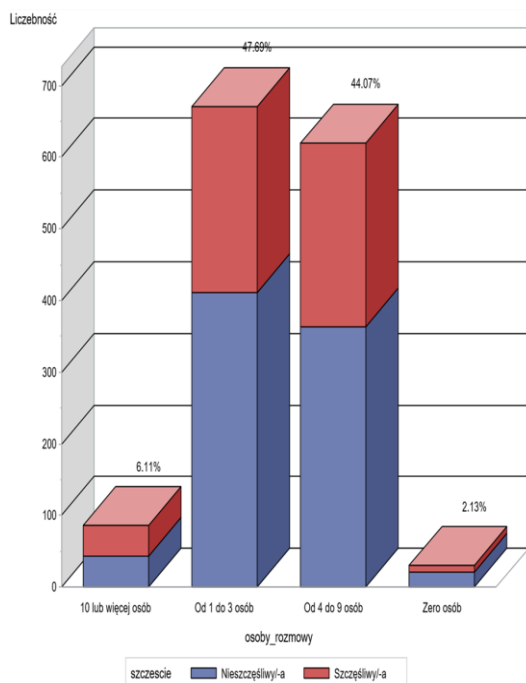
Wykres 10 Zmienna spotkania

Źródło: Opracowanie własne przy użyciu narzędzi SAS



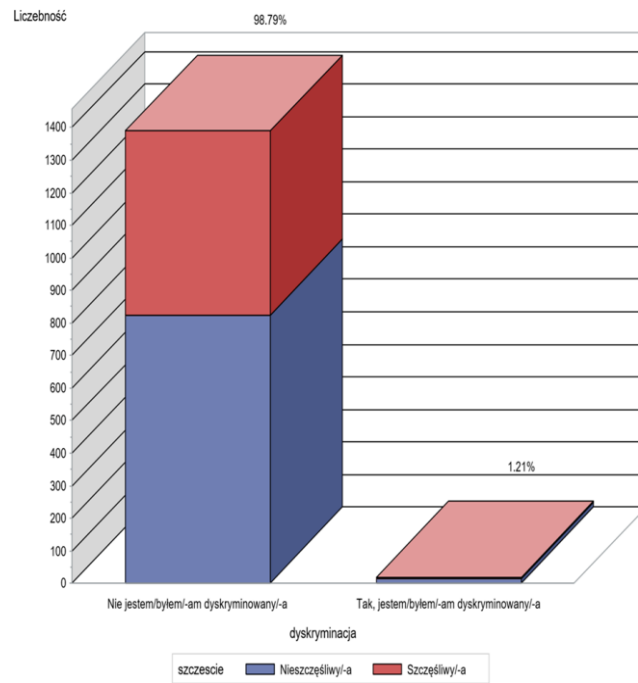
Wykres 12 Zmienna religia

Źródło: Opracowanie własne przy użyciu narzędzi SAS



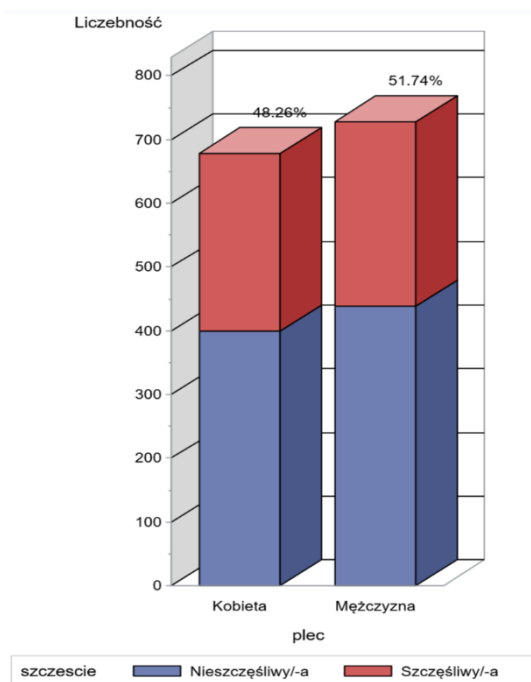
Wykres 11 Zmienna osoby\_rozmowy

Źródło: Opracowanie własne przy użyciu narzędzi SAS



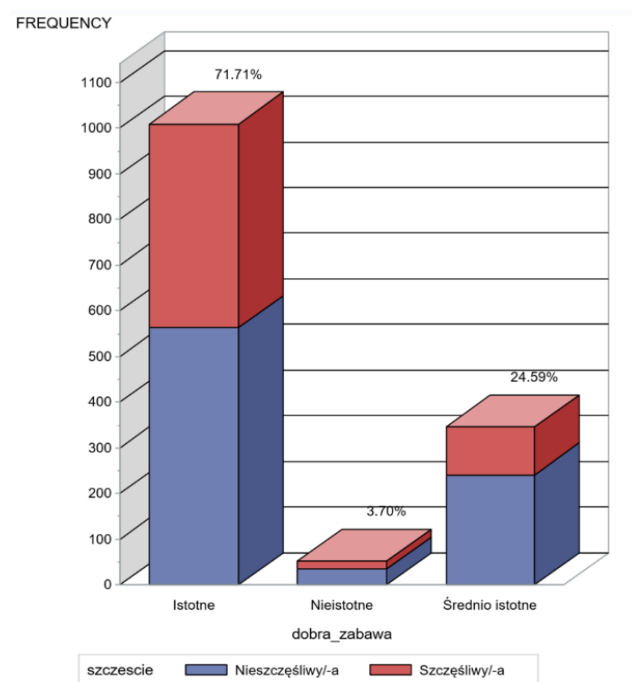
Wykres 13 Zmienna dyskryminacja

Źródło: Opracowanie własne przy użyciu narzędzi SAS



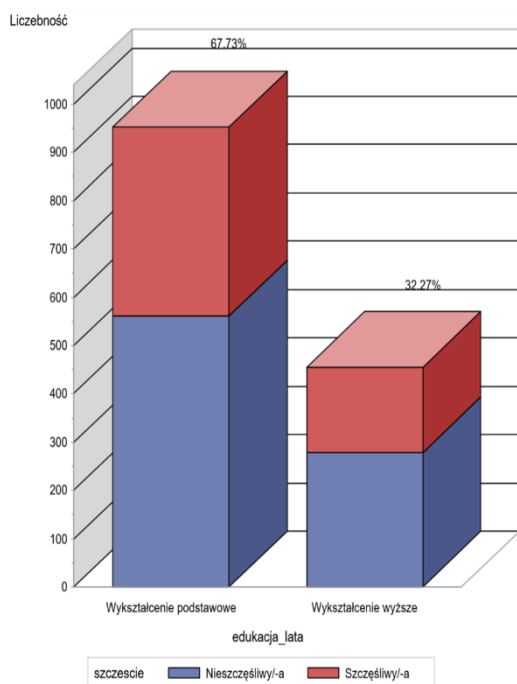
Wykres 14 Zmienna plec

Źródło: Opracowanie własne przy użyciu narzędzi SAS



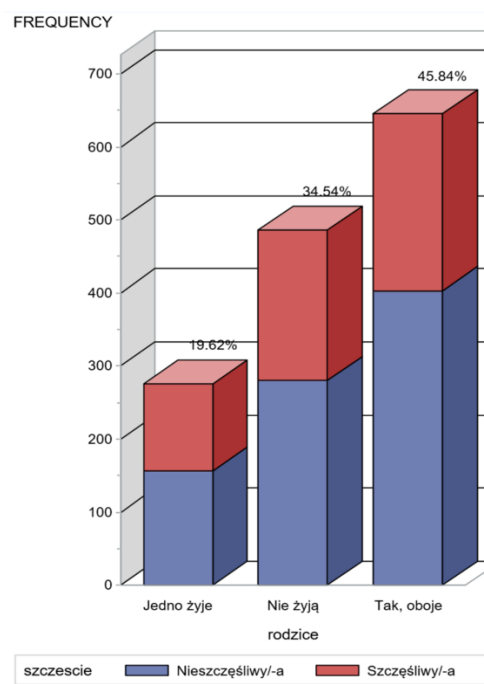
Wykres 16 Zmienna dobra\_zabawa

Źródło: Opracowanie własne przy użyciu narzędzi SAS



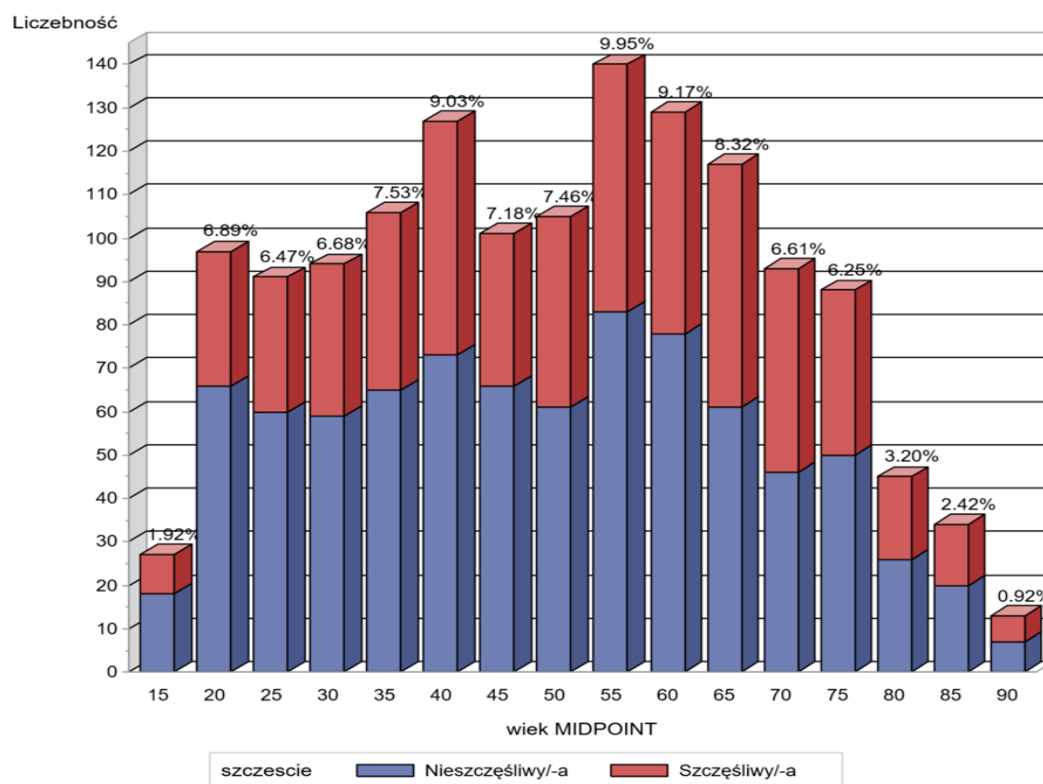
Wykres 15 Zmienna edukacja\_lata

Źródło: Opracowanie własne przy użyciu narzędzi SAS

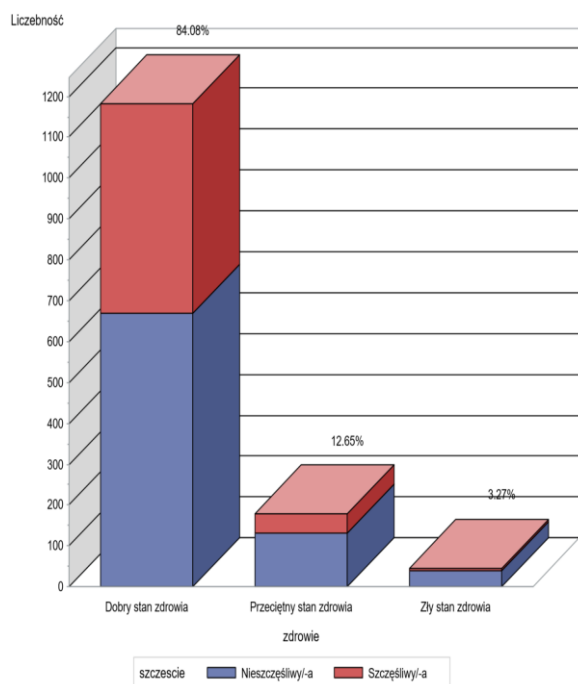


Wykres 17 Zmienna rodzic

Źródło: Opracowanie własne przy użyciu narzędzi SAS

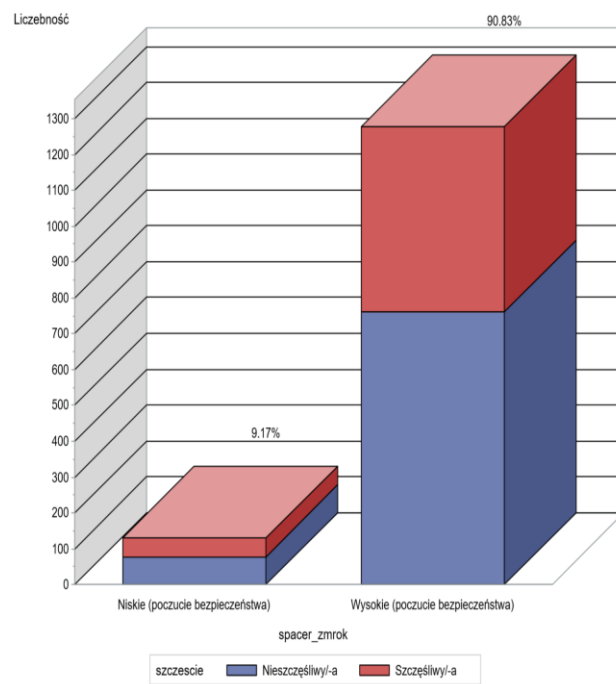


Wykres 18 Zmienna wiek  
 Źródło: Opracowanie własne przy użyciu narzędzi SAS



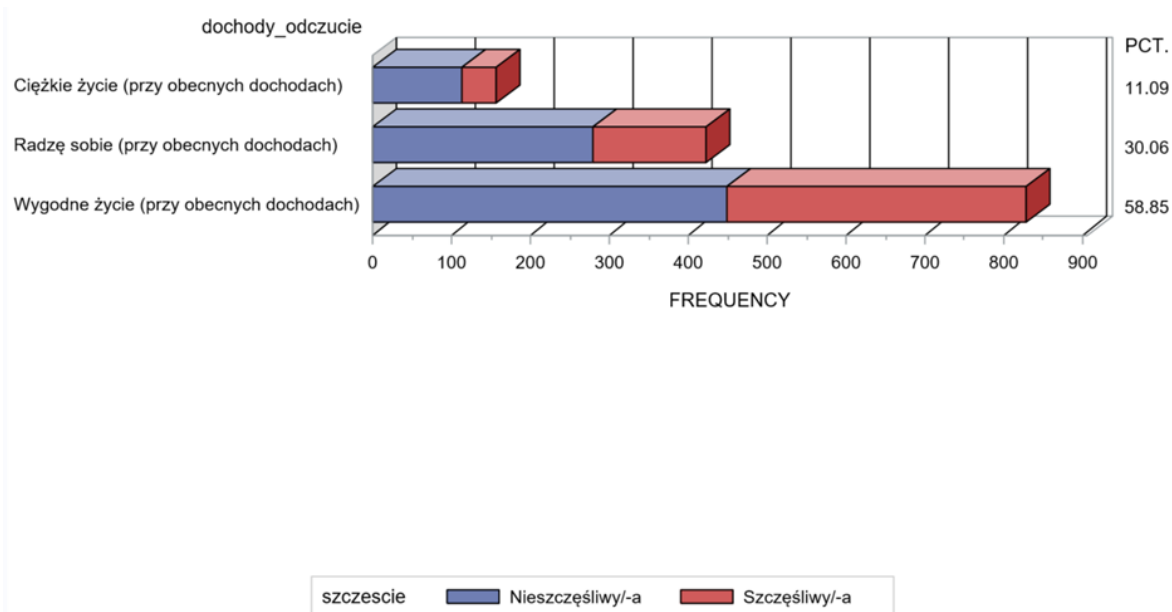
Wykres 19 Zmienna zdrowie

Źródło: Opracowanie własne przy użyciu narzędzi SAS



Wykres 20 Zmienna spacer\_zmrok

Źródło: Opracowanie własne przy użyciu narzędzi SAS



Wykres 21 Zmienna dochody\_odczucie

Źródło: Opracowanie własne przy użyciu narzędzi SAS

## 2.4 Kodowanie zmiennych

Początkowe kodowanie zmiennych niezależnych ma istotny wpływ na przeprowadzaną analizę, a szczególnie największy na ostateczną interpretację wyników. Poniżej przedstawiono tabelę, która określa etykietę oraz kodowanie poszczególnej zmiennej. W niniejszym projekcie przeważają zmienne z kodowaniem referencyjnym.

ZMIENNA	ETYKIETA	KODOWANIE
happy	Czy ankietowany jest szczęśliwy	Zmienna celu
netusoft	Częstość korzystania z internetu	Referencyjne
trstplc	Poziom ufności policji	Referencyjne
stfec0	Satysfakcja z sytuacji ekonomicznej w kraju	Referencyjne
stfhlth	Obecny stan usług zdrowotnych w kraju	Referencyjne
sclmeet	Intensywność spotkań ze znajomymi/krewnymi	Referencyjne
inprdsc	Liczba osób, z którymi respondent może rozmawiać o sprawach intymnych i osobistych	Referencyjne
sclact	Udział w aktywnościach społecznych w porównaniu z innymi osobami w tym samym wieku	Referencyjne
aesfdrk	Poczucie bezpieczeństwa podczas samotnych spacerów w okolicy po zmroku	Referencyjne
health	Stan zdrowia	Referencyjne
rlgblg	Przynależność do określonej religii lub wyznania	Referencyjne
dscrrce	Dyskryminacja grupy respondenta: kolor skóry lub rasa	Referencyjne
gndr	Płeć	Referencyjne
agea	Wiek	Zmienna ilościowa
eisced	Najwyższy poziom wykształcenia, ES - ISCED	Referencyjne
eduysr	Ukończone lata edukacji w pełnym wymiarze godzin	Porządkowe
hincfel	Obecne odczucia dotyczące dochodów gospodarstwa domowego	Referencyjne
livpnt	Czy rodzice wciąż żyją	Referencyjne
ipgdtim	Ważne, by dobrze się bawić	Referencyjne

*Tabela 4 Kodowanie zmiennych*

*Źródło: Opracowanie własne*



### 3. Model regresji logistycznej dla Szwajcarii

#### 3.1 Opis modelu

Celem badania jak zaznaczono wcześniej jest zrozumienie czynników wpływających na poziom szczęścia respondentów. Metoda ta pozwala ocenić wpływ różnych zmiennych niezależnych na prawdopodobieństwo wystąpienia interesującego nas zdarzenia.

W badaniu zastosowano różne zmienne niezależne, takie jak wiek, płeć, poziom wykształcenia, dochód, zdrowie fizyczne i emocjonalne respondentów. Po zebraniu danych od respondentów, przeprowadzono analizę regresji logistycznej, aby określić, które zmienne miały istotny wpływ na poziom szczęścia.

Wyniki analizy regresji logistycznej mogą dostarczyć informacji na temat siły i kierunku wpływu poszczególnych zmiennych na poziom szczęścia respondentów. Na podstawie tych wyników można dokonać predykcji i zidentyfikować czynniki, które mają największe znaczenie dla osiągnięcia wyższego poziomu szczęścia.

Poniżej zostały przedstawione informacje na temat wykonanego modelu:

Rezultaty regresji logistycznej	
Procedura LOGISTIC	
Informacje o modelach	
Zbiór	WORK.SORTTEMPTABLESORTED
Zmienna objaśniana	szczęście
Liczba poziomów odpowiedzi	2
Model	logit binarny
Metoda optymalizacji	Ocena Fishera

*Tabela 5 Informacje o modelu logistycznym*

*Źródło: Opracowanie własne przy użyciu narzędzi SAS*

Powyższa tabela przedstawia podstawowe informacje o modelu logistycznym. Zmienną objaśnianą jest zmienna o nazwie „szczęście”, która przyjmuje dwa poziomy odpowiedzi – szczęśliwy i nieszczęśliwy. W opracowaniu został zastosowany model logit binarny, który ma na celu przewidywanie prawdopodobieństwa przynależności do jednej z dwóch kategorii.

Technika optymalizacji, która została zastosowana w tej analizie, to metoda punktacji Fishera. Metoda ta jest iteracyjną procedurą służącą do estymacji parametrów modelu poprzez maksymalizację funkcji wiarygodności. Jest powszechnie stosowaną techniką w analizie regresji logistycznej.

<b>Liczba obserwacji wczytanych</b>	1407
<b>Liczba obserwacji użytych</b>	1407

*Tabela 6 Liczba wczytanych i użytych obserwacji*

*Źródło: Opracowanie własne przy użyciu narzędzi SAS*

W przeprowadzonej analizie regresji logistycznej uwzględniono łącznie 1407 obserwacji. Żadne obserwacje nie zostały pominięte i wszystkie zostały uwzględnione w analizie. Profil odpowiedzi przedstawia podział obserwacji według wartości zmiennej odpowiedzi "szczęście". Spośród wszystkich obserwacji, 839 osób zaklasyfikowano jako "Nieszczęśliwy/-a", natomiast 568 osób zaklasyfikowano jako "Szczęśliwy/-a".

<b>Profil odpowiedzi</b>		
<b>Wartość uporządkowana</b>	<b>szczęście</b>	<b>Całkowita liczebność</b>
<b>1</b>	Nieszczęśliwy/-a	839
<b>2</b>	Szczęśliwy/-a	568

*Tabela 7 Tabela odpowiedzi*

*Źródło: Opracowanie własne przy użyciu narzędzi SAS*

Powyższa tabela przedstawia liczebność każdej kategorii – Nieszczęśliwy/-a oraz Szczęśliwy/-a. Liczebność z każdej kategorii wynosi odpowiednio 839 i 568.

<b>Status zbieżności</b>
<b>Kryterium zbieżności (GCONV=1E-8) spełnione.</b>

Wynik "Kryterium zbieżności spełnione" wskazuje, że różnica między wynikami kolejnych iteracji jest wystarczająco mała, aby uznać, że algorytm optymalizacji osiągnął zbieżność. Oznacza to, że osiągnięto stabilne i wiarygodne estymacje parametrów modelu metodą Największej Wiarygodności.

### 3.2 Ocena jakości modelu binarnego

Dla ustawionego poziomu istotności wynoszącego  $\alpha = 0.05$  wyniki testu globalnego BETA wskazują, że należy odrzucić hipotezę zerową, która mówi o braku statystycznej istotności całego modelu. Poniższa tabelka przedstawia, że wszystkie trzy testy: Chi-kwadrat, iloczynu Wiarygodności (Mn. Lagrange'a) i Walda, wskazują na istotne różnice między modelem a zerową hipotezą.

Testowanie globalnej hipotezy zerowej: BETA=0			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kw.
Il. wiarygodn.	181.7881	31	<.0001
Mn. Lagrange'a	166.0309	31	<.0001
Walda	146.5928	31	<.0001

Tabela 8 Testowanie globalnej hipotezy zerowej

Źródło: Opracowanie własne przy użyciu narzędzi SAS

Poniższa tabela przedstawia statystyki dewiancji i dobroci dopasowania Pearsona.

Statystyki dewiancji i dobroci dopasowania Pearsona				
Kryterium	Wartość	DF	Wartość/DF	Pr. > chi-kw.
Dewiancja	1713.4310	1373	1.2479	<.0001
Pearsona	1388.5290	1373	1.0113	0.3791

Tabela 9 Statystyki dewiancji i dobroci dopasowania Pearsona

Źródło: Opracowanie własne przy użyciu narzędzi SAS

Wyniki tych statystyk na poziomie istotności  $\alpha = 0.05$  dla statystyki dewiancji sugerują odrzucenie hipotezy zerowej w przeciwieństwie do statystyki Pearsona. Model nie jest idealnie dopasowany, jednak jest statystycznie istotny. Dobroć dopasowania Pearsona wskazuje, że model ma akceptowalną jakość dopasowania do danych, ponieważ nie ma statystycznie istotnych dowodów na odrzucenie hipotezy o dobrej dopasowaniu modelu.

Przyjmując poziom istotności  $\alpha=0.5$  można stwierdzić, że model jest dobrze dopasowany do danych. Testujemy zatem hipotezę zerową, mówiącą o dobrym dopasowaniu modelu do danych, wobec hipotezy alternatywnej o złym odwzorowaniu.

Poniżej przedstawiamy również wyniki Testu zgodności Hosmera i Lemeshowa:

Test zgodności Hosmera i Lemeshowa		
Chi-kwadrat	DF	Pr. > chi-kw.
3.7894	8	0.8756

Tabela 10 Test Zgodności

Źródło: Opracowanie własne przy użyciu narzędzi SAS

Patrząc na Test zgodności Hosmera i Lemeshowa możemy ocenić dopasowanie modelu regresji logistycznej do danych.

H0: model jest dobrze dopasowany do danych

H1: zaprzeczenie hipotezy zerowej

Badamy czy wartość p-value wynosząca 0.8756 jest większa od wartości 0.05.

$0.8756 > 0.05$  zatem: brak podstaw do odrzucenia hipotezy zerowej.

Wnioskując z tego wyniku, można stwierdzić, nie ma istotnych dowodów na to, że model regresji logistycznej źle dopasowuje się do danych. Wartość p-value większa niż 0.05 sugeruje brak istotnej niezgodności między modelem a danymi, co wskazuje na akceptowalne dopasowanie modelu.

Statystyki dopasowania			
Kryterium	Tylko wyraz wolny		Wyraz wolny i współzmiennie
AIC	1899.992		1780.204
SC	1905.241		1948.178
-2 log L	1897.992		1716.204

R-kwadrat	0.1212	Maksymalnie przeskalowane R-kwadrat	0.1637
-----------	--------	-------------------------------------	--------

Tabela 11 Statystyki dopasowania

Źródło: Opracowanie własne przy użyciu narzędzi SAS

R kwadrat (Coxa-Snella) wynosi 0.1212. Oznacza to, że 12,12% zmiennej „Szczęście” może być wyjaśniane przez zmienne niezależne, które zostały użyte w modelu. Maksymalnie przeskalowany R-kwadrat wynosi 0.1637. Oznacza to, gdyby wszystkie zmienne objaśniające w pełni objaśniały zmienną zależną, R-kwadrat wyniósłby 16.37%. Za pomocą tych statystyk można stwierdzić, że model regresji logistycznej ma umiarkowane dopasowanie do danych.

Skojarzenie prognozowanych prawdopodobieństw i obserwowanych odpowiedzi			
Procent zgodnych	70.1	D Somersa	0.402
Procent niezgodnych	29.9	Gamma	0.402
Procent równych	0.0	Tau-a	0.194
Pary	476552	c	0.701

Tabela 12 Skojarzenie prognozowanych prawdopodobieństw i obserwowanych odpowiedzi

Źródło: Opracowanie własne przy użyciu narzędzi SAS

W powyższej tabeli zostały przedstawione wyniki skojarzenia prognozowanych prawdopodobieństw i obserwowanych odpowiedzi. Procent zgodnych obserwacji wynosi 70,1%. Oznacza to, że około 70.1% obserwacji ma zgodną prognozowaną klasę lub odpowiedź. Procent niezgodnych obserwacji wynosi 29.9%

- przeciwieństwo wyżej wspomnianej definicji. D Somersa i Gamma, które wynoszą 0.402 oznaczają, że istnieje umiarkowana zależność między prognozowanymi prawdopodobieństwami a odpowiedziami.

### 3.3 Interpretacja parametrów modelu binarnego

Tabela przedstawia statystyki Chi-kwadrat Walda oraz odpowiadające im poziomy istotności krytycznych. Dla statystyki na poziomie istotności  $\alpha = 0.05$ , istotnie statystycznie są zmienne: policja\_zaufanie, ekonomia\_satysfakcja, usługi\_zdrowotne\_stan, dochody\_odczucie, dobra\_zabawa, zdrowie, dyskryminacja.

Analiza efektów typu 3			
Efekt	DF	Chi-kwadrat Walda	Pr. > chi-kw.
wiek	1	3.6167	0.0572
internet_czestosc	2	4.3562	0.1133
policja_zaufanie	2	7.2902	0.0261
ekonomia_satysfakcja	2	36.9896	<.0001
uslugi_zdrowotne_sta	2	7.4947	0.0236
spotkania	2	1.2688	0.5302
edukacja	2	1.3043	0.5209
edukacja_lata	1	3.1202	0.0773
dochody_odczucie	2	9.6517	0.0080
rodzice	2	0.8549	0.6522
dobra_zabawa	2	16.0398	0.0003
osoby_rozmowy	3	1.9031	0.5928
akt_spol	2	1.2971	0.5228
spacer_zmrok	1	0.6175	0.4320
zdrowie	2	25.4043	<.0001
religia	1	0.7293	0.3931
dyskryminacja	1	3.8762	0.0490
plec	1	1.9470	0.1629

Tabela 13 Analiza efektów zmiennych

Źródło: Opracowanie własne przy użyciu narzędzi SAS

Analizując tabelę można stwierdzić brak korelacji wieku, płci i udziału w aktywnościach społecznych w porównaniu z innymi osobami w tym samym wieku (akt\_spol) ze zmienną szczęście. Mimo wszechobecności internetu, częstotliwość aktywności z nim związanych nie ma istotnego wpływu na odczuwanie szczęścia.

Można zauważyć, że największą istotnością w poziomie odczuwalnego szczęścia odznaczają się satysfakcja z sytuacji ekonomicznej w kraju oraz zdrowie. Dalej można stwierdzić, że osoby z wyższymi

dochodami są szczęśliwsi. Najbardziej zaskakującym faktem jest, że posiadanie rodzica oraz spotkania z rodziną i przyjaciółmi nie mają istotnego wpływu na poziom odczuwalnego szczęścia.

### 3.4 Kodowanie odniesienia

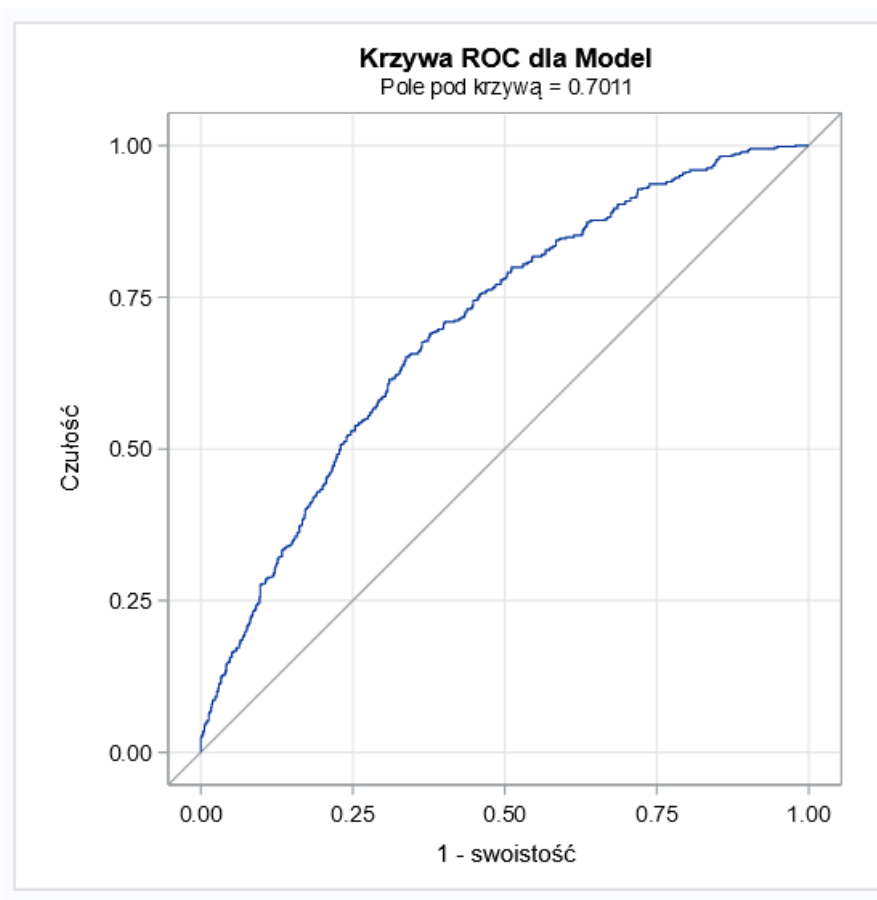
Poniżej przedstawione są estymacje parametrów modelu wraz z ich błędami, a także wartości statystyk Chi-kwadrat Walda wraz z odpowiadającymi im krytycznymi poziomami istotności.

Analiza ocen maksymalnej wiarygodności						
Parametr		DF	Ocena	Błąd standardowy	Chi-kwadrat Walda	Pr. > chi-kw.
Intercept		1	-3.9895	1.3765	8.4002	0.0038
wiek		1	0.0104	0.00545	3.6167	0.0572
internet_czestosc	Czasami	1	0.1602	0.3439	0.2170	0.6414
internet_czestosc	Często	1	-0.3008	0.2890	1.0834	0.2979
policja_zaufanie	Brak zaufania	1	0.1091	0.3416	0.1021	0.7494
policja_zaufanie	Przeciętne zaufanie	1	-0.3265	0.1262	6.6924	0.0097
ekonomia_satysfakcja	Niska satysfakcja	1	-0.9626	0.4675	4.2407	0.0395
ekonomia_satysfakcja	Przeciętna satysfakcja	1	-0.7503	0.1256	35.6664	<.0001
uslugi_zdrowotne_sta	Bardzo dobry stan	1	0.1298	0.4129	0.0989	0.7532
uslugi_zdrowotne_sta	Przeciętny stan	1	-0.2268	0.4166	0.2964	0.5861
spotkania	Czasami	1	-0.2205	0.3065	0.5173	0.4720
spotkania	Często	1	-0.0829	0.2997	0.0765	0.7821
edukacja	Niskie wykształcenie (szkoła średnia i poniżej)	1	-0.0981	0.1609	0.3713	0.5423
edukacja	Wyższe wykształcenie (licencjat, magister)	1	0.1019	0.1985	0.2639	0.6075
edukacja_lata	Wykształcenie podstawowe	1	0.2807	0.1589	3.1202	0.0773
dochody_odczucie	Ciężkie życie (przy obecnych dochodach)	1	-0.5346	0.2124	6.3347	0.0118
dochody_odczucie	Radzę sobie (przy obecnych dochodach)	1	-0.3246	0.1357	5.7217	0.0168
rodzice	Jedno żyje	1	0.1079	0.1794	0.3614	0.5477
rodzice	Nie żyją	1	-0.0521	0.2190	0.0565	0.8120
dobra_zabawa	Istotne	1	0.5762	0.1439	16.0284	<.0001
dobra_zabawa	Nieistotne	1	0.4638	0.3480	1.7765	0.1826
osoby_rozmowy	10 lub więcej osób	1	0.2985	0.4958	0.3625	0.5471
osoby_rozmowy	Od 1 do 3 osób	1	-0.0309	0.4410	0.0049	0.9442
osoby_rozmowy	Od 4 do 9 osób	1	0.0716	0.4445	0.0260	0.8719
akt_spol	Mniejszy (niż innych osób)	1	-0.2080	0.1826	1.2969	0.2548
akt_spol	Taki sam (jak innych osób)	1	-0.1343	0.1687	0.6337	0.4260
spacer_zmrok	Niskie (poczucie bezpieczeństwa)	1	0.1686	0.2145	0.6175	0.4320
zdrowie	Dobry stan zdrowia	1	1.4638	0.4466	10.7436	0.0010
zdrowie	Przeciętny stan zdrowia	1	0.6560	0.4715	1.9357	0.1641
religia	Nie przynależę	1	0.1038	0.1215	0.7293	0.3931
dyskryminacja	Nie jestem/byłem/-am dyskryminowany/-a	1	2.0826	1.0578	3.8762	0.0490
plec	Kobieta	1	0.1682	0.1206	1.9470	0.1629

Tabela 14 Analiza ocen maksymalnej wiarygodności

Źródło: Opracowanie własne przy użyciu narzędzi SAS

Możemy zauważyć, że osoby, które czasem korzystają z internetu są szczęśliwsze. Brak zaufania do policji znacząco wpływa na obniżenie poziomu zadowolenia (-40%). Wraz ze wzrostem satysfakcji ekonomicznej w kraju rośnie poziom szczęścia wśród społeczności o 20%. Nie będzie zaskoczeniem stwierdzenie, że bardzo dobry stan usług zdrowia poprawia poziom zadowolenia o 34%. Osoby, które stwierdziły, że mają ciężkie życie przy obecnych dochodach są mniej szczęśliwe o 20%. Podobny poziom zadowolenia wskazuje istotność dobrej zabawy. Można zauważyć, że ankietowani wskazali, że posiadanie 10 i więcej osób do rozmów intymnych podnosi ich poziom szczęścia. Najważniejszym czynnikiem do bycia szczęśliwym okazał się brak dyskryminacji. Dobry stan zdrowia jest drugim najważniejszym czynnikiem dla ankietowanych.



Wykres 22 Krzywa ROC dla pierwszego modelu

Źródło: Opracowanie własne przy użyciu narzędzi SAS

Krzywa ROC obrazuje zależność między prawdopodobieństwem, że osoba zaklasyfikowana jako "szczęśliwa" jest rzeczywiście szczęśliwa (true positives), a prawdopodobieństwem, że osoba zaklasyfikowana jako "szczęśliwa" jest w rzeczywistości nieszczęśliwa (false positives). Na osi odciętych przedstawione jest prawdopodobieństwo true positives, a na osi pionowej prawdopodobieństwo false positives.



## 4. Model regresji logistycznej z efektami głównymi oraz interakcjami par zmiennych

Model z efektem głównym poddano analizie w celu detekcji zmiennych zakłócających. W tym celu krokowo usuwano z modelu zmienne, których p-value jest największe i większe od 0,05. Następnie badano oszacowania parametrów, a dokładnie czy nastąpiła ich zmiana o więcej niż 15% - w przypadku pozytywnego zajścia usunięta zmienna oznaczała potencjalną zmienną zakłócającą. Zmienne usuwano w następującej kolejności:

- 1) *rodzice* (p-value równe 0,6522) – zmienna zakłócająca;
- 2) *osoby\_rozmowy* (p-value równe 0,6218);
- 3) *spacer\_zmrok* (p-value równe 0,5044) – zmienna zakłócająca;
- 4) *edukacja* (p-value równe 0,5014) – zmienna zakłócająca;
- 5) *spotkania* (p-value równe 0,4805);
- 6) *aktywność\_spoeczna* (p-value równe 0,3092) – zmienna zakłócająca;
- 7) *religia* (p-value równe 0,3370);
- 8) *edukacja\_lata* (p-value równe 0,2712);
- 9) *internet\_czestosc* (p-value równe 0,1196) – zmienna zakłócająca;
- 10) *plec* (p-value równe 0,1191);
- 11) *uslugi\_zdrowotne\_stan* (p-value równe 0,0509);
- 12) *dyskryminacja* (p-value równe 0,0508).

Po zakończeniu krokowego procesu usuwania zmiennych nieistotnych statystycznie pozostały następujące zmienne:

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
wiek	1	12.8471	0.0003
policja_zaufanie	2	9.7186	0.0078
ekonomia_satysfakcja	2	43.5126	<.0001
dochody_odczucie	2	11.5297	0.0031
dobra_zabawa	2	19.7426	<.0001
zdrowie	2	24.4518	<.0001

Tabela 15 Pozostałe zmienne po krokowym usuwaniu zmiennych nieistotnych statystycznie

Źródło: Opracowanie własne przy użyciu narzędzi SAS



Porównując model uwzględniający wszystkie zmienne niezależne (Model 1) oraz model po usunięciu zmiennych nieistotnych statystycznie (Model 2) wzięliśmy pod uwagę statystyki AIC (Akaike's Information Criterion), SC (Schwarz's Bayesian Criterion).

Model 1:

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1899.992	1780.204
SC	1905.241	1948.178
-2 Log L	1897.992	1716.204

*Tabela 16 Statystyki dopasowania dla Modelu 1*

*Źródło: Opracowanie własne przy użyciu narzędzi SAS*

Model 2:

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1899.992	1769.808
SC	1905.241	1832.799
-2 Log L	1897.992	1745.808

*Tabela 17 Statystyki dopasowania dla Modelu 2*

*Źródło: Opracowanie własne przy użyciu narzędzi SAS*

Statystyki w Modelu 2 mają niższe wartości niż w Modelu 1. Na podstawie tych wyników można wnioskować, że Model 2, ma lepsze dopasowanie do danych niż Model 1.

## 4.1 Proces regresji logistycznej z efektami głównymi przy wykorzystaniu metody stepwise

Metoda stepwise pozwoliła na identyfikację najważniejszych zmiennych niezależnych w modelu regresji logistycznej. Polega na iteracyjnym dodawaniu i usuwaniu zmiennych na podstawie ich istotności statystycznej w modelu.

Etap 0. Podano wstawiane efekty

Status zbieżności			
Kryterium zbieżności (GCONV=1E-8) spełnione.			

Statystyki dopasowania		
Kryterium	Tylko wyraz wolny	Wyraz wolny i współzmiennie
AIC	1899.992	1780.204
SC	1905.241	1948.178
-2 log L	1897.992	1716.204

R-kwadrat	0.1212	Maksymalnie przeskalowane R-kwadrat	0.1637
-----------	--------	-------------------------------------	--------

Testowanie globalnej hipotezy zerowej: BETA=0			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kw.
Il. wiarygodn.	181.7881	31	<.0001
Mn. Lagrange'a	166.0309	31	<.0001
Walda	146.5928	31	<.0001

Test resztowy chi-kwadrat		
Chi-kwadrat	DF	Pr. > chi-kw.
439.6006	431	0.3769

Dla modelu z wyrazem wolnym i współzmiennymi, wartości statystyk dopasowania są mniejsze niż dla modelu z samym wyrazem wolnym. Oznacza to, że model z wyrazem wolnym i współzmiennymi lepiej dopasowuje się do danych niż model tylko z wyrazem wolnym. Przyjęcie hipotezy zerowej (brak wpływu współzmiennych) można odrzucić, ponieważ wartości p są mniejsze niż

poziom istotności. Test resztowy chi-kwadrat sugeruje brak istotności statystycznej.

Etap 1. Wstawiono efekt rodzice\*osoby\_rozmowy

Status zbieżności			
Wykryto quasi-całkowitą rozdzielną punków danych.			

Statystyki dopasowania		
Kryterium	Tylko wyraz wolny	Wyraz wolny i współzmiennie
AIC	1899.992	1774.151
SC	1905.241	1973.622
-2 log L	1897.992	1698.151

R-kwadrat	0.1324	Maksymalnie przeskalowane R-kwadrat	0.1788
-----------	--------	-------------------------------------	--------

Testowanie globalnej hipotezy zerowej: BETA=0			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kw.
Il. wiarygodn.	199.8403	37	<.0001
Mn. Lagrange'a	178.3355	37	<.0001
Walda	154.4268	37	<.0001

Test resztowy chi-kwadrat		
Chi-kwadrat	DF	Pr. > chi-kw.
435.9078	425	0.3469

Uwaga: Nie usunięto żadnych efektów z modelu w kroku 1.

W etapie 1 dodaliśmy efekt interakcji pomiędzy zmiennymi "rodzice" i "osoby\_rozmow". Model z efektem interakcji lepiej dopasowuje się do danych niż model bez tego efektu ze względu na mniejsze statystyk dopasowania wyrazu wolnego i współzmiennych. Test resztowy chi-kwadrat

wskazuje, że nie ma istotności statystycznej, czyli dostosowanie modelu jest odpowiednie. W kroku 1 nie usunięto żadnych efektów z modelu. Oznacza to, że efekt interakcji "rodzice\*osoby\_rozmow" został zachowany w modelu.

się do danych niż model bez tego efektu ze względu na statystyki dopasowania. Odrzucamy hipotezę zerową (brak wpływu efektu interakcji). Test resztowy chi-kwadrat wskazuje, że nie ma istotności statystycznej. Efekt interakcji dodany na tym etapie został zachowany w modelu.

Identyczne wnioski wynikają również z etapów 3, 4 i 5 dla kolejnych interakcji.

## Etap 2. Wstawiono efekt spacer\_zmrok\*religia

Status zbieżności			
Wykryto quasi-całkowitą rozdzielną punków danych.			

Statystyki dopasowania		
Kryterium	Tylko wyraz wolny	Wyraz wolny i współzmiennie
AIC	1899.992	1770.369
SC	1905.241	1975.088
-2 log L	1897.992	1692.369

R-kwadrat	0.1360	Maksymalnie przeskalowane R-kwadrat	0.1836
-----------	--------	-------------------------------------	--------

Testowanie globalnej hipotezy zerowej: BETA=0			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kw.
Il. wiarygodn.	205.6230	38	<.0001
Mn. Lagrange'a	183.3542	38	<.0001
Walda	158.1985	38	<.0001

Test resztowy chi-kwadrat		
Chi-kwadrat	DF	Pr. > chi-kw.
432.7979	424	0.3734

**Uwaga:** Nie usunięto żadnych efektów z modelu w kroku 2.

Podobnie jak w poprzednim etapie model z dodanym efektem interakcji lepiej dopasowuje

## Etap 3. Wstawiono efekt usługi\_zdrowotne\_stan\*rodzice

Status zbieżności			
Wykryto quasi-całkowitą rozdzielną punków danych.			

Statystyki dopasowania		
Kryterium	Tylko wyraz wolny	Wyraz wolny i współzmiennie
AIC	1899.992	1765.584
SC	1905.241	1991.301
-2 log L	1897.992	1679.584

R-kwadrat	0.1438	Maksymalnie przeskalowane R-kwadrat	0.1942
-----------	--------	-------------------------------------	--------

Testowanie globalnej hipotezy zerowej: BETA=0			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kw.
Il. wiarygodn.	218.4072	42	<.0001
Mn. Lagrange'a	193.5318	42	<.0001
Walda	164.2500	42	<.0001

Test resztowy chi-kwadrat		
Chi-kwadrat	DF	Pr. > chi-kw.
429.1770	420	0.3679

**Uwaga:** Nie usunięto żadnych efektów z modelu w kroku 3.

## Etap 4. Wstawiono efekt dobra\_zabawa\*spacer\_zmrok

Status zbieżności
Wykryto quasi-całkowitą rozdzielną punków danych.

Statystyki dopasowania		
Kryterium	Tylko wyraz wolny	Wyraz wolny i współzmiennie
AIC	1899.992	1761.637
SC	1905.241	1997.851
-2 log L	1897.992	1671.637

R-kwadrat	0.1486	Maksymalnie przeskalowane R-kwadrat	0.2007
-----------	--------	-------------------------------------	--------

Testowanie globalnej hipotezy zerowej: BETA=0			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kw.
Il. wiarygodn.	226.3549	44	<.0001
Mn. Lagrange'a	200.5174	44	<.0001
Walda	168.3124	44	<.0001

Test resztowy chi-kwadrat		
Chi-kwadrat	DF	Pr. > chi-kw.
423.8147	418	0.4117

Uwaga: Nie usunięto żadnych efektów z modelu w kroku 4.

Etap 5. Wstawiono efekt  
usługi\_zdrowotne\*edukacja

Etap 6. Efekt rodzice\*osoby\_rozmowy został  
usunięty

Status zbieżności
Wykryto quasi-całkowitą rozdzielną punków danych.

Statystyki dopasowania		
Kryterium	Tylko wyraz wolny	Wyraz wolny i współzmiennie
AIC	1899.992	1756.776
SC	1905.241	2013.988
-2 log L	1897.992	1658.776

R-kwadrat	0.1564	Maksymalnie przeskalowane R-kwadrat	0.2111
-----------	--------	-------------------------------------	--------

Testowanie globalnej hipotezy zerowej: BETA=0			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kw.
Il. wiarygodn.	239.2155	48	<.0001
Mn. Lagrange'a	210.3521	48	<.0001
Walda	172.7149	48	<.0001

Test resztowy chi-kwadrat		
Chi-kwadrat	DF	Pr. > chi-kw.
412.2157	414	0.5155

Etap 6. Efekt rodzice\*osoby\_rozmowy został usunięty:

Status zbieżności
Wykryto quasi-całkowitą rozdzielną punków danych.

Statystyki dopasowania		
Kryterium	Tylko wyraz wolny	Wyraz wolny i współzmiennie
AIC	1899.992	1761.093
SC	1905.241	1986.810
-2 log L	1897.992	1675.093

R-kwadrat	0.1465	Maksymalnie przeskalowane R-kwadrat	0.1979
-----------	--------	-------------------------------------	--------

Testowanie globalnej hipotezy zerowej: BETA=0			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kw.
Il. wiarygodn.	222.8982	42	<.0001
Mn. Lagrange'a	198.4024	42	<.0001
Walda	166.9228	42	<.0001

Test resztowy chi-kwadrat		
Chi-kwadrat	DF	Pr. > chi-kw.
417.6768	420	0.5228

Uwaga: Nie usunięto żadnych efektów z modelu w kroku 6.

Dopiero w etapie szóstym usunęliśmy efekt usługi\_zd\*spacer\_zmr został usunięty.

Dla modelu z wyrazem wolnym i współzmiennymi bez efektu "usługi\_zd\*spacer\_zmr", wartości statystyk dopasowania są nieco wyższe niż dla modelu z efektem interakcji. Jednak różnica jest niewielka. Ze względu na statystyki globalnej hipotezy odrzucamy hipotezę zerową (brak wpływu efektu

"uslugi\_zd\*spacer\_zmr"). Test resztowy chi-kwadrat wskazuje, że nie ma istotności statystycznej, co sugeruje odpowiednie dopasowanie modelu.

W modelu pozostały inne efekty, ale efekt "uslugi\_zd\*spacer\_zmr" został usunięty ze względu na kryterium statystyki Walda.

Ostateczny przebieg i wyniki procesu zostały przedstawiony w poniższej tabeli.

Podsumowanie selekcji krokowej							
Krok	Efekt		DF	Liczba włączonych	Chi-kwadrat ocen	Chi-kwadrat Walda	Pr. > chi-kw.
	Wstawione	Usunięte					
1	rodzice*osoby_rozmow		6	19	16.7429		0.0103
2	spacer_zmrok*religia		1	20	5.7837		0.0162
3	uslugi_zdrow*rodzice		4	21	12.2071		0.0159
4	dobra_zab*spacer_zmr		2	22	7.8662		0.0196
5	uslugi_zdro*edukacja		4	23	12.0974		0.0166
6		rodzice*osoby_rozmow	6	22		11.7748	0.0672

Tabela 18 Podsumowanie selekcji krokowej

Źródło: Opracowanie własne przy użyciu narzędzi SAS

Wynika z niej, że dodanie interakcji w krokach od 1 do 5 poprawiło dopasowanie modelu. Usunięcie efektu uslugi\_zd\*spacer\_zmr nieznacznie pogorszyło dopasowanie modelu.

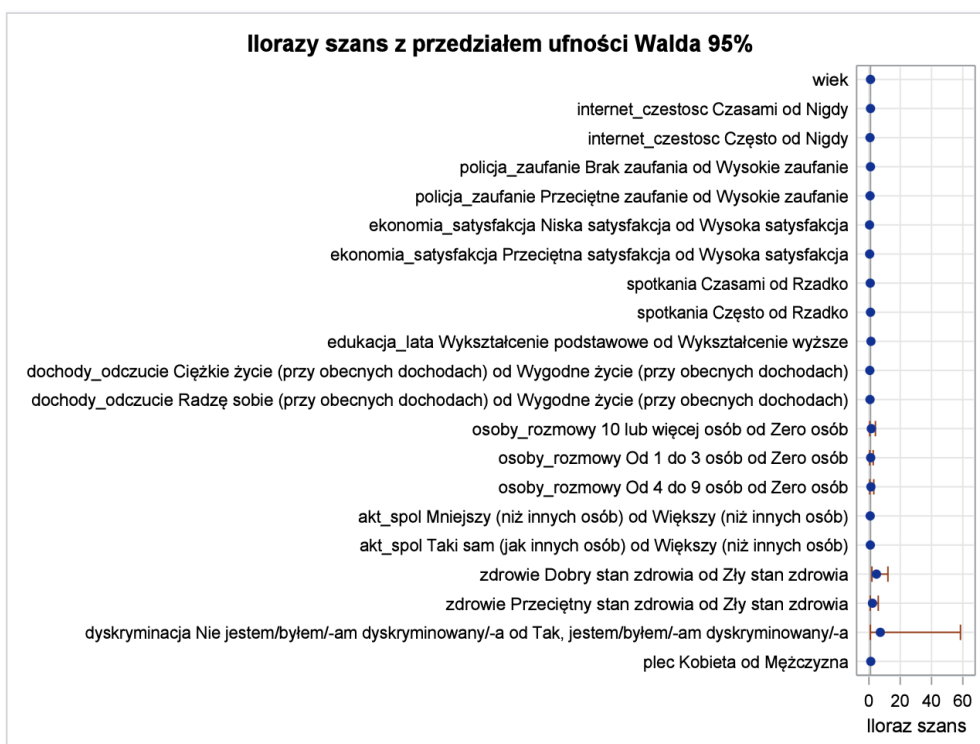
Statystyki dewiancji i dobroci dopasowania Pearsona				
Kryterium	Wartość	DF	Wartość/DF	Pr. > chi-kw.
Dewiancja	1672.3209	1362	1.2278	<.0001
Pearsona	1381.5018	1362	1.0143	0.3502

Liczba unikatowych profili: 1405

Tabela 19 Statystyki dewiancji i dobroci dopasowania Pearsona

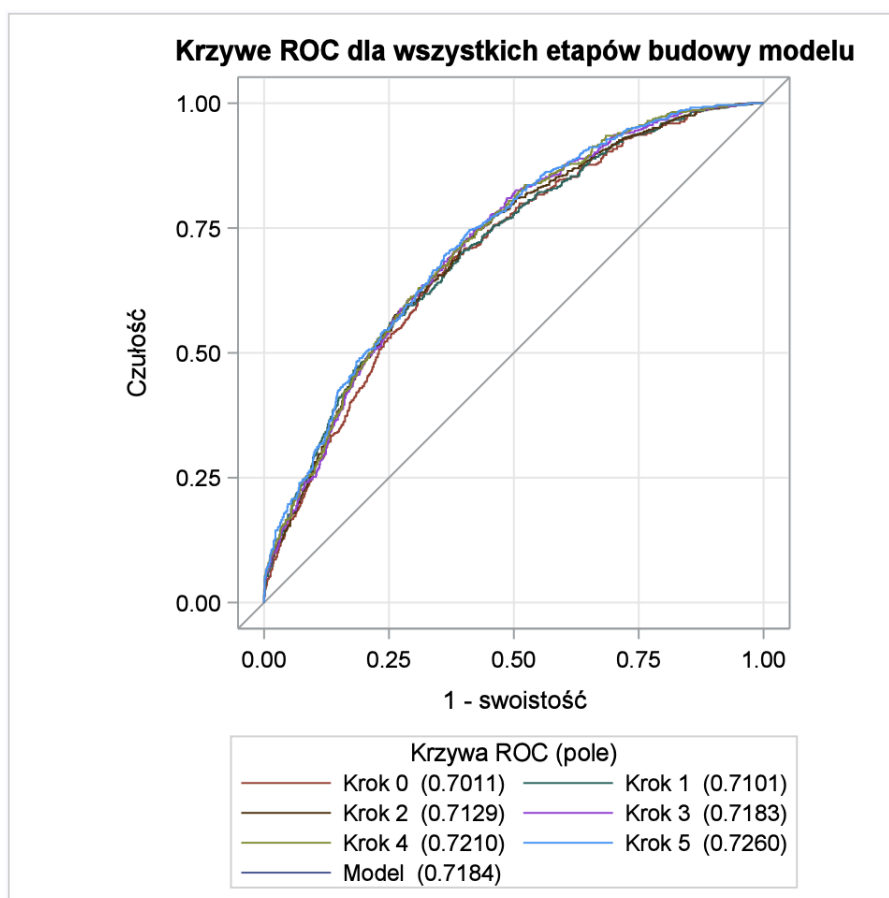
Źródło: Opracowanie własne przy użyciu narzędzi SAS

Wartość/DF bliska 1 wskazuje na dobre dopasowanie modelu. W danych występuje 1401 różnych kombinacji wartości zmiennych niezależnych. Na podstawie statystyk dewiancji i dobroci dopasowania Pearsona można stwierdzić, że model regresji logistycznej ma istotne dopasowanie do danych.



Wykres 23 Iloraz szans z przedziałem ufności Walda

Źródło: Opracowanie własne przy użyciu narzędzi SAS



Wykres 24 Krzywa ROC dla modelu z wykorzystaniem metody stepwise

Źródło: Opracowanie własne przy użyciu narzędzi SAS

Model z efektami głównymi oraz interakcjami zmiennych ponownie poddano analizie w celu detekcji zmiennych zakłócających. W tym celu krokowo usuwano z modelu zmienne, których p-value jest największe i większe od 0,05 oraz nie są one w interakcji z innymi zmiennymi. Następnie badano oszacowania parametrów, a dokładnie czy nastąpiła ich zmiana o więcej niż 15% - w przypadku pozytywnego zajścia usunięta zmienna oznaczać może potencjalną zmienną zakłócającą. W modelach badano także wartość statystyk: AIC oraz SC. Odpowiednio z modelu usuwano następujące zmienne:

1. *spotkania* (p-value równe 0.9765) – potencjalna zmienna zakłócająca;
2. *osoby\_rozmowy* (p-value równe 0.5354) – potencjalna zmienna zakłócająca;
3. *rodzice* (p-value równe 0.7408) – zmienna zakłócająca;
4. *edukacja* (p-value równe 0.5889) – zmienna zakłócająca;
5. *akt\_spol* (p-value równe 0.2455) – zmienna zakłócająca;
6. *internet\_czestosc* (p-value równe 0.1355).

Po redukcji modelu i usunięciu prawdopodobnych zmiennych zakłócających (bez usuwania zmiennych w interakcjach) w modelu pozostały następujące zmienne:

Testy łączne			
Efekt	DF	Chi-kwadrat Walda	Pr. > chi-kw.
wiek	1	5.9420	0.0148
policja_zaufanie	2	3.2433	0.1976
ekonomia_satysfakcja	2	38.0452	<.0001
uslugi_zdrowotne_sta	2	4.0152	0.1343
edukacja_lata	1	6.3698	0.0116
dochody_odczucie	2	12.5598	0.0019
dobra_zabawa	2	15.6341	0.0004
spacer_zmrok	1	3.9605	0.0466
zdrowie	2	27.6542	<.0001
religia	1	0.0235	0.8782
dyskryminacja	1	3.9015	0.0482
plec	1	7.3709	0.0066
wiek*policja_zaufani	2	6.9457	0.0310
wiek*uslugi_zdrowotn	2	6.8474	0.0326
edukacja_lata*plec	1	4.7030	0.0301
dobra_zab*spacer_zmr	2	7.5443	0.0230
spacer_zmrok*religia	1	5.9429	0.0148

Tabela 20 Podsumowanie selekcji krokowej dla modelu z interakcjami

Źródło: Opracowanie własne przy użyciu narzędzi SAS

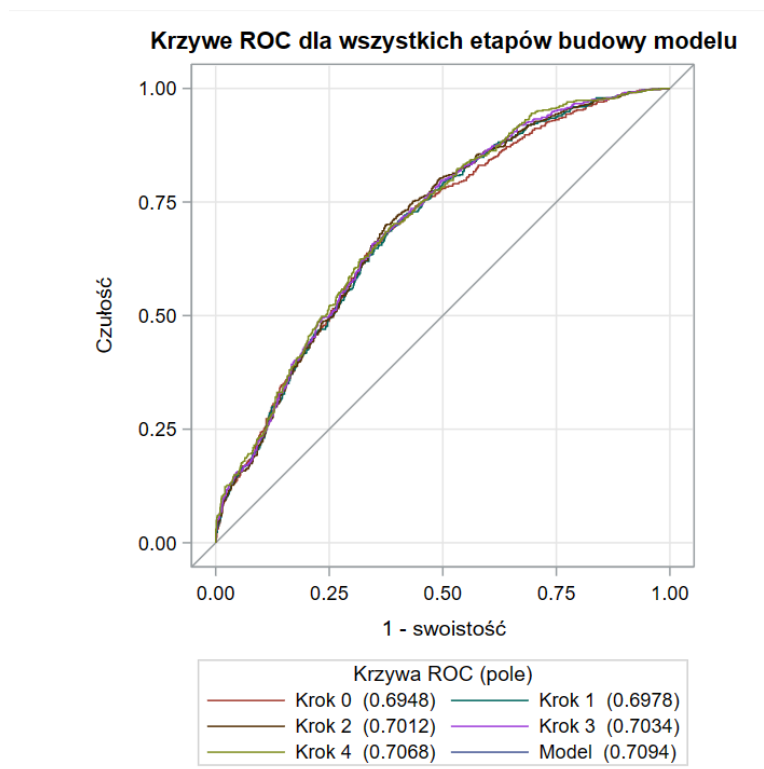


Tabela 21 Krzywa ROC dla modelu z wykorzystaniem metody stepwise po redukcji

Źródło: Opracowanie własne przy użyciu narzędzi SAS

W przypadku poniższych wyników odnośnie Modelu 1 (model z interakcjami) i Modelu 2 (model z interakcjami poddany redukcji odrzucając zmienne nieistotnie statystycznie) możemy stwierdzić, że w przypadku obu modeli, wartość  $Pr. > \chi^2$  jest bardzo mała, co wskazuje na istotność statystyczną dla obu modeli. Model 2 ma nieznacznie lepsze dopasowanie do danych w porównaniu do Modelu 1 jako sugerowane przez niższe wartości Dewiancji i Wartości/DF.

Model 1:

Statystyki dewiancji i dobroci dopasowania Pearsona				
Kryterium	Wartość	DF	Wartość/DF	Pr. > $\chi^2$
Dewiancja	1672.3209	1362	1.2278	<.0001
Pearsona	1381.5018	1362	1.0143	0.3502

Tabela 22 Statystyki dopasowania dla Modelu 1 w metodzie stepwise

Źródło: Opracowanie własne przy użyciu narzędzi SAS

Model 2:



Statystyki dewiencji i dobroci dopasowania Pearsona				
Kryterium	Wartość	DF	Wartość/DF	Pr. > chi-kw.
Dewiencja	1618.4352	1325	1.2215	<.0001
Pearsona	1333.2839	1325	1.0063	0.4311

Tabela 23 Statystyki dopasowania dla Modelu 2 w metodzie stepwise

Źródło: Opracowanie własne przy użyciu narzędzi SAS

Porównując wyniki odnośnie skojarzenia prognozowanych prawdopodobieństw i obserwowanych odpowiedzi możemy zauważyć, że miary te są bardzo podobne. Oba modele mają podobne wartości procentu zgodnych, D Somersa, procentu niezgodnych, Gamma, procentu równych, Tau-a oraz liczbę par. Różnice między tymi wartościami są niewielkie. Model 1 ma nieznacznie wyższy współczynnik C, co sugeruje nieco lepsze skojarzenie między tymi zmiennymi.

#### Model 1:

Skojarzenie prognozowanych prawdopodobieństw i obserwowanych odpowiedzi			
Procent zgodnych	71.8	D Somersa	0.437
Procent niezgodnych	28.2	Gamma	0.437
Procent równych	0.0	Tau-a	0.210
Pary	476552	c	0.718

Tabela 24 Skojarzenie prognozowanych prawdopodobieństw i obserwowanych odpowiedzi dla Modelu 1 w metodzie stepwise

Źródło: Opracowanie własne przy użyciu narzędzi SAS

#### Model 2:

Skojarzenie prognozowanych prawdopodobieństw i obserwowanych odpowiedzi			
Procent zgodnych	70.9	D Somersa	0.419
Procent niezgodnych	29.1	Gamma	0.419
Procent równych	0.0	Tau-a	0.202
Pary	476552	c	0.709

Tabela 25 Skojarzenie prognozowanych prawdopodobieństw i obserwowanych odpowiedzi dla Modelu 2 w metodzie stepwise

Źródło: Opracowanie własne przy użyciu narzędzi SAS

Poniżej przedstawione są dane odnośnie Modelu 2 z interakcjami po wykonaniu redukcji krokowej. Możemy zauważyć przykładowo, że osoby o dobrym stanie zdrowia mają większe szanse na odczuwanie szczęścia w porównaniu do osób ze złym stanem zdrowia przy założeniu, że porównywane grupy osób nie różnią się pod względem pozostałych zmiennych niezależnych (zasada ceteris paribus). Przedział ufności wskazuje, że ta różnica jest znacząca i istotna statystycznie.

Oceny ilorazów szans i przedziały ufności Walda				
Efekt	Jednostka	Ocena	Przedział ufności 95%	
ekonomia_satysfakcja Niska satysfakcja od Wysoka satysfakcja	1.0000	0.352	0.138	0.896
ekonomia_satysfakcja Przeciętna satysfakcja od Wysoka satysfakcja	1.0000	0.468	0.385	0.599
dochody_odczucie Ciężkie życie (przy obecnych dochodach) od Wygodne życie (przy obecnych dochodach)	1.0000	0.557	0.389	0.839
dochody_odczucie Radzę sobie (przy obecnych dochodach) od Wygodne życie (przy obecnych dochodach)	1.0000	0.693	0.533	0.901
zdrowie Dobry stan zdrowia od Zły stan zdrowia	1.0000	5.990	2.385	15.171
zdrowie Przeciętny stan zdrowia od Zły stan zdrowia	1.0000	2.766	1.038	7.371
dyskryminacja Nie jestem/byłem/-am dyskryminowany/-a od Tak, jestem/byłem/-am dyskryminowany/-a	1.0000	8.134	1.016	65.099

*Tabela 26 Oceny ilorazów szans i przedziały ufności Walda*

*Źródło: Opracowanie własne przy użyciu narzędzi SAS*

## 5. Podsumowanie

Przeprowadzony projekt miał na celu analizę czynników wpływających na stosunek Szwajcarii do poziomu szczęścia, wykorzystując model regresji logistycznej. Na podstawie zebranych danych można wyciągnąć kilka istotnych wniosków.

Główna hipoteza badawcza sugerowała, że aktywność społeczna Szwajcarów, takie jak intensywność spotkań ze znajomymi lub krewnymi, udział w aktywnościach społecznych oraz liczba osób, z którymi mogą rozmawiać o sprawach intymnych i osobistych, mają wpływ na ich poziom szczęścia. Wyniki analizy potwierdziły tę hipotezę, wskazując na dodatnią korelację między aktywnością społeczną a poziomem szczęścia respondentów. Im wyższa aktywność społeczna, tym respondent czuł się szczęśliwszy. Ankietowani wskazali, że posiadanie 10 i więcej osób do rozmów intymnych podnosi ich poziom szczęścia.

Kolejna zbadana hipoteza sugerowała, że lepsza sytuacja polityczno-społeczna kraju ma związek ze szczęśliwością jego obywateli. Czynniki takie jak satysfakcja z sytuacji ekonomicznej, ocena stanu usług zdrowotnych, zaufanie do służb porządkowych (np. policji) oraz poczucie bezpieczeństwa podczas samotnych spacerów w okolicy po zmroku mogą wpływać na komfort życia w kraju. Analiza danych potwierdziła tę hipotezę, wskazując na istotną korelację między tymi czynnikami a poziomem szczęścia. Respondenci odczuwali większe zadowolenie o 20%, gdy sytuacja ekonomiczna była lepsza, również o 34% gdy stan usług zdrowotnych w kraju był dobry. Natomiast brak zaufania do policji istotnie obniżał poziom zadowolenia o 40%, podobnie jak poczucie bezpieczeństwa podczas samotnych spacerów w okolicy po zmroku.

Ponadto, wyniki analizy ujawniły kilka innych czynników, które mają istotny wpływ na poziom szczęścia respondentów. Osoby korzystające czasem z internetu wykazywały większą szczęśliwość. Brak dyskryminacji został uznany za najważniejszy czynnik wpływający na bycie szczęśliwym, a dobry stan zdrowia był drugim najważniejszym czynnikiem.

## 6. Fragmenty Kodu SAS

Poniżej przedstawiono elementy kodu zawierającego import danych, kategoryzacje zmiennych oraz blok kodu PROC LOGISTIC pokazującego proces przeprowadzania regresji logistycznej (model z efektami głównymi).

```
DATA WORK.CH_final;
```

```
  LENGTH
```

name	\$ 8
essround	8
edition	8
proddate	8
idno	8
cntry	\$ 2
dweight	8
pspwght	8
pweight	8
anweight	8
prob	8
stratum	8
psu	8
netusoft	8
trstplc	8
stfeco	8
stfhlth	8
happy	8
sclmeet	8
inprdsc	8
sclact	8
aesfdrk	8
health	8
rlgblg	8
dscrrce	8
gndr	8
agea	8
eisced	8

eduyrs	8
hincfel	8
livpnt	8
ipgdtim	8 ;

#### FORMAT

name	\$CHAR8.
essround	BEST2.
edition	BEST3.
proddate	DDMMYY10.
idno	BEST5.
cntry	\$CHAR2.
dweight	BEST1.
pspwght	BEST10.
pweight	BEST10.
anweight	BEST10.
prob	BEST13.
stratum	BEST3.
psu	BEST4.
netusoft	BEST1.
trstplc	BEST2.
stfeco	BEST2.
stfhlth	BEST2.
happy	BEST2.
sclmeet	BEST2.
inprdsc	BEST2.
sclact	BEST1.
aesfdrk	BEST1.
health	BEST1.
rlgblg	BEST1.
dscrrce	BEST1.
gndr	BEST1.
agea	BEST3.
eisced	BEST2.
eduyrs	BEST2.
hincfel	BEST1.
livpnt	BEST1.
ipgdtim	BEST1. ;

#### INFORMAT

name	\$CHAR8.
essround	BEST2.
edition	BEST3.
proddate	DDMMYY10.
idno	BEST5.
cntry	\$CHAR2.
dweight	BEST1.
pspwght	BEST10.
pweight	BEST10.
anweight	BEST10.
prob	BEST13.
stratum	BEST3.
psu	BEST4.
netusoft	BEST1.
trstplc	BEST2.
stfeco	BEST2.
stfhlth	BEST2.
happy	BEST2.
sclmeet	BEST2.
inprdsc	BEST2.
sclact	BEST1.
aesfdrk	BEST1.
health	BEST1.
rlgblg	BEST1.
dscrrce	BEST1.
gndr	BEST1.
agea	BEST3.
eisced	BEST2.
eduyrs	BEST2.
hincfel	BEST1.
livpnt	BEST1.
ipgdtim	BEST1. ;

INFILE 'C:\Users\wikto\Desktop\Nowy folder (2)\CH\_final.csv'

LRECL=137

ENCODING="WLATIN2"

TERMSTR=CRLF

DLM='7F'x

MISSOVER

```

DSD ;
INPUT
name           : $CHAR8.
essround       : ?? BEST2.
edition        : ?? COMMA3.
proddate       : ?? DDMMYY10.
idno           : ?? BEST5.
cntry          : $CHAR2.
dweight        : ?? BEST1.
pspwght        : ?? COMMA10.
pweight        : ?? COMMA10.
anweight       : ?? COMMA10.
prob           : ?? COMMA13.
stratum        : ?? BEST3.
psu            : ?? BEST4.
netusoft       : ?? BEST1.
trstplc        : ?? BEST2.
stfeco         : ?? BEST2.
stfhlth        : ?? BEST2.
happy          : ?? BEST2.
sclmeet        : ?? BEST2.
inprdsc        : ?? BEST2.
sclact         : ?? BEST1.
aesfdrk        : ?? BEST1.
health         : ?? BEST1.
rlgblg         : ?? BEST1.
dscrrce        : ?? BEST1.
gndr           : ?? BEST1.
agea           : ?? BEST3.
eiscd          : ?? BEST2.
eduyrs         : ?? BEST2.
hincfel        : ?? BEST1.
livpnt         : ?? BEST1.
ipgdtim        : ?? BEST1. ;

```

```

RUN;

```

```

%LET _CLIENTTASKLABEL=;

```

```

%LET _CLIENTPROCESSFLOWNAME=;
%LET _CLIENTPROJECTPATH=;
%LET _CLIENTPROJECTPATHHOST=;
%LET _CLIENTPROJECTNAME=;

/*   POCZĄTEK WĘZŁA: Rekodowanie zmiennych SQL   */
%LET _CLIENTTASKLABEL='Rekodowanie zmiennych SQL';
%LET _CLIENTPROCESSFLOWNAME='Przebieg procesu';
%LET _CLIENTPROJECTPATH='C:\Users\poczt\OneDrive -
SGH\Dokumenty\Regresja\Projekt\Szwajcaria_wiek_numeryczna_29.05_najnowszy.egp
';
%LET _CLIENTPROJECTPATHHOST='PROOBOOK';
%LET _CLIENTPROJECTNAME='Szwajcaria_wiek_numeryczna_29.05_najnowszy.egp';
%LET _SASPROGRAMFILE='';
%LET _SASPROGRAMFILEHOST='';

```

```

PROC SQL;

```

```

CREATE TABLE szwajcaria AS

```

```

SELECT happy,

```

```

CASE

```

```

    WHEN happy <= 8 THEN 'Nieszczęśliwy/-a'

```

```

    WHEN happy >= 9 THEN 'Szczęśliwy/-a'

```

```

END AS szczescie,

```

```

CASE

```

```

    WHEN netusoft = 1 THEN 'Nigdy'

```

```

    WHEN netusoft IN (2,3) THEN 'Czasami'

```

```

    WHEN netusoft IN (4,5) THEN 'Często'

```

```

END AS internet_czestosc,

```

```

CASE

```

```

    WHEN trstplc BETWEEN 0 AND 3 THEN 'Brak zaufania'

```

```

    WHEN trstplc BETWEEN 4 AND 7 THEN 'Przeciętne zaufanie'

```

```

    WHEN trstplc BETWEEN 8 AND 10 THEN 'Wysokie zaufanie'

```

```

END AS policja_zaufanie,

```

```

CASE
  WHEN stfeco BETWEEN 0 AND 3 THEN 'Niska satysfakcja'
  WHEN stfeco BETWEEN 4 AND 7 THEN 'Przeciętna satysfakcja'
  WHEN stfeco BETWEEN 8 AND 10 THEN 'Wysoka satysfakcja'
END AS ekonomia_satysfakcja,

```

```

CASE
  WHEN stfhlth BETWEEN 0 AND 3 THEN 'Zły stan'
  WHEN stfhlth BETWEEN 4 AND 7 THEN 'Przeciętny stan'
  WHEN stfhlth BETWEEN 8 AND 10 THEN 'Bardzo dobry stan'
END AS usługi_zdrowotne_stan,

```

```

CASE
  WHEN sclmeet IN (1,2) THEN 'Rzadko'
  WHEN sclmeet IN (3,4) THEN 'Czasami'
  WHEN sclmeet IN (5,6,7) THEN 'Często'
END AS spotkania,

```

```

CASE
  WHEN inprdsc = 0 THEN 'Zero osób'
  WHEN inprdsc BETWEEN 1 AND 3 THEN 'Od 1 do 3 osób'
  WHEN inprdsc IN (4,5) THEN 'Od 4 do 9 osób'
  WHEN inprdsc = 6 THEN '10 lub więcej osób'
END AS osoby_rozmowy,

```

```

CASE
  WHEN sclact IN (1,2) THEN 'Mniejszy (niż innych osób)'
  WHEN sclact = 3 THEN 'Taki sam (jak innych osób)'
  WHEN sclact IN (4,5) THEN 'Większy (niż innych osób)'
END AS akt_spol,

```

```

CASE
  WHEN aesfdrk IN (1,2) THEN 'Wysokie (poczucie bezpieczeństwa)'
  WHEN aesfdrk IN (3,4) THEN 'Niskie (poczucie bezpieczeństwa)'
END AS spacer_zmrok,

```

```

CASE
  WHEN health IN (1,2) THEN 'Dobry stan zdrowia'

```



```

    WHEN health = 3 THEN 'Przeciętny stan zdrowia'
    WHEN health IN (4,5) THEN 'Zły stan zdrowia'
END AS zdrowie,

CASE
    WHEN rlgblg = 1 THEN 'Tak, przynależę'
    WHEN rlgblg = 2 THEN 'Nie przynależę'
END AS religia,

CASE
    WHEN dsccrce = 1 THEN 'Tak, jestem/byłem/-am dyskryminowany/-a'
    WHEN dsccrce = 0 THEN 'Nie jestem/byłem/-am dyskryminowany/-a'
END AS dyskryminacja,

CASE
    WHEN gndr = 1 THEN 'Mężczyzna'
    WHEN gndr = 2 THEN 'Kobieta'
END AS plec,

/* case*/
/*    when agea <= 38 then 'młody wiek'*/
/*    when agea between 39 and 60 then 'średni wiek'*/
/*    when agea >= 60 then 'stary wiek'*/
/* end as */
agea as wiek,

CASE
    WHEN eisced BETWEEN 1 AND 4 THEN 'Niskie wykształcenie (szkoła średnia i
poniżej) '
    WHEN eisced = 5 THEN 'Średnie wykształcenie (zaawansowane zawodowe,
podyplomowe) '
    WHEN eisced IN (6,7) THEN 'Wyższe wykształcenie (licencjat, magister) '
END AS edukacja,

CASE
    WHEN eduys <= 12 THEN 'Wykształcenie podstawowe'
    WHEN eduys > 12 THEN 'Wykształcenie wyższe'
END AS edukacja_lata,
```

```

CASE
    WHEN hincfel = 1 THEN 'Wygodne życie (przy obecnych dochodach)'
    WHEN hincfel = 2 THEN 'Radzę sobie (przy obecnych dochodach)'
    WHEN hincfel IN (3,4) THEN 'Ciężkie życie (przy obecnych dochodach)'
END AS dochody_odczucie,

CASE
    WHEN livpnt = 1 THEN 'Tak, oboje'
    WHEN livpnt IN (2,3) THEN 'Jedno żyje'
    WHEN livpnt = 4 THEN 'Nie żyją'
END AS rodzice,

CASE
    WHEN ipgdtim IN (1,2) THEN 'Istotne'
    WHEN ipgdtim IN (3,4) THEN 'Średnio istotne'
    WHEN ipgdtim IN (5,6) THEN 'Nieistotne'
END AS dobra_zabawa
FROM CH_final
WHERE
    (happy BETWEEN 0 AND 10) AND
    (netusoft BETWEEN 1 AND 5) AND
    (trstplc BETWEEN 0 AND 10) AND
    (stfeco BETWEEN 0 AND 10) AND
    (stfhlth BETWEEN 0 AND 10) AND
    (sclmeet BETWEEN 1 AND 7) AND
    (inprdsc BETWEEN 0 AND 6) AND
    (sclact BETWEEN 1 AND 5) AND
    (aesfdrk BETWEEN 1 AND 4) AND
    (health BETWEEN 1 AND 5) AND
    (rlgblg BETWEEN 1 AND 2) AND
    (dscrrce BETWEEN 0 AND 1) AND
    (gnr BETWEEN 1 AND 2) AND
    (agea > 0) AND
    (agea <> 999) AND
    (eiscd BETWEEN 1 AND 7) AND
    (eduyrs <> 88) AND
    (hincfel BETWEEN 1 AND 4) AND

```

```

(livpnt BETWEEN 1 AND 4) AND
(ipgdtim BETWEEN 1 AND 6);

QUIT;

/*  POCZĄTEK WĘZŁA: Pie Chart  */
%LET _CLIENTTASKLABEL='Pie Chart';
%LET _CLIENTPROCESSFLOWNAME='Przebieg procesu';
%LET _CLIENTPROJECTPATH='C:\Users\poczt\OneDrive -
SGH\Dokumenty\Regresja\Projekt\Szwajcaria_wiek_numeryczna_29.05_najnowszy.egp
';
%LET _CLIENTPROJECTPATHHOST='PROOBOOK';
%LET _CLIENTPROJECTNAME='Szwajcaria_wiek_numeryczna_29.05_najnowszy.egp';

/*  POCZĄTEK WĘZŁA: Logistic Regression  */
%LET _CLIENTTASKLABEL='Logistic Regression ';
%LET _CLIENTPROCESSFLOWNAME='Przebieg procesu';
%LET _CLIENTPROJECTPATH='C:\Users\poczt\OneDrive -
SGH\Dokumenty\Regresja\Projekt\Szwajcaria_wiek_numeryczna_29.05_najnowszy.egp
';
%LET _CLIENTPROJECTPATHHOST='PROOBOOK';
%LET _CLIENTPROJECTNAME='Szwajcaria_wiek_numeryczna_29.05_najnowszy.egp';

/* -----
Kod wygenerowany przez zadanie SAS-a

Wygenerowany dnia: poniedziałek, 29 maja 2023 o godz. 19:08:06
Przez zadanie: Logistic Regression

Dane wejściowe: Local:WORK.SZWAJCARIA
Serwer: Local
----- */
ODS GRAPHICS ON;

%_eg_conditional_dropds(WORK.SORTTempTableSorted);
/* -----
Sortowanie zbioru Local:WORK.SZWAJCARIA
----- */

```

```

PROC SQL;
CREATE VIEW WORK.SORTTempTableSorted AS
SELECT    T.szczescie,    T.wiek,    T.internet_czestosc,    T.policja_zaufanie,
T.ekonomia_satysfakcja,    T.uslugi_zdrowotne_stan,    T.spotkania,    T.edukacja,
T.edukacja_lata,    T.dochody_odczucie,    T.rodzice,    T.dobra_zabawa,
T.osoby_rozmowy, T.akt_spol
        , T.spacer_zmrok, T.zdrowie, T.religia, T.dyskryminacja, T.plec
FROM WORK.SZWAJCARIA as T
;
QUIT;
TITLE;
TITLE1 "Rezultaty regresji logistycznej";
FOOTNOTE;
FOOTNOTE1 "Wygenerowane przez SAS-a (&_SASSERVERNAME, &SYSSCPL) dnia
%TRIM(%QSYFUNK (DATE(), NLDATE20.)) o godz. %TRIM(%QSYFUNK (TIME(),
NLTIMAP25.))";
PROC LOGISTIC DATA=WORK.SORTTempTableSorted
PLOTS (ONLY)=ODDSRATIO
PLOTS (ONLY)=ROC
;
CLASS internet_czestosc    (PARAM=REF) policja_zaufanie    (PARAM=REF)
ekonomia_satysfakcja (PARAM=REF) uslugi_zdrowotne_stan    (PARAM=REF)
spotkania (PARAM=REF) edukacja (PARAM=REF) edukacja_lata (PARAM=REF)
dochody_odczucie    (PARAM=REF)
    rodzice (PARAM=REF) dobra_zabawa    (PARAM=REF) osoby_rozmowy (PARAM=REF)
akt_spol    (PARAM=REF) spacer_zmrok    (PARAM=REF) zdrowie    (PARAM=REF)
religia    (PARAM=REF) dyskryminacja (PARAM=REF) plec    (PARAM=REF);
MODEL    szczescie (Event = 'Szczęśliwy/-a ')=wiek    internet_czestosc
policja_zaufanie ekonomia_satysfakcja uslugi_zdrowotne_stan spotkania edukacja
edukacja_lata dochody_odczucie rodzice dobra_zabawa osoby_rozmowy akt_spol
spacer_zmrok zdrowie religia dyskryminacja plec    /
SELECTION=NONE
LACKFIT
AGGREGATE SCALE=NONE
RSQUARE
CTABLE
LINK=LOGIT

```

CLPARM=WALD

CLODDS=WALD

ALPHA=0.05

;

RUN;

QUIT;

## 7. Bibliografia

1. Amati, Viviana, i inni "Social relations and life satisfaction: The role of friends." *Genus* 74 (2018): 1-18.
2. Blanchflower, David G., and Andrew J. Oswald. "Is well-being U-shaped over the life cycle?." *Social science & medicine* 66.8 (2008): 1733-1749.
3. Brakus, J. Joško, i inni "Experiences and happiness: The role of gender." *Psychology & Marketing* 39.8 (2022): 1646-1659.
4. Diener, Ed, and Robert Biswas-Diener. "Will money increase subjective well-being?." *Social indicators research* 57 (2002): 119-169.
5. Dumitrache, Cristina G., Laura Rubio, and Ramona Rubio-Herrera. "Perceived health status and life satisfaction in old age, and the moderating role of social support." *Aging & mental health* 21.7 (2017): 751-757.
6. Huta, Veronika, and Richard M. Ryan. "Pursuing pleasure or virtue: The differential and overlapping well-being benefits of hedonic and eudaimonic motives." *Journal of happiness studies* 11 (2010): 735-762.
7. Kim, Junhyoung, i inni "The contribution of physical and social activity participation to social support and happiness among people with physical disabilities." *Disability and health journal* 14.1 (2021): 100974.
8. Oerlemans, Wido GM, Arnold B. Bakker, and Ruut Veenhoven. "Finding the key to happy aging: A day reconstruction study of happiness." *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 66.6 (2011): 665-674.
9. R. Williams, David, and Ruth Williams-Morris. "Racism and mental health: The African American experience." *Ethnicity & health* 5.3-4 (2000): 243-268.

10. Rizvi, M. A. K., & Hossain, M. Z. (2016). Relationship Between Religious Belief and Happiness: A Systematic Literature Review. *Journal of Religion and Health*, 56(5), 1561–1582. doi:10.1007/s10943-016-0332-6