Warszawa, 08.05.2019 r.

**REGRESJA LOGISTYCZNA Z WYKORZYSTANIEM NARZĘDZI SAS**

**223480-1528**

SEMESTR LETNI 2018/2019

**Dr inż. Aleksandra Zofia Iwanicka**

Projekt zaliczeniowy I – Model regresji logistycznej – modelowanie poziomu szczęścia na przykładzie Norwegii

**Lista autorów:**

Michał Hajdan 68501

Anna Kozak 68570

Marcin Mandziej 68373

Eryk Mazuś 68835

Edyta Rybicka 67670

Sylwia Semeniuk 67978

Justyna Stanek 64235

Paweł Tafliński 67500

**Spis treści**

[Wstęp 3](#_Toc8216976)

[Hipotezy badawcze 5](#_Toc8216977)

[Eksploracja zbioru danych 11](#_Toc8216978)

[Analiza zmiennej objaśnianej 14](#_Toc8216979)

[Analiza zmiennych objaśniających 17](#_Toc8216980)

[Zmienne ciągłe 17](#_Toc8216981)

[Zmienne nominalne 23](#_Toc8216982)

[Zmienne porządkowe 46](#_Toc8216983)

[Analiza współliniowości 59](#_Toc8216984)

[Tolerancja zmiennej 60](#_Toc8216985)

[Wskaźnik inflacji wariancji 61](#_Toc8216986)

[Weryfikacja modelu regresji logistycznej 63](#_Toc8216987)

[Model uwzględniający tylko efekty główne 66](#_Toc8216988)

[Modele uzyskane metodą selekcji forward i stepwise 68](#_Toc8216989)

[Interpretacja parametrów modelu 69](#_Toc8216990)

[Ocena jakości modelu 74](#_Toc8216991)

[Test Hosmera-Lemeshowa 74](#_Toc8216992)

[Dyskryminacja modelu 75](#_Toc8216993)

[Krzywa ROC 76](#_Toc8216994)

[Analiza reszt, miary wpływu 77](#_Toc8216995)

[Podsumowanie i wnioski 79](#_Toc8216996)

[Bibliografia 81](#_Toc8216997)

[Spis tabel 82](#_Toc8216998)

[Spis wykresów 84](#_Toc8216999)

[Spis rysunków 85](#_Toc8217000)

[Wykorzystany kod 86](#_Toc8217001)

# Wstęp

Celem raportu przygotowanego przez zespół projektowy jest zbadanie determinant szczęścia wśród mieszkańców Norwegii oraz próba odpowiedzi na pytania postawione w hipotezach badawczych. W tym celu przygotowany został model regresji logistycznej, który został oparty na danych pochodzących z 8 rundy badania European Social Survey.

Badanie zostało przeprowadzone w 23 państwach Europy. Jego głównym celem jest monitorowanie i interpretacja postaw, zachowań i wartości społecznych występujących w poszczególnych krajach, które uczestniczyły w badaniu. Oprócz tego badanie ma na celu ulepszanie metod badania zjawisk społecznych oraz opracowanie wskaźników postaw społecznych i porównanie wyników ich wartości dla poszczególnych krajów.

Dane wykorzystane w modelu zebrane dla obywateli Norwegii pochodzą z 2016 r. Do analizy zdychotomizowanej zmiennej objaśnianej - deklarowanego **poziomu szczęścia** - wykorzystano 13 zmiennych objaśniających, które według przypuszczeń zespołu projektowego mają największy wpływ na odczuwany poziom szczęścia oraz których wpływ może wydawać się nieintuicyjny i ciekawy z punktu widzenia państwa jakim jest Norwegia, takich jak poziom religijności oraz liczba godzin poświęcanych na pracę w ciągu tygodnia.

Według badań przeprowadzonych przez ośrodek badań rynkowych i opinii społecznych Ipsos w społeczeństwie Norwegii w 2016 r. po raz pierwszy liczba osób deklarujących ateizm przewyższyła liczbę osób deklarujących przynależność do jakiejkolwiek religii. Od tego czasu według badań liczba osób deklarujących ateizm stale wzrasta w porównaniu do osób deklarujących przynależność religijną. Przypuszczeniem zespołu projektowego jest, że Norwegowie deklarujący ateizm mogą statystycznie różnić się poziomem szczęścia od osób wyznających wiarę.

Zespół projektowy przypuszcza, że liczba przepracowanych godzin w tygodniu ma wpływ na poczucie odczuwanego szczęścia nawet w państwie takim jak Norwegia, gdzie sytuacja finansowa obywateli znajduje się w większości na wysokim i bezpiecznym poziomie. Według danych z 2017 r. Norwegia jest 5 państwem na świecie z największym współczynnikiem siły nabywczej, dlatego może się wydawać, że Norwegowie nie muszą poświęcać dużej ilości czasu na zapewnienie sobie dostatnich warunków życia i być może osoby poświęcające dużo czasu na pracę czynią to z powodów innych niż finansowe, np. spełniania swoich ambicji i samorealizacji. Zespół projektowy przypuszcza, że osoby takie mogą statystycznie różnić się pod względem poziomu odczuwanego szczęścia od osób, które poświęcają przeciętną liczbę godzin w tygodniu na pracę.

# Hipotezy badawcze

Pierwszą hipotezą badawczą postawioną przez zespół projektowy jest przypuszczenie, że wysoki poziom religijności zwiększa szansę na poczucie wysokiego poziomu szczęścia. W opinii zespołu projektowego jest to interesujące założenie. Zgodnie z intuicją może się wydawać, że zaangażowanie w rozwój duchowy i religijny może wpływać na odczuwanie pozytywnych emocji, zwiększoną świadomość i stabilne relacje psychiczne, co powinno wpływać na zwiększone poczucie szczęścia przez daną osobę. Przeprowadzona analiza zbioru danych ma na celu wskazać, czy postawiona teza znajduje odzwierciedlenie w rzeczywistości.

W opinii zespołu projektowego przeprowadzone badanie może wskazać interesujące rezultaty, zwłaszcza w przypadku państwa takiego jak Norwegia, gdzie jak w pozostałych krajach skandynawskich religii przypisuje się marginalną rolę i państwa te charakteryzują się wysokim odsetkiem osób deklarujących ateizm lub niewielkie zaangażowanie w życie religijne.

Drugą hipotezą badawczą postawioną przez zespół projektowy jest przypuszczenie, że duża liczba godzin w tygodniu spędzonych na pracy zmniejsza szansę na poczucie wysokiego poziomu szczęścia. Intuicyjnie może się wydawać, że duża liczba godzin poświęconych na pracę powinna wpływać na zmniejszenie odczuwanego poziomu szczęścia.

Głównym przypuszczeniem zespołu projektowego jest to, że aktywność zawodowa odbiera danej osobie czas, który mogłaby ona poświęcić na czynności, które wydają się bardziej powiązane z wysokim poziomem odczuwanego szczęścia, np. spędzanie czasu z rodziną i innymi bliskimi osobami, aktywność fizyczna, rozwój pasji i zainteresowań.

Z drugiej strony, wydawać się może, że w Norwegii jako państwie o wysokiej kulturze pracy i poszanowania pracowników, wysokim poziomie zaawansowania cywilizacyjnego oraz wysokich płacach realnych, gdzie obywatele zazwyczaj nie mają problemów finansowych osoby, które poświęcają na aktywność zawodową dużą ilość swojego czasu mogą robić to, ponieważ odczuwają np. wysoki poziom zadowolenia i satysfakcji z wykonywanej pracy i bardziej niż z powodów finansowych wykonują ją, ponieważ jest to swego rodzaju pasja i zainteresowanie. To z kolei powinno zwiększać odczuwany przez nich poziom szczęścia. Analiza zbioru danych ma na celu wskazać, czy tak faktycznie jest w rzeczywistości.

Obie wartości, których dotyczą hipotezy badawcze od dawna stanowią ważną część życia zdecydowanej większości osób. Praca wykonana przez zespół projektowy stanowi próbę odpowiedzi na pytanie, czy wysoki poziom religijności i przywiązywanie wagi do wartości duchowych oraz podejmowanie i poświęcanie dużej ilości czasu aktywności zawodowej przekładają się na poziom szczęścia obywateli Norwegii i jeśli tak – to w jaki sposób wpływa na jego poziom.

W literaturze obie wartości są wskazywane jako ważne elementy składowe ludzkiego życia i przypisuje im się dużą wagę jako determinantom szczęścia.

Temat religijności jako determinantu szczęścia porusza między innymi Tatiana Kamasz, w swoim dziele *“Uwarunkowania szczęścia”.* Autorka powołuje się na przeprowadzone badania empiryczne, które wskazują na to, że osoby zaangażowanie w życie duchowe i religijne mogą być bardziej szczęśliwe od osób, które tego nie robią. Statystycznie osoby uczęszczające do kościoła są bardziej zadowolone z życia w porównaniu do osób, które tego nie robią. Nie ma jednak dowodów na to, że przedstawiciele konkretnych religii są przeciętnie bardziej szczęśliwi od osób wyznających inne wierzenia.

Według autorki stopień religijności wpływa na wyższy poziom odczuwanego szczęścia z 3 głównych powodów. Po pierwsze, zaangażowanie w życie religijne pozwala odnaleźć wspólny system przekonań, który ułatwia odnalezienie sensu życia i pozytywnego nastawienie do przyszłości. Co więcej osoby angażujące się we wspólnotach religijnych odczuwają przynależność do określonej grupy społecznej i wsparcie jej uczestników. Dodatkowo, dobre praktyki, które wiążą się z zaangażowaniem religijnym (np. umiarkowana konsumpcja, post, jałmużna, modlitwa, medytacja) wpływają pozytywnie na samopoczucie pod względem fizycznym jak i psychicznym[[1]](#footnote-1).

Temat wpływu religijności na odczuwane szczęście został także poruszony w pracy pt. *“World happiness report”*. Autorzy publikacji wskazują na badania, które dowodzą dodatniej korelacji (0,14) pomiędzy korzystaniem z “usług religijnych” (religious services) a ogólnym poczuciem szczęścia i zadowolenia z życia. Badania zostały przeprowadzone w latach 2013 - 2016, jednym z ich rezultatów jest to, że osoby zaangażowane w jakiś sposób w życie religijne cechują się wyższym poziomem szczęścia w porównaniu do osób, które takiej aktywności nie deklarują. Autorzy badania wskazują także ciekawy, ale i kontrowersyjny wniosek. Mianowicie, religijność pozwala wykształcić danej osobie cechy takie jak, np. sumienność, pracowitość, które pomagają w życiu podczas realizacji postawionych celów. Przynależność religijną warto także rozważyć w kontekście historycznym. Autorzy twierdzą, że uwarunkowania historyczne wpływają na to, że osoby religijne są w pewien sposób postrzegane lepiej, bardziej pożądane do sprawowania wysokich pozycji, np. urzędniczych czy naukowych. To z kolei przekładać się może na wyższą pozycję społeczną tychże osób, a co za tym idzie - ich odczuwany poziom szczęścia[[2]](#footnote-2).

Wpływ religijności na odczuwany poziom szczęścia został opisany i uzasadniony przez autorów pracy *“A Compass Towards a Just and Harmonious Society”*. Badacze wskazują, że wysoki poziom religijności danej osoby zazwyczaj pozytywnie wpływa na jej samopoczucie psychiczne z powodu wyciszenia, poczucia spokoju i kontroli nad swoimi emocjami. Bardzo duży wpływ na taki stan rzeczy mają rytuały związane z praktykami religijnymi, np. modlitwa, pielgrzymki, medytacje, udział w kazaniach, czy nawet świadome zadawanie sobie cierpienia poprzez głód lub okaleczanie fizyczne. Tego rodzaju obrzędy, rytuały mogą stanowić swego rodzaju plan, strukturę życiową, która przynosi ze sobą poczucie sensu życia.

Autorzy wskazują także na korelację pomiędzy deklarowanym stopniem religijności a udziałem w inicjatywach wolontaryjnych i pomocy innym ludziom. Te 2 aktywności przyczyniają się do kontaktu i zacieśniania więzi z innymi ludźmi oraz satysfakcji z pomocy i wykonanej dla nich pracy, co przekłada się na wyższy poziom odczuwanego szczęścia. Autorzy pracy wskazują jednak, że deklarowany wysoki stopień religijności może wpływać negatywnie na odczuwany poziom szczęścia np. z powodu prześladowań na tle religijnym lub innych traumatycznych doświadczeń doświadczonych podczas wyznawania swojej wiary, bądź spowodowanych przez przynależność do określonej grupy religijnej[[3]](#footnote-3).

Eksperci badający wpływ pracy i aktywności zawodowej na odczuwany poziom szczęścia przez człowieka wskazują, że są to ważne czynniki warunkujące dobrostan psychiczny i fizyczny deklarowany przez badane osoby. Panowie Thomas A. Wright, Kymberley K. Benett i Ty Dun w artykule pt. “*Life and Job Satisfaction*” po przeprowadzeniu badań na ten temat wyodrębnili 3 zasadnicze koncepcje, swego rodzaju typy podejścia do pracy, które łączą satysfakcję (lub jej brak) związaną z aktywnością zawodową a poziomem odczuwanego szczęścia.

Jako pierwsze podejście autorzy wyznaczyli kompensację - podejście, które deklarują osoby nie odczuwające dużego poziomu satysfakcji z wykonywanej pracy - jest ona dla nich swego rodzaju złem koniecznym - środkiem pozwalającym realizować się i osiągać szczęścia w innych obszarach życia, np. w życiu rodzinnym i towarzyskim, w rozwoju i realizowaniu pasji i zainteresowań.

Następnie wyróżnić można podejście określane jako segmentacja - rozdzielenie pracy zawodowej i innych aspektów życia związanych z aktywnością pozazawodową (autorzy wymieniają jako przykładowe obszary wypoczynek i czas wolny, życie rodzinne, wykształcenie). W tym przypadku osoby przywiązane do takiego podejścia deklarują zazwyczaj, że satysfakcja z poszczególnych obszarów życia ma wpływ na odczuwane przez nich szczęście, choć nie wskazują aktywności zawodowej jako kluczowy czynnik determinujący ich poczucie szczęścia.

Jako ostatnie i zdaniem zespołu projektowego najbardziej intuicyjne podejście autorzy przytaczają koncepcję oddziaływania, według której, satysfakcja z wykonywanej pracy oddziałuje na inne obszary życia człowieka. Dzieje się tak dlatego, ponieważ zadowolenie lub jego brak z pracy zawodowej wpływa bezpośrednio na dobrostan fizyczny i psychiczny danej osoby, który wpływa dalej na pozostałe aspekty życia. Co za tym idzie, bardziej prawdopodobne jest, że osoba usatysfakcjonowana zawodowo będzie deklarować wyższy poziom satysfakcji w pozostałych dziedzinach życia, jak np. zadowolenie z relacji rodzinnych, samoocena, zadowolenie z miejsca zamieszkania czy kraju. Autorzy przestrzegają jednak przed przypisywaniem pracy zawodowej zbyt dużej wagi podczas jej szacowania jako determinanty szczęścia. Przeprowadzone przez nich badania empiryczne wskazują, że istnieją ważniejsze czynniki takie jak np. kontakty towarzyskie, wypoczynek i rekreacja oraz aktywność pozazawodowa[[4]](#footnote-4).

Tematyka satysfakcji zawodowej i jej wpływ na odczuwane przez człowieka szczęście została poruszona także w artykule naukowym *“Satysfakcja z pracy jako źródło dobrostanu psychicznego u polskich emigrantów konsumpcyjnych w Kanadzie”.* Dzieło opisuje sytuację emigrantów wyjeżdżających z Polski do Kanady w poszukiwaniu lepszej pracy i warunków życia i wskazuje aktywność zawodową jako kluczowy czynnik determinujący satysfakcję życiową tych osób. Przemyślenia zawarte w tekście znajdują jednak odzwierciedlenie w przypadku każdej innej sytuacji.

Autorzy przy wysuwaniu przypuszczenia o istotności wpływu satysfakcji zawodowej na odczuwany poziom szczęścia powołują się na badania przeprowadzone przez psychologów pracy w Stanach Zjednoczonych oraz krajach skandynawskich. Według nich na satysfakcję z wykonywanej pracy składa się wiele czynników, z czego do najważniejszych należą możliwość aktywnego działania i spełniania swoich ambicji, niezależność i możliwość wypełniania różnorodnych i ambitnych zadań przy wykorzystaniu swoich zdolności i predyspozycji, zdobycie uznania i szacunku z powodu dobrze wykonywanej pracy oraz stosunki i relacje interpersonalne ze współpracownikami i przełożonymi. Autorzy dopiero jako czynnik drugiego rzędu wskazują rekompensatę finansową za wykonywaną pracę oraz fizyczne warunki pracy.

Zdaniem autorów praca zawodowa jest kluczowym czynnikiem, który wpływa na poczucie szczęścia lub jego brak, czego główną przyczyną jest fakt, że zazwyczaj ludzie spędzają na pracy znaczącą część swojego życia. Dodatkowo, wykonywanie pracy wymaga często znacznego zaangażowania emocjonalnego ze strony pracownika. Dlatego zadowolenie z wykonywanej pracy powinno wpływać na odczuwanie szczęścia przez człowieka, natomiast niezadowolenie powinno warunkować brak szczęścia u danej osoby[[5]](#footnote-5).

Temat wykonywanej pracy jako determinantu szczęścia został poruszony także w artykule “*Health, Wealth and Happiness: Why Pursue a Higher Education?*”, gdzie wraz z odpowiednią edukacją i zdrowiem aktywność zawodowa została wskazana jako jeden z kluczowych czynników, które wpływają na to, czy dana osoba jest szczęśliwa, czy nie. Autorzy badania wysuwają przypuszczenie, że podejmowanie aktywności zawodowej z jednej strony może wpływać pozytywnie na odczuwaną satysfakcję z życia dla danej osoby m. in. z powodu poczucia celu wynikającego z wykonywanej pracy, poczucia wspólnoty z współpracownikami oraz satysfakcji z efektów pracy. Z drugiej strony wykonywanie pracy może wiązać się z podwyższonym poziomem stresu, zmęczeniem fizycznym lub psychicznym oraz brakiem motywacji i satysfakcji wynikającym z nieprzyjaznego środowiska pracy[[6]](#footnote-6).

# Eksploracja zbioru danych

Dane wykorzystane w poniższym raporcie pochodzą z European Social Survey – Europejskiego Sondażu Społecznego z 2016 r. Badanie jest jednym z największych i najważniejszych europejskich projektów badawczych. Inicjatorem był Komitet Ekspertów Europejskiej Fundacji Nauki (European Science Foundation). Celem badania jest obserwacja zmian społecznych i porównywanie wyników między krajami. Pierwsza edycja badania miała miejsce w 2002 r., kolejne odbywały się co dwa lata. Ostatnia miała miejsce w 2016 r.   
W poszczególnych rundach wzięło udział od 22 do 31 krajów.

Wyniki uzyskane w Europejskim Sondażu Społecznym są powszechnie wykorzystywane. Oryginalny zbiór danych wykorzystany w raporcie liczył 1545 obserwacji i 534 zmienne.

Po analizie literatury, którą wyszczególniono przy opisie poszczególnej zmiennej, zdecydowano się wybrać 13 zmiennych objaśniających. Wybrano 4 zmienne porządkowe, 4 zmienne ciągłe oraz 5 zmiennych nominalnych. Wszystkie zostały opisane w poniższej tabeli.

|  |  |  |
| --- | --- | --- |
| RODZAJ ZMIENNEJ | NAZWA ZMIENNEJ | OPIS ZMIENNEJ |
| Porządkowa objaśniana | happy | How happy are you |
| Porządkowa objaśniająca | health | Subjective general health |
| stfdem | How satisfied with the way democracy works in country |
| atchctr | How emotionally attached to [country] |
| rlgdgr | How religious are you |
| Ciągła objaśniająca | hhmmb | Number of people living regularly as member of household |
| agea | Age of respondent, calculated |
| wkhtot | Total hours normally worked per week in main job overtime included |
| eduyrs | Years of full-time education completed |
| Nominalna objaśniająca | gndr | Gender |
| jbspv | Responsible for supervising other employees |
| tporgwk | What type of organisation work/worked for |
| hinctnta | Household's total net income, all sources |
| region | Region |

Tabela 1. Zestawienie zmiennych wykorzystanych w modelu.

Wyjściowy zbiór liczył 1545 obserwacji. Po eksploracji zbioru, zdecydowano się usunąć odpowiedzi typu „Nie wiem”, „Odmowa odpowiedzi”, „Brak danych”. Nie odnotowano braków danych w zbiorze przygotowanym w wyniku Europejskiego Sondażu Społecznego, co przedstawia Tabela nr 2. Do dalszych analiz przejęto zbiór liczący 1440 obserwacji. Zmienną region przeformułowano ze zmiennej znakowej na zmienną numeryczną, aby można było wykorzystać ją w modelu.

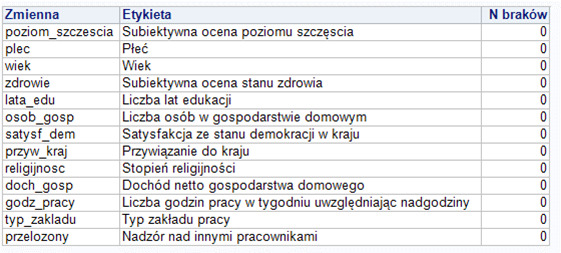


Tabela 2. Zestawienie braków danych.

Przed rozpoczęciem dalszych obliczeń zmieniono nazwy zmiennych na bardziej intuicyjne. Dodano również etykiety, a po kategoryzacji nadano format dla każdej z kategorii. Zmienne *osob\_gosp* oraz *godz\_pracy* postanowiono pozostawić w charakterze zmiennych ciągłych. Poniższa tabela przedstawia nowe nazwy i nowe kategorie wszystkich zmiennych.

|  |  |  |  |
| --- | --- | --- | --- |
| **RODZAJ ZMIENNEJ** | **NAZWA ZMIENNEJ** | **NOWA NAZWA ZMIENNEJ** | **NOWE KATEGORIE ZMIENNYCH** |
| **Nominalna**  **objaśniana** | Happy | poziom\_szczescia | 0 - Nieszczęśliwa/y  1 – Szczęśliwa/y |
| **Ciągła objaśniająca** | Hhmmb | osob\_gosp | - |
| **Ciągła objaśniająca** | Wkhtot | godz\_pracy | - |
| **Nominalna objaśniająca** | Gndr | plec | 0 - Mężczyźni  1 – Kobiety |
| **Nominalna**  **objaśniająca** | Agea | wiek | 1 - 22 lat i mniej  2 - 23-30 lat  3 - 31-45 lat  4 - 46-66 lat  5 - 67 lat i więcej |
| **Nominalna objaśniająca** | Jbspv | przelozony | 0 – Nie  1 - Tak |
| **Nominalna objaśniająca** | Tporgwk | typ\_zakladu | 1 - Sektor państwowy  2 - Sektor prywatny  3 - Samozatrudnienie  4 - Inne |
| **Nominalna objaśniająca** | Eduyrs | lata\_edu | 1 - 12 lat i mniej  2 - 13-14 lat  3 - 15-17 lat  4 - 18 lat i więcej |
| **Nominalna**  **objaśniająca** | Rlgdgr | religijnosc | 1 – Niewierząca/y  2 – Wierząca/y  3 – Bardzo wierząca/y |
| **Nominalna objaśniająca** | Region | region | 1 – NO01  2 – NO02  3 – NO03  4 – NO04  5 – NO05  6 – NO06  7 – NO07 |
| **Porządkowa objaśniająca** | health | zdrowie | 1 – Bardzo dobre  2 – Dobre  3 – Umiarkowane  4 - Złe |
| **Porządkowa objaśniająca** | stfdem | satysf\_dem | 1 – Mała  2 - Średnia  3 - Duża |
| **Porządkowa objaśniająca** | atchctr | przyw\_kraj | 1 – Małe  2 - Średnie  3 - Duże |
| **Porządkowa objaśniająca** | hinctnta | doch\_gosp | 1 – Niski  2 – Średni  3 - Wysoki |

Tabela 3. Zestawienie wartości przyjmowanych przez zmienne wykorzystane w modelu.

## Analiza zmiennej objaśnianej

Wybraną zmienną objaśnianą jest poziom szczęścia, która odpowiada na pytanie jak bardzo respondent czuje się szczęśliwy. W zbiorze danych poziom szczęścia został określony jako zmienna *poziom\_szczescia*.

Zmienna ta nie została wybrana przypadkowo. Norwegia to jeden z najszczęśliwszych narodów na świecie. Norweska metoda na szczęście zawarta jest w jednym słowie *hygge*, które jest nieprzetłumaczalne na język polski. Hygge oznacza ogólne poczucie błogości spokoju i relaksu. To między innymi przyjemna atmosfera w otoczeniu rodziny i przyjaciół. Moda na hygge została rozpowszechniona głównie przez Duńczyków[[7]](#footnote-7).

Natomiast słowo ma pochodzenie norweskie i wywodzi się od staronordyckiego słowa hyggja, co oznacza „myśleć” oraz „być z czegoś zadowolonym”. Obecnie wiele krajów naśladuje te narody i wykorzystuję metodę relaksowania się po skandynawsku. To właśnie kraje skandynawskie zajmują czołowe miejsca w rankingu „World Happiness Report”[[8]](#footnote-8).

Dodatkowo szczęście od lat jest dążeniem ludzkości. Jest ono analizowane przez wielu badaczy, a każdy odbiera szczęście w inny sposób w zależności od osobowości. Każda dziedzina nauki skupia się na innych determinantach szczęścia i podaje różne definicje. Psychologowie skupiają szczęście na osobowości, biolodzy analizują procesy neurofizjologiczne, a filozofowie zwracają uwagę na moralność. Szczęście wciąż pozostaje centrum zainteresowania. Z powyżej opisanych przyczyn postanowiono przyjrzeć się bliżej tej zmiennej i przeanalizować czy wybrane zmienne objaśniające rzeczywiście wpływają na szczęście[[9]](#footnote-9).

W ankiecie respondenci oceniali swoje poczucie szczęścia w skali od 0 do 10, przy czym 0 określało bardzo nieszczęśliwą osobę, a 10 osobę bardzo szczęśliwą. Na podstawie Tabeli nr 4 oraz Wykresu nr 1, które przedstawiają odpowiednio częstość i liczebność oraz rozkład zmiennej *poziom\_szczescia* można wnioskować, że Norwegowie to naród szczęśliwy. Histogram przedstawia rozkład zmiennej, który jest lewostronnie asymetryczny. Największe wartości uzyskano dla 3 najwyższych kategorii 8-10, wskazało je aż 75,63% respondentów. Medianą oraz dominantą okazała się kategoria 8, która została wybrana przez 490 respondentów. Drugą kategorią o największej liczebności okazała się kategoria 9. Najniższe kategorie uzyskały liczebności poniżej 10.

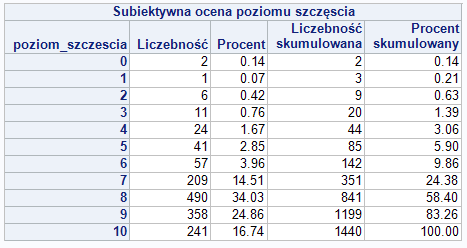
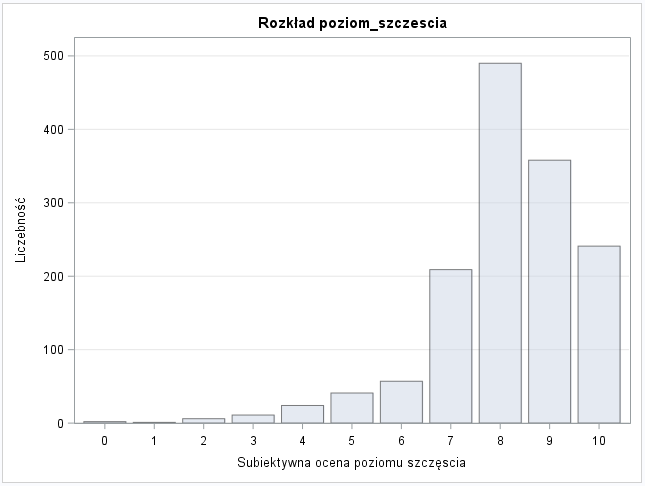


Tabela 4. Liczebność i częstość zmiennej poziom\_szczescia.



Wykres 1. Histogram zmiennej poziom\_szczescia.

Do dalszego badania zmiennej w regresji logistyczną dokonano dychotomizacji zmiennej poziom\_szczescia, tak aby wskazywała ona tylko dwie wartości: dla 0 osobę nieszczęśliwą oraz dla 1 osobę szczęśliwą. Aby nie zniekształcić ogólnego rozkładu zmiennej, do nieszczęśliwych zaliczono te osoby, które oceniły swój poziom szczęścia w kategoriach od 0 do 7 włącznie, pozostałe kategorie będą odzwierciedlały szczęśliwych obywateli Norwegii. Po kategoryzacji otrzymujemy dwie grupy o liczebnościach 351 – osoby nieszczęśliwe (24,38%) oraz 1089 – osoby szczęśliwe (75,63%).

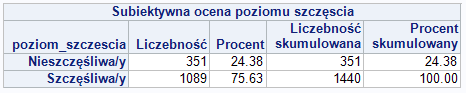
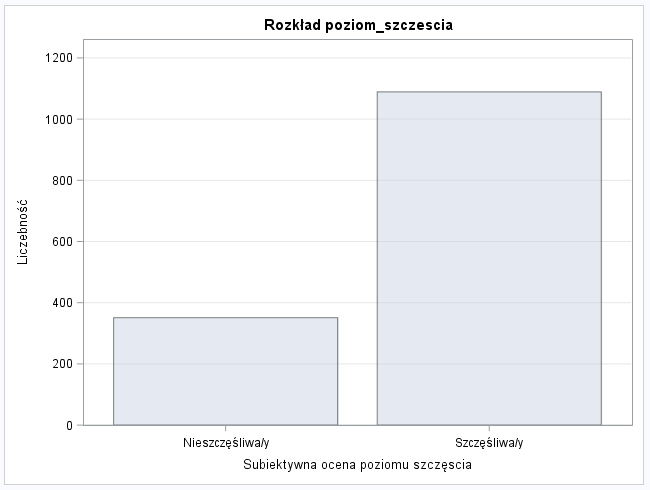


Tabela 5. Liczebność i częstość zmiennej poziom\_szczescia po dychotomizacji.



Wykres 2. Histogram zmiennej poziom\_szczescia po dychotomizacji.

## Analiza zmiennych objaśniających

### Zmienne ciągłe

**Zmienna osob\_gosp**

Liczba osób w gospodarstwie domowym ma bezpośredni wpływ na stan dobrobytu - im większe gospodarstwo domowe, tym mniejszy dochód na osobę, a więc mniejszy dobrobyt. Potencjalny wpływ tej zmiennej na odczuwany poziom szczęścia jest zatem intuicyjny.

Należy jednak spojrzeć też na inne, pozamaterialne aspekty posiadania dużej liczby osób w domostwie - z jednej strony można liczyć na większe wsparcie psychologiczne w razie posiadania problemów, a z drugiej jest to potencjalne podłoże do większej liczby konfliktów. Duża rodzina to również większa ilość obowiązków. Jednocześnie samotność powoduje oczywisty spadek poczucia szczęścia[[10]](#footnote-10).

Jak wskazuje portal Daily Mail osoby posiadające 2-3 dzieci odczuwają wyższy poziom szczęścia niż single i rodziny z 1 dzieckiem[[11]](#footnote-11).

Poniżej zaprezentowano rozkład badanej cechy w ujęciu tabelarycznym jak i na wykresie.

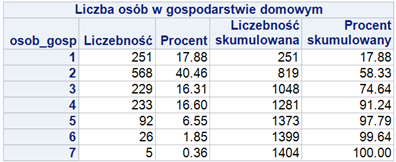
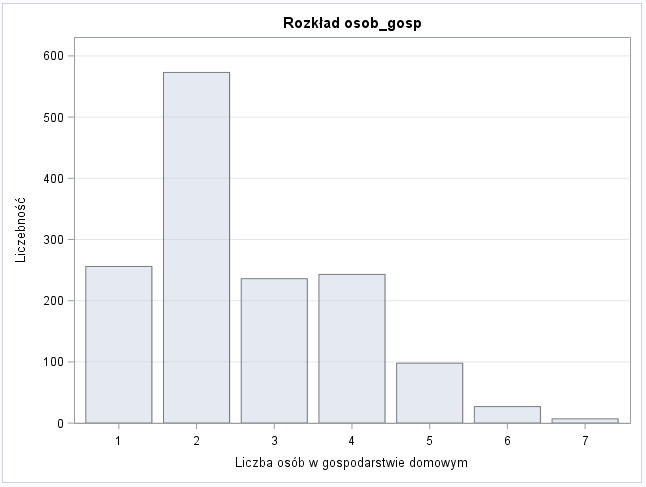


Tabela 6. Liczebność i częstość zmiennej osob\_gosp.



Wykres 3. Histogram dla zmiennej osob\_gosp.

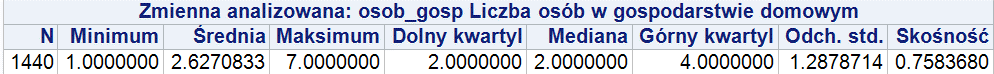
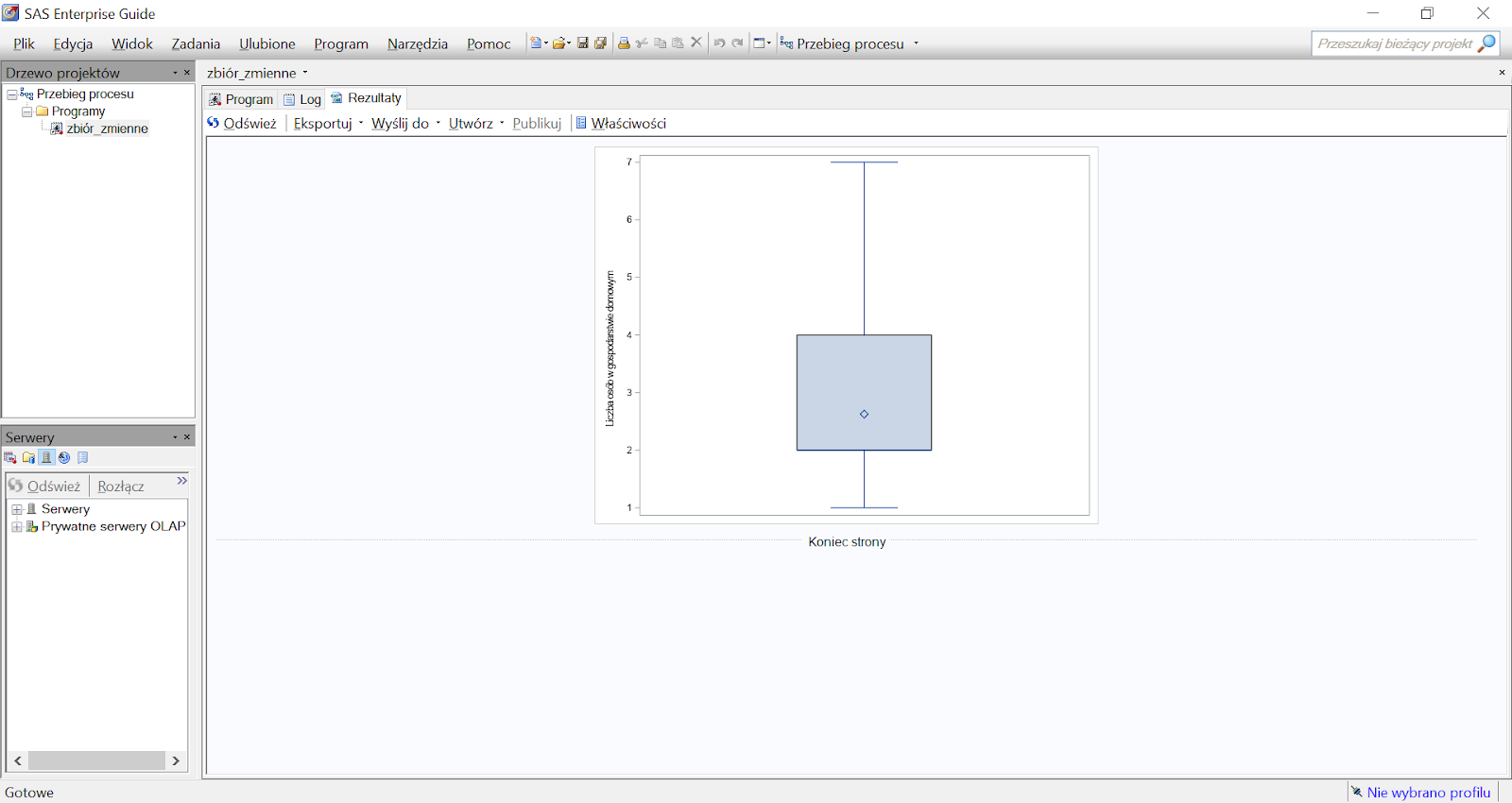


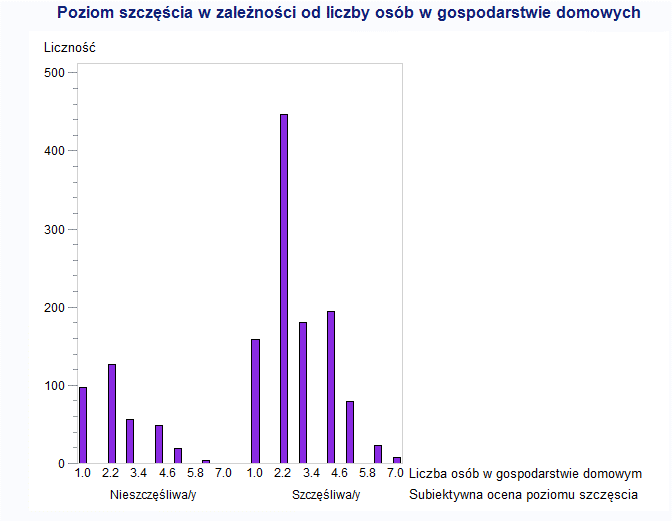
Tabela 7. Statystyki dla zmiennej osob\_gosp.



Wykres 4. Wykres pudełkowy dla zmiennej osob\_gosp.

Ze względu na spodziewane istotne różnice we wpływie na badane zjawisko pomiędzy poszczególnymi kategoriami, zespół projektowy zdecydował się potraktowanie tej zmiennej w charakterze ilościowym. Zdecydowana większość Norwegów posiada 2-osobowe gospodarstwa domowe, a ponad 90% posiada do 3 domowników. Średnia liczba domowników wyniosła 3 osoby. Odchylenie standardowe wyniosło 1 osobę a współczynnik skośności wynoszący 0,76 wskazuje na silną prawostronną asymetrię. Powyższe fakty potwierdza Wykres nr 4 ilustrujący skoncentrowanie wartości zmiennej w jej niższych kategoriach. Przedstawiony rozkład to klasyczny model społeczeństwa kraju rozwiniętego gospodarczo, gdzie ludzie nie dążą do zakładania dużych rodzin.

Wygenerowano również histogram z podziałem na skategoryzowaną zmienną objaśnianą. Wykres pokazuje, że znaczna większość osób zadeklarowała, że są szczęśliwi. Najwięcej gospodarstw składa się z dwóch osób.



Wykres 5. Histogram dla zmiennej osob\_gosp przy skategoryzowanej zmiennej poziom\_szcescia.

**Zmienna godz\_pracy**

Zmienna *godz\_pracy* jest zmienną ciągłą, wskazuje na liczbę godzin poświęconych w tygodniu na pracę przez respondenta.

Według literatury liczba godzin pracy w tygodniu wpływa nie tylko na produktywność pracownika, ale też na jego poczucie szczęście. Maksymalną zalecaną tygodniową liczbą godzin pracujących według specjalistów jest 50 godzin - ze względu na drastyczny spadek produktywności badacze nie zalecają przekraczania tej granicy. Duża liczba godzin pracy w tygodniu wpływa także na małą ilość czasu wolnego, wysoki poziom stresu i przemęczenie. Według badań, zalecana dzienna liczba godzin poświęconych pracy powinna wynosić 7,6 h. Według badaczy jest to wartość graniczna, powyżej której statystycznie pracownicy deklarują niewielką ilość czasu wolnego, który mogą poświęcić na aktywności pozazawodowe. Bardzo podobną liczbę godzin wynosi tydzień pracy w Danii - według World Happiness Report jednego z najszczęśliwszych państw na świecie. Duńczycy pracują przez 38 godzin w tygodniu, a więc jedynie godzinę dłużej niż rekomendują naukowcy. Podobne statystyki występują w przypadku Norwegii i pozostałych państw skandynawskich[[12]](#footnote-12).

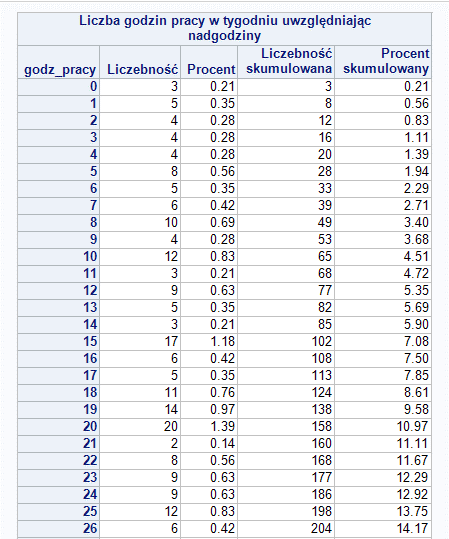
Wartość zmiennej uwzględnia ewentualne nadgodziny, jeśli respondent zadeklarował takowe. Maksymalna możliwa wartość zmiennej to 126 godzin, czyli całkowita liczba godzin w tygodniu. Zmienna nie została skategoryzowana.

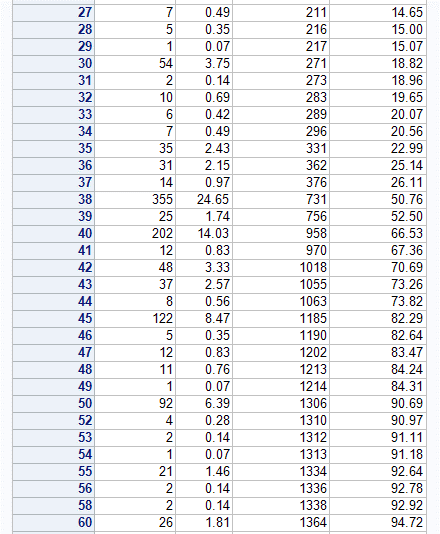


Tabela 8. Statystyki dla zmiennej godz\_pracy.

Średnia liczba przepracowanych godzin w tygodniu wyniosła trochę ponad 37 h. Mediana wyniosła natomiast 38 h. Fakt, iż mediana zmiennej jest niższa od jej średniej świadczy o lewostronnej asymetrii rozkładu zmiennej *godz\_pracy*. Odchylenie standardowe wyniosło niecałe 14 godzin. Świadczy to o dużej liczbie obserwacji odstających, gdzie respondenci deklarowali bardzo dużą liczbę przepracowanych godzin w tygodniu lub gdzie nie deklarowali aktywności zawodowej.

Poniższa tabela prezentuje rozkład kategorii dla wartości omawianej zmiennej. Poniższy wykres wskazuje liczebności i częstości dla kategorii omawianej zmiennej.





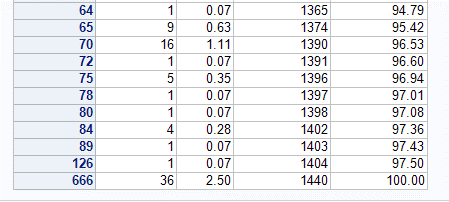
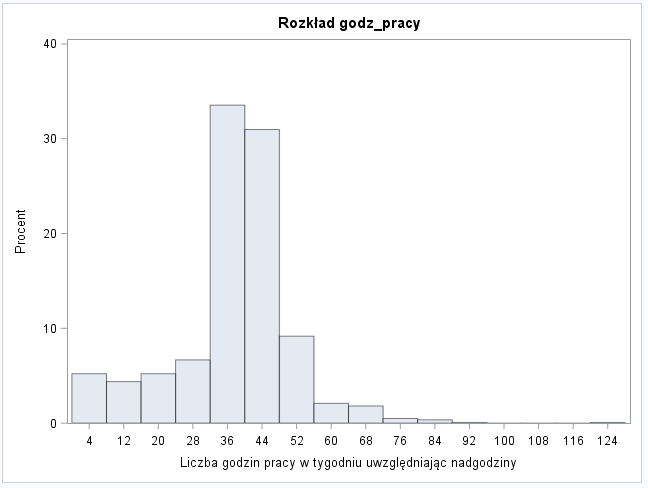
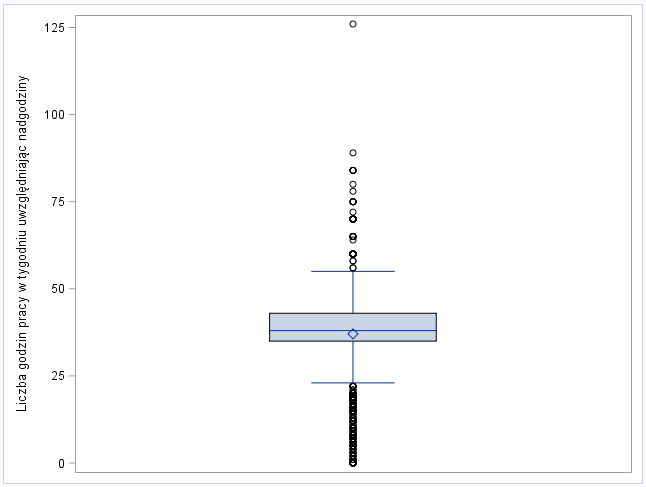


Tabela 9. Liczebność i częstość zmiennej godz\_pracy.



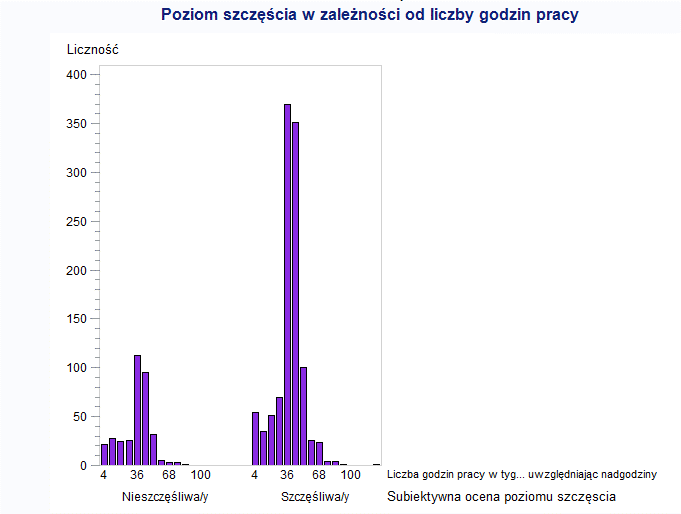
Wykres 6. Histogram dla zmiennej godz\_pracy.

Ze względu na znaczące różnice w liczebnościach dla poszczególnych wartości zmiennej, zespół projektowy zdecydował się na potraktowanie tej zmiennej w charakterze ilościowym. Z wykresu Nr 6 można odczytać, że zdecydowana większość Norwegów w ciągu tygodnia pracuje od 35 do 43 godzin. W badanej populacji wystąpiły osoby pracujące znacznie więcej niż wynosi ustawowa długość tygodnia pracy. Jedna z ankietowanych osób przepracowała w tygodniu ponad 125 godzin. Wystąpiły także obserwacje, gdzie osoby nie zadeklarowały aktywności zawodowej.



Wykres 7. Wykres pudełkowy dla zmiennej godz\_pracy.

Poniżej przedstawiono histogram z podziałem na zmienną poziom\_szczescia. Najwięcej badanych osób pracuje w tygodniu około 40 godzin.



Wykres 8. Histogram dla zmiennej godz\_pracy przy skategoryzowanej zmiennej poziom\_szcescia.

### Zmienne nominalne

**Zmienna: plec**

Zmienna *plec* określa płeć respondenta. W badaniu przyjmuje wartości 1 i 2, które wskazują odpowiednio na mężczyznę oraz kobietę. W badaniu nie wystąpiły osoby, które zadeklarowały inne wartości lub odmówiły odpowiedzi na pytanie.

Według badań psychologów płeć odgrywa rolę jako determinant szczęścia i jego poziom zazwyczaj różni się od płci danej osoby. Badania przeprowadzone w różnych państwach i regionach kulturowych, jak np. Stany Zjednoczone i Iran dowodzą, że mężczyźni statystycznie deklarują wyższy poziom odczuwanego szczęścia, niż kobiety na przestrzeni całego życia. Prawidłowość ta jest szczególnie zauważalna w późniejszym okresie życia[[13]](#footnote-13).

Do ciekawych wniosków doszli autorzy artykułu “*The Paradox of Declining Female Happiness*”, Stevenson i Wolfers. Po przeprowadzeniu badań na populacji Stanów Zjednoczonych okazuje się, że w ciągu ostatnich kilku dekad mimo znacznego wzrostu poziomu życia a także społecznego znaczenia kobiet deklarowany poziom szczęścia wśród kobiet obniżył się zarówno pod względem wartości bezwzględnej, jak i względem analogicznego wskaźnika dla mężczyzn[[14]](#footnote-14).

Poniższa tabela prezentuje liczebności i częstości dla kategorii omawianej zmiennej. Z zamieszczonej poniżej tabeli można odczytać, że mężczyźni (830 osób) stanowili 53,72% badanej populacji, podczas gdy kobiety (715 osób) stanowiły 46,28% badanych respondentów. Rozkład zmiennej płeć pozwala sądzić, że próba wykorzystana do badania jest zbilansowana pod kątem proporcji obydwu płci.

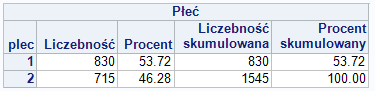
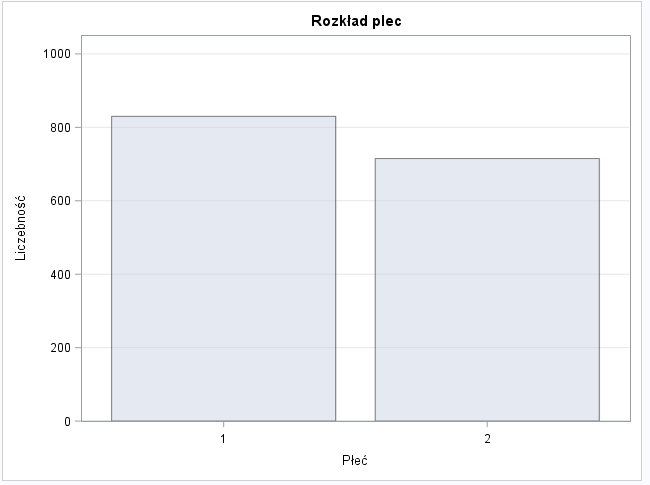


Tabela 10. Liczebność i częstość zmiennej plec.

Poniższy wykres wskazuje liczebności i częstości dla kategorii omawianej zmiennej.



Wykres 9. Histogram dla zmiennej plec.

Tabela nr 11 zawiera tablicę kontyngencji dla zmiennych *poziom\_szczescia* oraz *plec*. Macierz kontyngencji wskazuje, że dla obu płci większość stanowią osoby szczęśliwe. Dalsza analiza tablicy pozwala wskazać, że większe prawdopodobieństwo bycia szczęśliwymi mają kobiety.

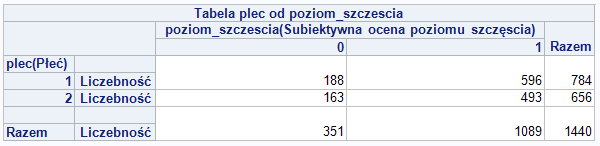


Tabela 11. Tablica kontyngencji dla zmiennej plec.

Po przeanalizowaniu i przeprocesowaniu zbioru danych wartości dla mężczyzn zostały przemianowane na 0, a dla kobiet na 1. Postanowiono pozostawić zmienną do obliczeń w obecnej postaci po przemianowaniu.

**Zmienna wiek**

Jedną ze zmiennych, które postanowiono włączyć do modelu jest wiek. Przypuszcza się, że subiektywny poziom odczuwanego szczęścia lub jego brak zależy od wieku, w jakim znajdują się respondenci. Przypuszcza się, że osoby młode mogą wskazywać na relatywnie wyższy poziom szczęścia, gdyż zazwyczaj cechują się dobrym stanem zdrowia, nie mają stałych zobowiązań, takich jak na przykład praca czy odpowiedzialność za wychowanie dzieci. Ludzie młodzi są ciekawi świata, otwarci na nowe wyzwania, a także łatwo przystosowują się do nowych sytuacji. Z wiekiem, gdy liczba obowiązków i poziom odpowiedzialności wzrastają subiektywny poziom odczuwalnego szczęścia może się obniżać. Na taki stan rzeczy może wpływać również pogarszające się z wiekiem zdrowie.

Powyższa teza znajduje również podparcie w przeglądzie literatury, którego dokonano w celu identyfikacji możliwych determinant szczęścia ludzi. Studia nad determinantami szczęścia doprowadziły do opublikowania w 2004 r. pracy Blanchflowera i Oswalda, którzy dowiedli, że zależność satysfakcji z życia od wieku jest U-kształtna. Najniższy poziom szczęścia jest wskazywany przez osoby w wieku około 40 lat. Powyższy wzorzec został potwierdzony w badaniu z 2008 r. przeprowadzonym na danych z wielu krajów[[15]](#footnote-15).

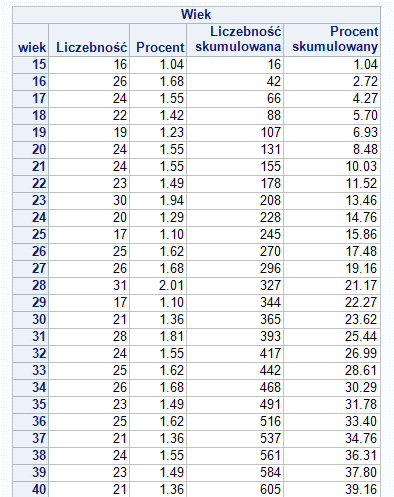
Krzywa odzwierciedlająca poziom szczęścia ilustruje, że od około 20 roku życia poczucie szczęścia powoli spada osiągając minimalny poziom w wieku około 40 lat. W okolicach 50-tego roku życia poziom szczęścia zaczyna ponownie rosnąć. Wzorzec ten był najlepiej widoczny w krajach zamożniejszych. Powyższe argumenty skłoniły zespół projektowy do włączenia zmiennej wiek do analizy[[16]](#footnote-16).

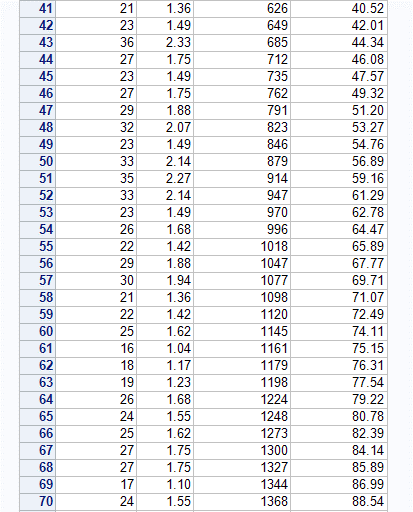
Zmienna określająca *wiek* respondentów jest zmienną ilościową ciągłą. Tabela nr 12 zawiera podstawowe statystyki opisowe zmiennej wiek. Najmłodszy respondent miał 15 lat, natomiast najstarszy 98 lat. Zarówno mediana, jak i średnia wynoszą ok. 47 lat. Wartość dolnego kwartyla wskazuje, że 25% respondentów ma 32 lata lub mniej. Natomiast wartość górnego kwartyla wskazuje, że 75 % respondentów ma 61 lub mniej lat. Miara skośności na poziomie 0,18 wskazuje na słabą prawostronną asymetrię rozkładu zmiennej.



Tabela 12. Statystyki opisowe dla zmiennej wiek.

Poniższa tabela przedstawia częstość i liczebność zmiennej wiek.





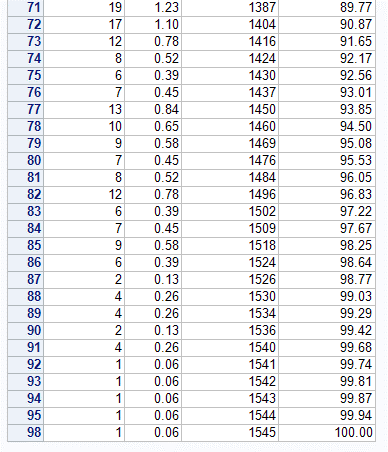
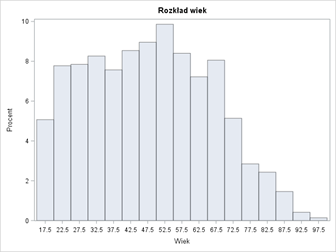


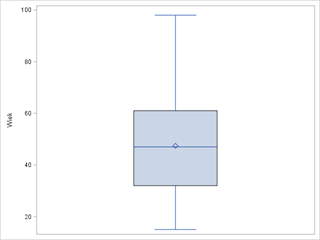
Tabela 13. Liczebność i częstość zmiennej wiek.

Poniższy histogram przestawia rozkład zmiennej wiek po zgrupowaniu jej w kategorie w celu poprawy czytelności rozkładu.

**

Wykres 10. Histogram dla zmiennej wiek.

Wykres Nr 10 przedstawia histogram zmiennej *wiek*, który pozwala zaobserwować prawostronną skośność zmiennej. Najliczniejszą grupą respondentów były osoby w wieku od 50 do 55 lat. Najmniej liczną grupą były osoby w wieku od 95 lat.



Wykres 11. Wykres pudełkowy dla zmiennej wiek.

Wykres Nr 11 przedstawia wykres pudełkowy dla zmiennej *wiek*. Długość prostokąta reprezentuje rozstęp ćwiartkowy. Zatem 50% środkowych obserwacji znajduje się w przedziale od 32 do 61 lat. Pozioma linia rozdzielająca pudełko wyznacza wartość mediany na poziomie 47 lat. Wąsy natomiast łączą pudełko z odpowiednio wartością minimalną i maksymalną. Dłuższy wąs łączący pudełko z wartością maksymalną wskazuje na prawostronną skośność rozkładu zmiennej.

W kolejnym kroku dokonano rekategoryzacji zmiennej wiek. Decyzję oparto na wcześniejszym przeglądzie literatury, z którego wynika, że przynależność do poszczególnych grup wiekowych determinuje poziom odczuwalnego szczęścia. Respondentów podzielono na 5 kategorii wiekowych zaprezentowanych poniżej:

* 1 – 22 lata i mniej
* 2 – 23-30 lat
* 3 – 31-45 lat
* 4 – 46-66 lat
* 5 – powyżej 67 lat

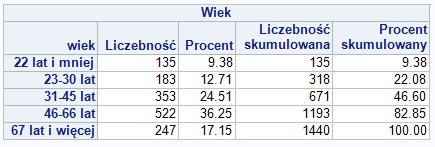
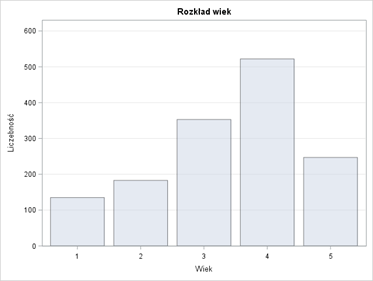


Tabela 14. Liczebność i częstość zmiennej wiek po kategoryzacji.

Najliczniejszą grupę respondentów stanowią Norwegowie w wieku od 46-66 lat (36,25%). Najmniej liczną grupę stanowią osoby w wieku poniżej 22 lat (9,38%). Respondenci w wieku 23-30 lat stanowią 12.71% próby, w wieku 31-45 lat stanowią 24,51%, a w wieku powyżej 67 lat 17,15%.



Wykres 12. Histogram dla zmiennej wiek.

Tabela nr 15 zawiera tablicę kontyngencji dla zmiennych poziom\_szczęścia oraz wiek. Macierz kontyngencji wskazuje, że we wszystkich grupach wiekowych większość stanowią osoby szczęśliwe.

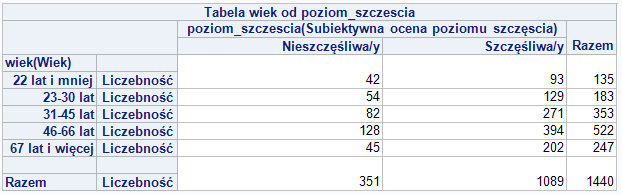
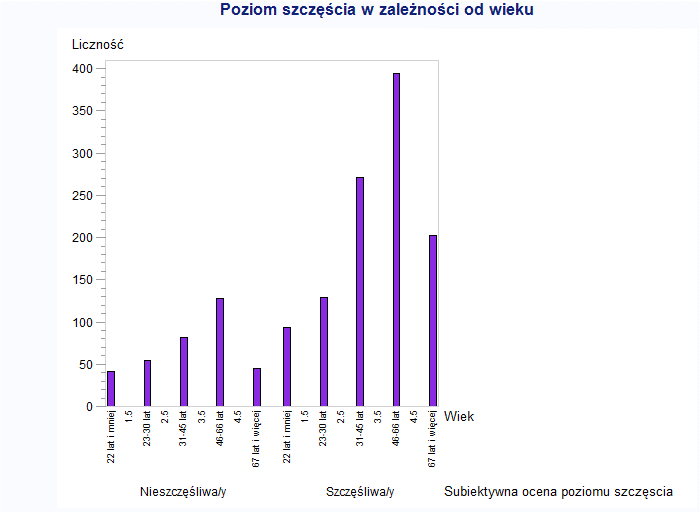


Tabela 15. Tabela kontyngencji dla zmiennych poziom\_szczescia i wiek.

Dla nowych kategorii przedstawiono liczebności w postaci histogramu z podziałem na binarną zmienną poziom\_szczescia. Wykres przedstawia znaczną różnicę w liczebności osób szczęśliwych i nieszczęśliwych szczególnie dla osób starszych.



Wykres 13. Histogram dla zmiennej wiek przy skategoryzowanej zmiennej poziom\_szcescia.

**Zmienna przelozony**

Uzasadnienie dla wyboru tej zmiennej do modelu możemy znaleźć w artykule Anny Strużyńskiej-Kujałowicz ‘’Czy władza szczęście daje? Władza a nastrój, emocje i satysfakcja z życia’’. Autorka w swoim tekście odwołuje się do tez Fiszke i Keltnera, którzy twierdzą, iż możliwość wywierania wpływu na innych jest zwykle postrzegana jako czynnik wpływający na wzrost zadowolenia. Oznacza to, że nadzór sprawowany nad podwładnymi może przyczynić się do poczucia szczęścia. Z kolei zgodnie z koncepcją Keltnera i Andersona sprawowanie władzy co oznacza zajmowanie nadrzędnej pozycji w hierarchii zwiększa szanse realizacji celów i zamierzeń, a to może przyczyniać się do odczuwania zadowolenia i poczucia szczęścia.

Wyniki badań, na które powołuje się autorka w artykule wskazują na fakt, iż osoby sprawujące władzę doświadczają bardziej pozytywne emocje niż ci nad którymi władza jest sprawowana[[17]](#footnote-17).

Zespół projektowy postanowił zatem sprawdzić, czy nadzór nad pracownikami wpływa na poczucie szczęścia.

Poniższa tabela oraz wykres przedstawiają oryginalny podział na osoby, które mają nadzór (1), nie mają nadzoru nad innymi pracownikami (2) oraz na odpowiedzi „Nie dotyczy” (6).

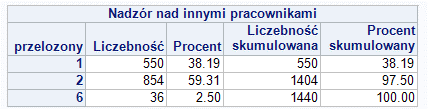
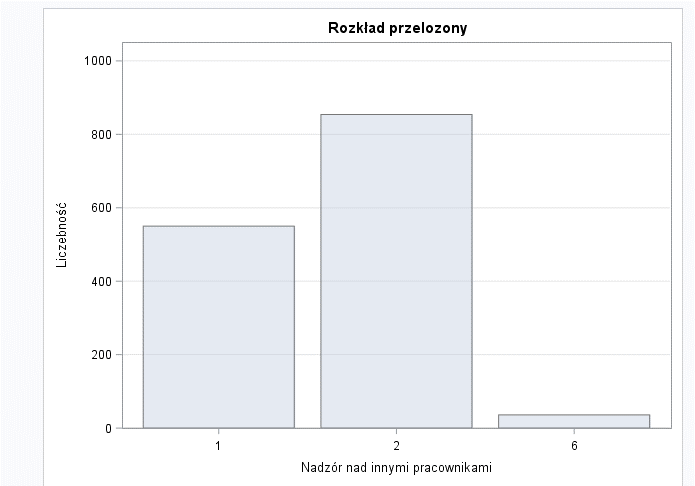


Tabela 16. Liczebność i częstość zmiennej przelozony przed kategoryzacją.



Wykres 14. Histogram dla zmiennej przelozony przed kategoryzacją.

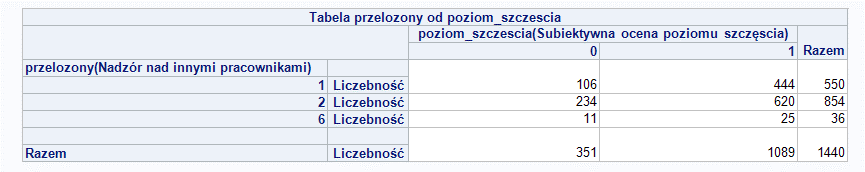


Tabela 17. Tabela kontyngencji dla zmiennych poziom\_szczescia i przelozony przed kategoryzacją.

Postanowiono kategorię ‘nie dotyczy’ włączyć do kategorii ‘nie posiada nadzoru’, co przedstawia poniższa tabela.

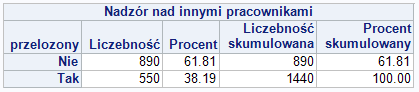
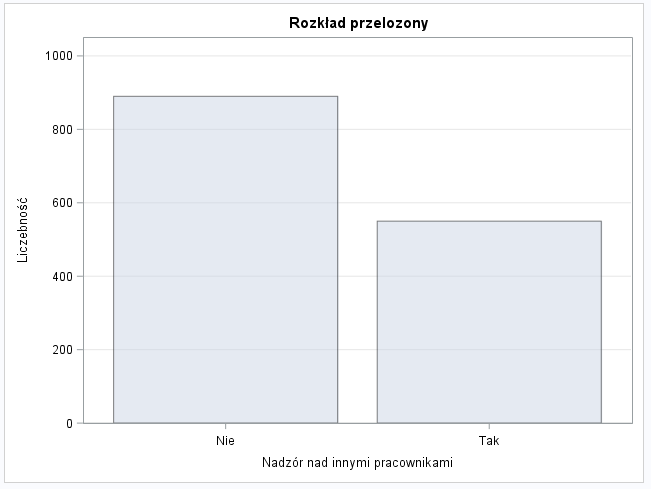


Tabela 18. Liczebność i częstość zmiennej przelozony po kategoryzacji.

Z danych zaprezentowanych w Tabeli nr 18 wynika, iż większość badanych nie posiada nadzoru nad pracownikami (61,81%). Nie odnotowano jednak znaczącej dysproporcji w liczebności pomiędzy tymi grupami, więc zmienna nie wymaga dalszej modyfikacji.



Wykres 15. Histogram dla zmiennej przelozony.

Potwierdza to również zależność kontyngencji danej zmiennej ze zmienną objaśnianą przedstawiona w Tabeli nr 19 z której wynika, że dla każdej kombinacji kategorii obu zmiennych mamy wystarczającą liczbę obserwacji.

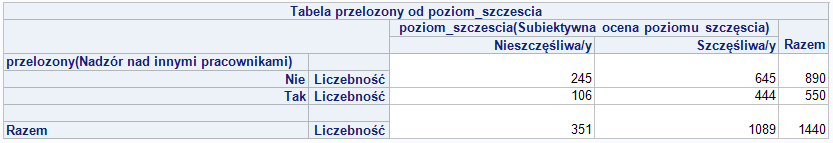


Tabela 19. Tablica kontyngencji dla zmiennych poziom\_szczescia i przelozony po kategoryzacji.

**Zmienna: typ\_zakladu**

Jako że na pracę zawodową ludzie przeznaczają sporą część swojego czasu, jest ona istotnym komponentem wpływającym na szczęście jednostki. Zespół projektowy zdecydował się na włączenie zmiennej typ zakładu pracy do analizy ze względu na znalezione przesłanki w literaturze.

W pracy „Happiness and Work” z 2014 r. wskazano na następujące wzorce:

* samozatrudnienie wpływa na deklarowanie większej satysfakcji życiowej, co jest prawdopodobnie związane z większą autonomią,
* praca w sektorze publicznym również wiąże się ze wzrostem satysfakcji.[[18]](#footnote-18)

W analizowanym badaniu wartości zmiennej typ zakładu pracy (*typ\_zakladu*) zostały zebrane w wyniku odpowiedzi respondentów na pytanie: Dla jakiego typu organizacji pracujesz/pracowałeś?

Możliwe odpowiedzi po usunięciu braków danych (odmowa odpowiedzi, nie wiem, brak odpowiedzi) przedstawiają się następująco:

* 1 - Lokalne lub centralne jednostki rządowe
* 2 - Inne jednostki sektora publicznego (np. służba zdrowia, jednostki edukacyjne)
* 3 - Przedsiębiorstwa należące do państwa
* 4 - Firmy prywatne
* 5 - Samozatrudnienie
* 6 - Inne
* 66 - Nie dotyczy

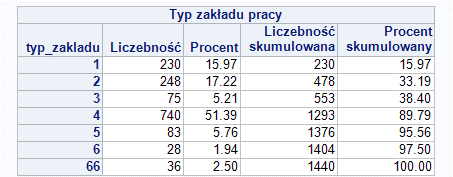
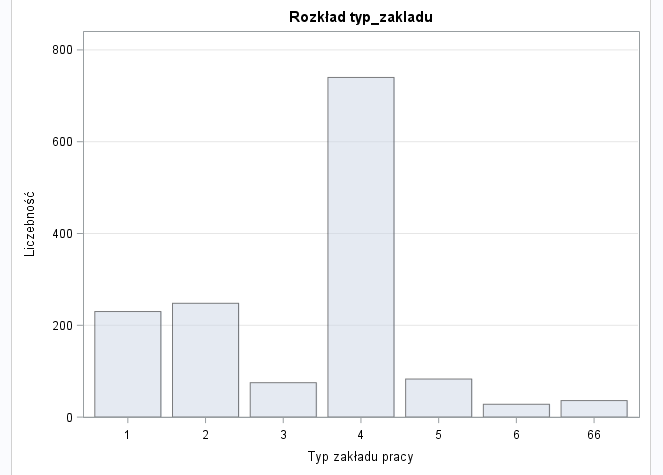
**

Tabela 20. Liczebność i częstość zmiennej typ\_zakladu przed kategoryzacją.

Rozkład zmiennej w badanej próbie przedstawia Wykres nr 16. Najliczniejszą grupą są osoby pracujące w firmach prywatnych, stanowią one ok. 51% respondentów. Następnie grupy osób pracujących w innych jednostkach sektora publicznego oraz w lokalnych lub centralnych jednostkach rządowych stanowią odpowiednio 17% oraz 16% respondentów. Najmniej osób wskazało odpowiedź ‘Inne’- ok. 2% respondentów.

**

Wykres 16. Histogram dla zmiennej typ\_zakladu.

Następnie sporządzono tablicę kontyngencji celem zbadania łącznych rozkładów zmiennej poziom szczęścia oraz typ zakładu pracy (Tabela nr 21). Analiza tablicy kontyngencji daje podstawy do połączenia dwóch pierwszych kategorii - lokalne lub centralne jednostki rządowe oraz inne jednostki sektora publicznego. Ponadto w analizowanej literaturze dotyczącej zależności poziomu szczęścia od typu zakładu pracy wskazuje się bardziej ogólny podział - brak podgrup sektora publicznego. Podjęto również decyzję o dalszej rekategoryzacji zmiennej. Ostatecznie połączono trzy pierwsze kategorie. Połączono również kategorie ‘Inne’ oraz ‘Nie dotyczy’, gdyż mogą obejmować one ludzi bezrobotnych, którzy nigdy nie podejmowali pracy oraz szarą strefę.

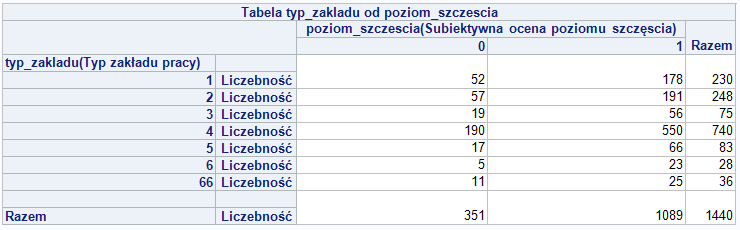
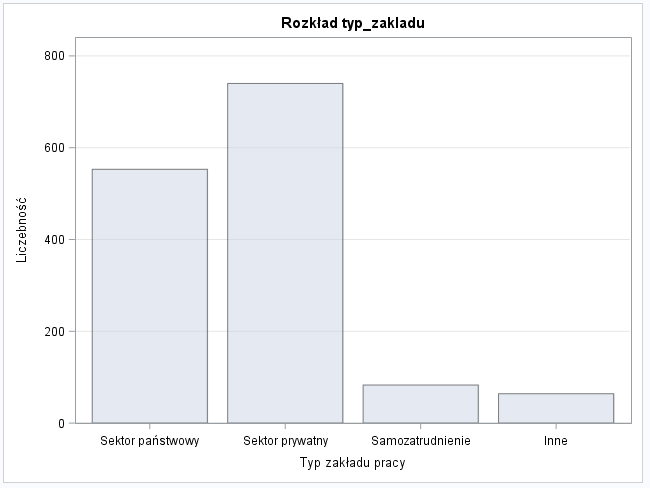


Tabela 21. Tablica kontyngencji dla zmiennych poziom\_szczescia i typ\_zakladu.

Po przeprowadzeniu rekategoryzacji nowe kategorie zmiennej przedstawiają się następująco:

* 1 - Sektor państwowy
* 2 - Sektor prywatny
* 3 - Samozatrudnienie
* 4 - Inne

Rozkład zmiennej typ zakładu pracy po rekategoryzacji przedstawia Wykres nr 17 oraz Tabela nr 22.



Wykres 17. Rozkład zmiennej typ\_zakladu po rekategoryzacji.

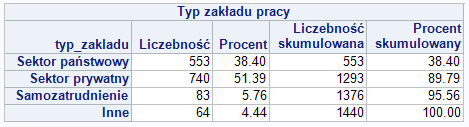


Tabela 22. Liczebność i częstość zmiennej typ\_zakladu po rekategoryzacji.

Najliczniejszą grupą respondentów są osoby zatrudnione w sektorze prywatnym, natomiast najmniej liczną jest kategoria Inne. Norwegia jest krajem wysoko rozwiniętym stąd taki wynik jest zgodny z intuicją. W następnym etapie zbudowano tablicę kontyngencji dla zmiennych typ zakładu pracy oraz subiektywnie odczuwany poziom szczęścia (Tabela nr 23).

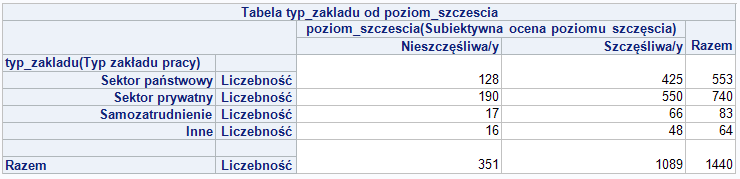


Tabela 23. Tablica kontyngencji dla zmiennych poziom\_szczescia i typ\_zakladu.

**Zmienna: lata\_edu**

Jedną z ciągłych zmiennych wybranych do modelu jest zmienna *lata\_edu* przedstawiająca liczbę lat edukacji. Z artykułu J. Hartoga i H. Oosterbeeka pt. „Health, wealth, and happiness: why pursue a higher education?” dowiadujemy się, że edukacja jest pozytywnie skorelowana ze szczęściem tylko do pewnego stopnia. Wyniki badań wykazały paraboliczny efekt edukacji na poziom szczęścia. Najwyższy stopień naukowy nie wpływa pozytywnie ani na wysokość dochodów, ani na poziom szczęścia. W krajach rozwinięty efekt wpływu liczby lat edukacji na szczęście może mieć nawet negatywny efekt[[19]](#footnote-19).

Według Tatiany Kanasz wykształcenie daje większe korzyści w krajach mniej rozwiniętych. W swojej pracy dodaje również, że odnotowano związek edukacji i poziomu szczęścia dla grup o niskich dochodach w krajach rozwiniętych oraz dla krajów biednych[[20]](#footnote-20).

Zdecydowano sprawdzić prawdziwość powyżej opisanych tez.

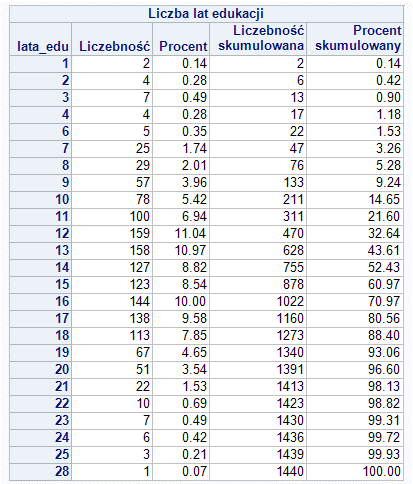
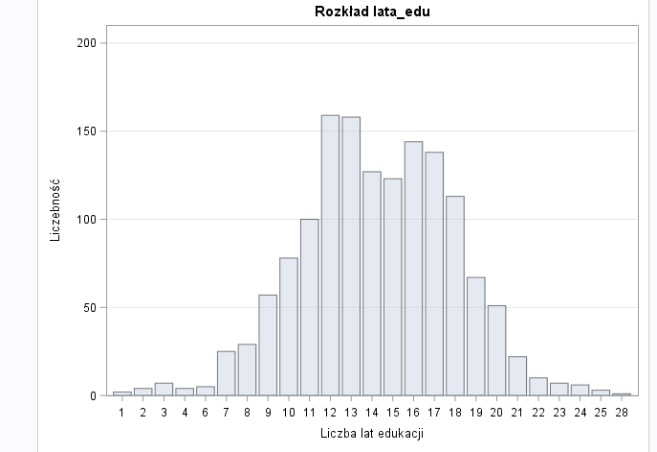
****

Tabela 24. Liczebność i częstość zmiennej lata\_edu.

****

Wykres 18. Histogram dla zmiennej lata\_edu.

Na podstawie histogramu można stwierdzić, że zmienna ma symetryczny rozkład. W badaniu wzięło udział dwóch respondentów o rocznym doświadczeniu oraz jeden respondent o maksymalnej liczbie lat edukacji równej 28. Średnio osoby uczestniczące w badaniu zadeklarowały 14 lat edukacji. 50% respondentów zadeklarowało 14 lub mniej lat odbytej edukacji, a 75% respondentów 17 lat lub mniej. Odchylenie standardowe wyniosło 3,7. Współczynnik asymetrii wyniósł -0,16 co wskazuje na słabą lewostronną asymetrię.

Dodatkowo wygenerowano wykres pudełkowy, który wskazuje na skoncentrowanie obserwacji w środkowych kategoriach. Na wykresie zostały również przedstawione obserwacje odstające dla wartości poniżej 5 oraz powyżej 25 lat.

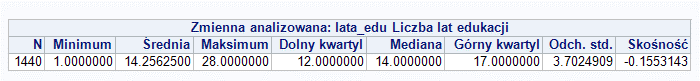
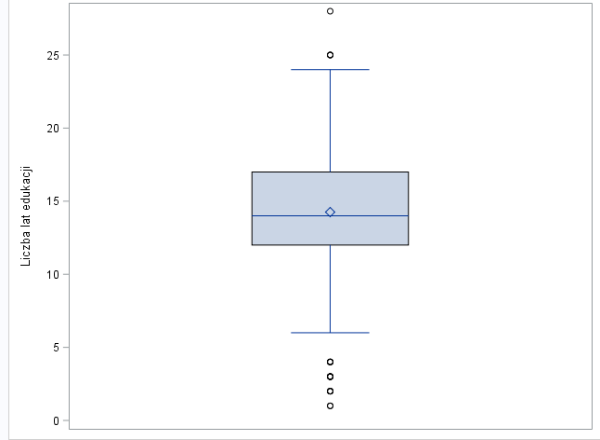


Tabela 25. Statystyki opisowe dla zmienna lata\_edu.



Wykres 19. Wykres pudełkowy dla zmiennej lata\_edu.

Zespół projektowy postanowił dokonać kategoryzacji tej zmiennej na podstawie kwartyli. Poniżej przedstawiono nowe kategorie dla zmiennej lata\_edu.

* 1 – 12 lat i mniej
* 2 – 13-14 lat
* 3 – 15-17 lat
* 4 - 18 lat i więcej

Po kategoryzacji uzyskano liczebności 470, 285, 405 oraz 280 dla kategorii odpowiednio 12 lat i mniej, 13-14 lat, 15-17 lat oraz 18 lat i więcej.

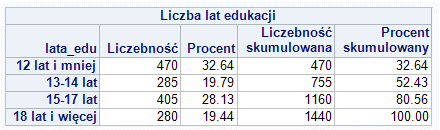
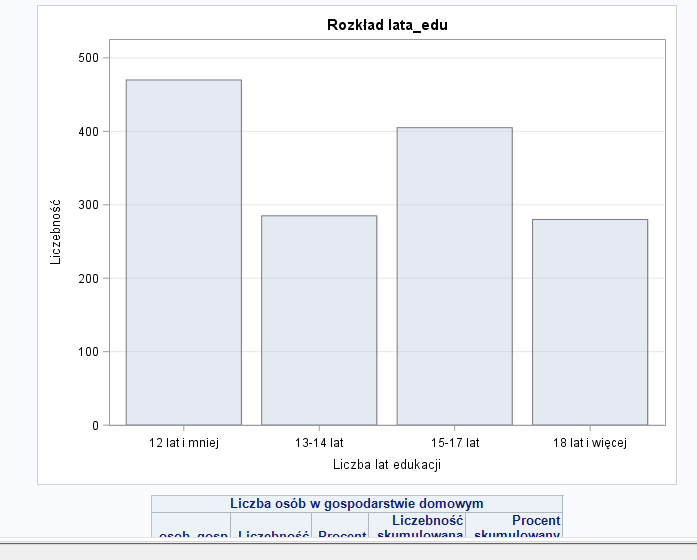


Tabela 26. Liczebność i częstość zmiennej lata\_edu po kategoryzacji.



Wykres 20. Histogram dla zmiennej lata\_edu.

Liczebności w wygenerowanej tablicy kontyngencji dla zmiennej lata\_edu po kategoryzacji są dostateczne dla dalszych obliczeń.

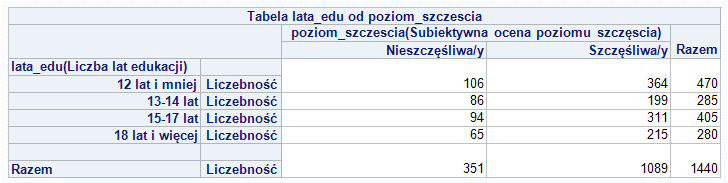
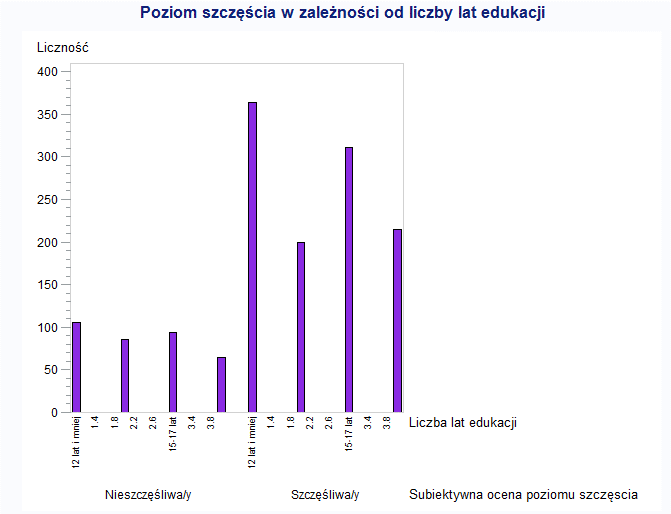


Tabela 27. Tablica kontyngencji dla zmiennej lata\_edu po kategoryzacji.

Dzięki poniższemu histogramowi, można zaobserwować znaczne różnice w poziomie szczęścia dla kategorii o 12 lat i mniej oraz 15-17 lat.



Wykres 21. Histogram dla zmiennej lata\_edu przy skategoryzowanej zmiennej poziom\_szcescia.

**Zmienna: religijnosc**

Kolejną zmienną, którą uwzględniono w modelu jest zmienna *religijnosc*, która pokazuje w jakim stopniu religia jest ważna dla respondenta. Pomimo braku dowodów na to, że religia bardziej wpływa na poczucie szczęścia niż inne czynniki, jest to ciekawa dziedzina ludzkiego codziennego życia. Z pewnością religia ma większe znaczenie dla wierzących, dodatkowo również dla kobiet niż mężczyzn oraz dla osób starszych niż młodych. Bardzo często ludzie odnajdują w wierze nadzieję na lepsze życie oraz czują wsparcie wśród społeczności osób wierzących. Dodatkowo modlitwa i medytacja pozwalają na wyciszenie, znalezienie spokoju oraz pomagają w trudnych chwilach.

W badaniu Tomasza Strózika z 2009 r., o którym wspomina Tatiana Kanasz w książce pt. „Uwarunkowania szczęścia”, studenci wierzący deklarowali wyższy poziom szczęścia od tych niewierzących[[21]](#footnote-21).

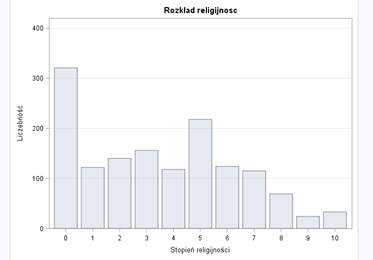
Ważny jest również stosunek do wiary i czy osoba pozostaje praktykująca lub nie. W Norwegii około 70% populacji należy do tzw. Kościoła Norweskiego. Jednak zaledwie kilka procent osób uczestniczy w nabożeństwach częściej niż raz w miesiącu[[22]](#footnote-22).

Dodatkowo w ostatnich czasach obserwujemy tendencję malejącą w liczbie osób wierzących i równocześnie praktykujących. Z tego względu grupa projektowa zdecydowała na wybranie tej zmiennej oraz oparciu o nią hipotezy czy religia wpływa na poziom szczęścia.

W zbiorze wyjściowym zmienna przyjmowała wartości od 0 do 10, gdzie 0 oznaczało osobę niewierzącą, a 10 osobę bardzo wierzącą. Na podstawie tabeli nr 28 oraz wykresu nr 22 można stwierdzić, że Norwegowie nie są wierzącym narodem. Potwierdzają to wysokie liczebności w kategoriach od 0 do 5, które razem stanowiły 74,65% wszystkich respondentów. Największą liczebność uzyskano dla kategorii 0, która wyniosła 321 respondentów, czyli 22,29% całej próby. Medianą okazała się kategoria 3. Najmniej osób wybrało kategorię 9, tylko 24 respondentów, co stanowiło 1,67% wszystkich osób biorących udział w badaniu.



Tabela 28. Liczebność i częstość zmiennej religijność przed kategoryzacją.



Wykres 22. Histogram dla zmiennej religijność po kategoryzacji.

Wygenerowano również tablicę kontyngencji w celu sprawdzenia liczebności dla poszczególnych kategorii według zmiennej *poziom\_szczescia*. Po krótkiej analizie, zdecydowano się na kategoryzację zmiennej *religijnosc* aby uniknąć niskich liczebności dla najwyższych kategorii, np. 4, 5 dla kategorii 9 i 10 dla osób nieszczęśliwych.

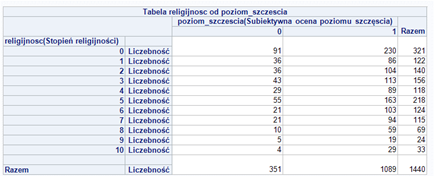


Tabela 29. Tablica kontyngencji dla zmiennej poziom\_szczescia i religijnosc przed kategoryzacją.

Postanowiono stworzyć 3 nowe kategorie dla stopnia religijności: 1 - osoba niewierząca, 2 - osoba wierząca oraz 3 - osoba bardzo wierząca. Do osób niewierzących zaliczono kategorie 0, 1 i 2. Osoby, które wybrały 3, 4, 5 i 6 zostały przydzielone do nowej kategorii osób wierzących. Pozostałych respondentów sklasyfikowano jako bardzo wierzących. Uznano, że taki podział pokaże, czy religia ma wpływ na poziom szczęścia oraz czy stopień zaangażowania podnosi tę subiektywną ocenę. Po kategoryzacji uzyskano liczebności odpowiednio 583, 616 oraz 241.

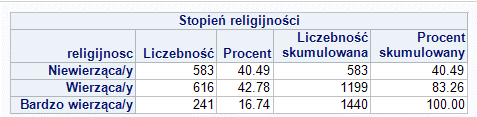
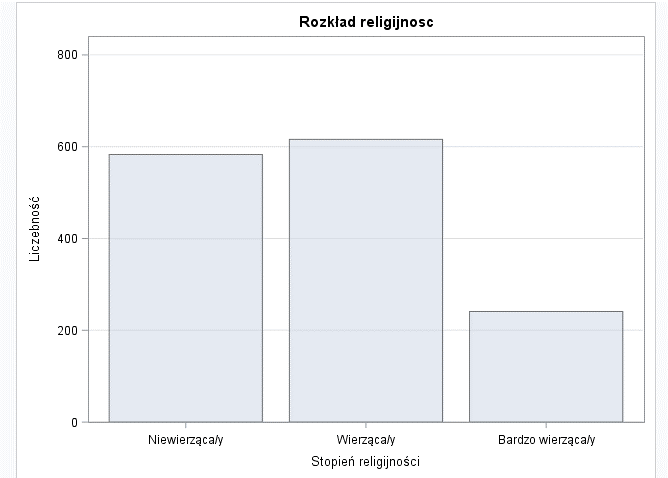


Tabela 30. Liczebność i częstość zmiennej religijnosc po kategoryzacji.



Wykres 23. Histogram dla zmiennej religijnosc po kategoryzacji.

Sprawdzono również, jak wygląda tablica kontyngencji dla nowych kategorii. Wszystkie komórki uzyskały liczebności powyżej 20 obserwacji.

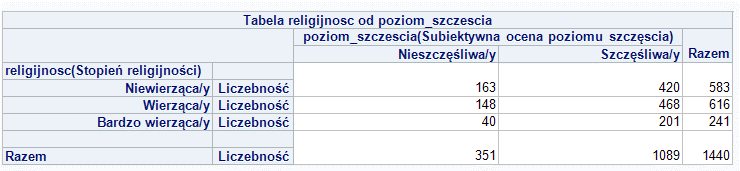
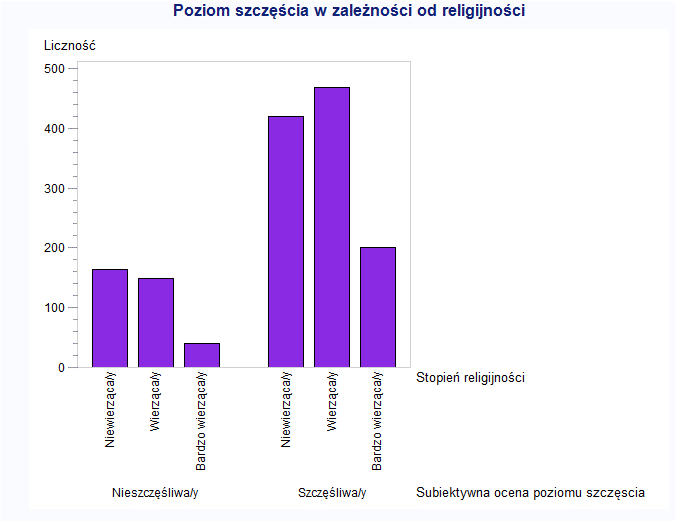


Tabela 31. Tablica kontyngencji dla zmiennej poziom\_szczescia i religijnosc po kategoryzacji.

Zdecydowano również wygenerować histogram dla zmiennej religijnosc, o którą oparta jest hipoteza. Porównywalną ilość osób szczęśliwych odnotowano dla kategorii wierzący i niewierzący, a znacznie mniej dla kategorii bardzo wierzący.



Wykres 24. Histogram dla zmiennej religijnosc przy skategoryzowanej zmiennej poziom\_szcescia.

**Zmienna: region**

Jest to zmienna przypisująca region zamieszkania respondenta na podstawie odpowiednika oficjalnej klasyfikacji statystycznej krajów Unii Europejskiej NUTS2. Powyższy podział ilustruje Rysunek nr 1 znajdujący się poniżej.



Rysunek 1. Podział administracyjny Norwegii na regiony.

W ten sposób Norwegia została podzielona na 7 obszarów, z oznaczonym jako NO01 miastem Oslo i okręgiem stołecznym.

Jak wskazują dane World Population Review w Norwegii występuje bardzo duże zróżnicowanie zagęszczenia ludności - wobec 14,7 os/km^2 w całym kraju w aglomeracji Oslo mieszka ponad 3000 os. /km^2. Ludność skupiona jest głównie w południowej części kraju, gdzie znajduje się też większość ośrodków miejskich. Klimat również jest zróżnicowany - od umiarkowanie chłodnego na południu do subpolarnego na krańcach północy[[23]](#footnote-23).

Powyższe fakty mogą rzutować na zróżnicowany poziom infrastruktury z gorszym jej poziomem na północy kraju. Ponadto mieszkańcy północy kraju zimą cierpią na deficyt światła słonecznego, co skutkuje jak wskazuje portal The Atlantic zapadalnością na

stany depresyjne na poziomie 10-15% populacji[[24]](#footnote-24).

Wobec powyższych czynników można przypuszczać, że w regionach północnych panują trudniejsze warunki do życia.

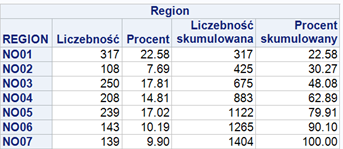
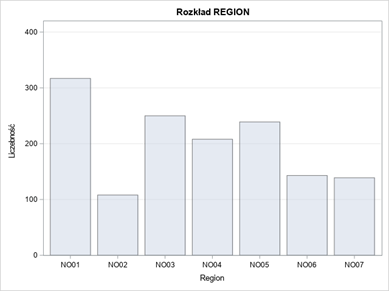


Tabela 32. Liczebność i częstość zmiennej region



Wykres 25. Histogram dla zmiennej region

Największy odsetek odpowiedzi został odnotowany z regionu NO01 - Oslo i okrąg stołeczny- ponad 22%. W pozostałych regionach liczebność respondentów jest równomierna, co odpowiada danym odnośnie zagęszczenia ludności i powierzchni regionów. Dostateczność liczby obserwacji w poszczególnych kategoriach potwierdza Tabela nr 33. Wobec powyższego zespół projektowy nie zdecydował się na ponowną kategoryzację zmiennej.

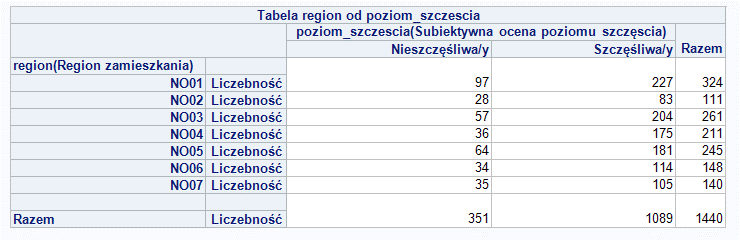


Tabela 33. Tablica kontyngencji dla zmiennej poziom\_szczescia i region.

### Zmienne porządkowe

**Zmienna zdrowie**

Zmienna *zdrowie* wskazuje na zadeklarowany przez respondenta ogólny stan zdrowia. Jest to subiektywna ocena respondenta i nie została poparta faktycznym badaniem stanu zdrowia ankietowanych.

Według World Happiness Report 2019 deklarowany stan zdrowia ma znaczący wpływ na odczuwany przez respondenta poziom szczęścia. Oprócz tego, determinuje on częściowo długość trwania życia, czyli okres, przez jaki człowiek może czerpać z życia satysfakcję i poczucie szczęścia. W badaniach przeprowadzonych w Meksyku respondenci wskazali zdrowie jako 4 najważniejszy faktor determinujący ich poczucie szczęścia. W 10 stopniowej skali określającej wagę kryteriów branych pod uwagę podczas oceny szczęścia danej osoby stanowi zdrowia ankietowani nadali średnią wagę 8,4. Wyżej oceniono tylko relacje z bliskimi, wykonywany zawód oraz poczucie spełnienia. Z tego powodu zespół projektowy uznał istotność tej zmiennej dla modelowania poczucia szczęścia i postanowił ją włączyć do analizy[[25]](#footnote-25).

Zmienna przyjmuje wartości od 1 do 5, które po przypisaniu bardziej intuicyjnych etykiet można rozumieć w następujący sposób:

* 1 - bardzo dobry,
* 2 - dobry,
* 3 - przeciętny,
* 4 - zły,
* 5 - bardzo zły

Podczas badania nie udzielono niepoprawnych odpowiedzi i odpowiedzi o wartościach innych niż powyższe. Nie zanotowano także odmowy odpowiedzi na pytanie.

Poniższa tabela prezentuje rozkład kategorii dla wartości omawianej zmiennej przed ponowną kategoryzacją.

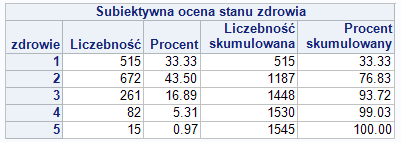
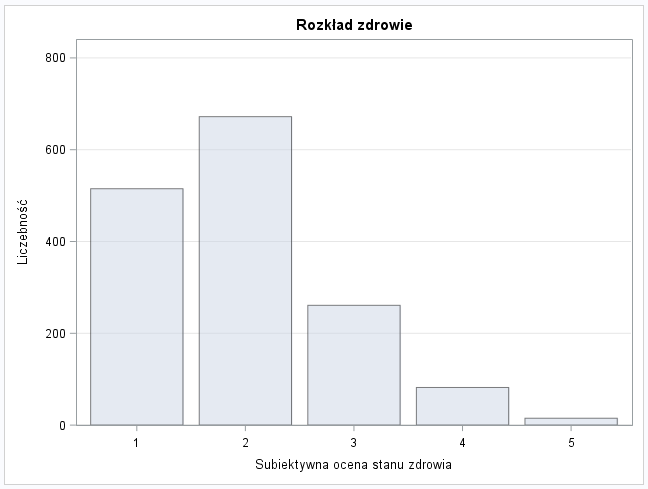


Tabela 34. Liczebność i częstość zmiennej zdrowie przed ponowną kategoryzacją.

Poniższy wykres wskazuje liczebności i częstości dla kategorii omawianej zmiennej przed ponowną kategoryzacją.



Wykres 26. Histogram dla zmiennej zdrowie przed ponowną kategoryzacją.

Wśród respondentów zdecydowaną większość stanowią osoby deklarujące bardzo dobry (515 odpowiedzi - 33,33%) lub dobry (672 odpowiedzi - 43,50%) stan zdrowia. Osoby deklarujące przeciętny stan zdrowia stanowią 16,89% badanych. Takiej odpowiedzi udzieliło 261 osób. Respondenci, którzy zadeklarowali zły lub bardzo zły stan zdrowia stanowią odpowiednio 5,31% i 0,97% badanej populacji. Takich odpowiedzi udzieliło odpowiednio 82 i 15 osób. Z wyników badania można wywnioskować, że Norwegowie są narodem, gdzie większość osób jest zadowolona ze swojego stanu zdrowia i jest on powyżej stanu przeciętnego.

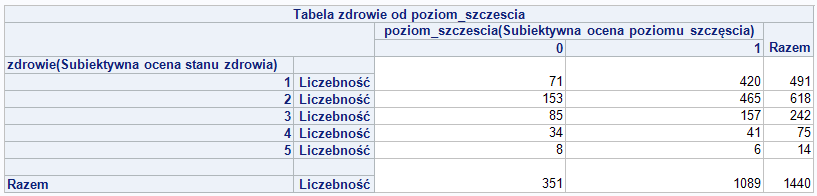


Tabela 35. Tablica kontyngencji zmiennych poziom\_szczescia i zdrowie przed ponowną kategoryzacją.

Zmienna została poddana kategoryzacji w celu zgrupowania kategorii o niewielkiej liczebności - 4 i 5 odpowiadających wartościom deklarowanego złego lub bardzo złego stanu zdrowia oraz zapewnienia odpowiedniej liczebności poszczególnych komórek, co jest warunkiem niezbędnym do wykonania modelu regresji logistycznej. Po ponownym skategoryzowaniu zbioru rozkład zmiennej *zdrowie* przedstawia się następująco:

Zmienna przyjmuje wartości od 1 do 4, które należy rozumieć w następujący sposób po przypisaniu bardziej intuicyjnych etykiet:

* 1 - bardzo dobry,
* 2 - dobry,
* 3 - umiarkowany,
* 4 - zły

Poniższa tabela prezentuje rozkład kategorii dla wartości omawianej zmiennej po ponownej kategoryzacji.

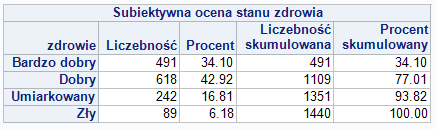
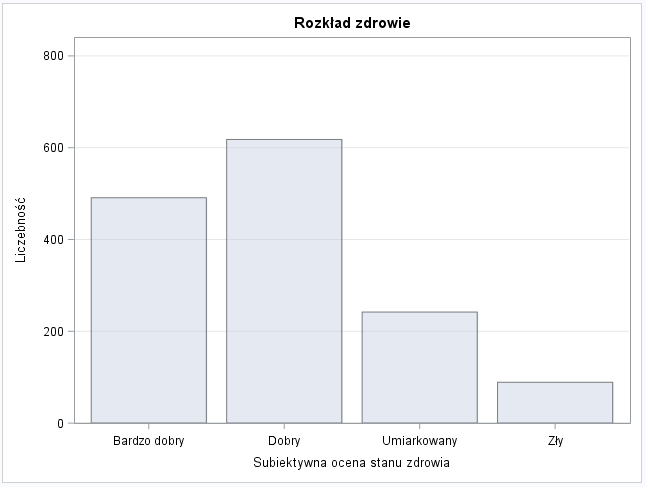


Tabela 36. Liczebność i częstość zmiennej zdrowie po ponownej kategoryzacji.

Poniższy wykres wskazuje liczebności i częstości dla omawianej zmiennej po ponownej kategoryzacji.



Wykres 27. Histogram dla zmiennej zdrowie po ponownej kategoryzacji.

Zdecydowaną większość obserwacji stanowią rekordy, gdzie stan zdrowia przyjmuje wartość 1 (bardzo dobry) i 2 (dobry). Takich obserwacji jest odpowiednio 491 (34,10%) i 618 (42,92%). Po kategoryzacji, liczba osób których stan zdrowia przyjmuje wartość 3 (umiarkowany) wynosi 242, czyli 16,81% badanej populacji. Liczba obserwacji, dla których stan zdrowia przyjmuje wartość 4 (zły) wynosi 89, czyli 6,18% badanej populacji. Poszczególne komórki mają liczebność przekraczającą 20, a więc prawidłową do wykonania modelu regresji logistycznej.

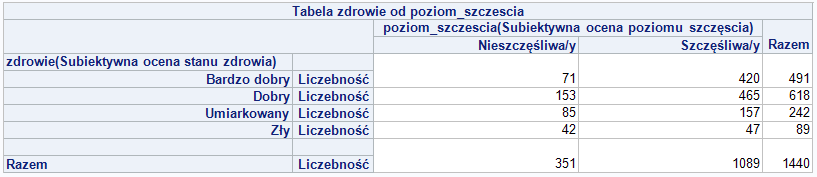


Tabela 37. Tablica kontyngencji dla zmiennych poziom\_szczescia i zdrowie po ponownej kategoryzacji.

**Zmienna: satysf\_dem**

Na podstawie przeglądu literatury można stwierdzić, że osoby, które uważają otaczający świat za sprawiedliwy i są przekonane o moralnej słuszności, lepiej oceniają własne życie. Powyższa teza autorstwa K. Skarżyńskiej i A. Cisłak została zawarta w książce Tatiany Kanasz ,,*Uwarunkowania szczęścia*”, gdzie autorka potwierdza ważność oceny stanu demokracji jako determinantę poziomu szczęścia. Powołuje się również na tezę głoszoną przez R. Bisward- Dieniera, który uważa, iż „warunki społeczne mają ogromne znaczenie dla poczucia szczęścia i że rządy państw – a nie tylko jednostki – wpływają na poziom doświadczanego przez ludzi dobrostanu. […] Życie w zamożnym, stabilnym i dobrze rządzonym kraju sprzyja poczuciu szczęścia”[[26]](#footnote-26).

Ponadto World Happiness Report 2019 potwierdza, że wysoka efektywność zarządzania demokracją wpływa pozytywnie na poziom szczęścia odczuwany przez społeczeństwo[[27]](#footnote-27).

Z tego powodu zespół projektowy uznał ważność tej zmiennej dla modelowanego zjawiska i postanowił ją włączyć do analizy. Poniżej zaprezentowano rozkład odpowiedzi na pytanie o demokrację wraz z liczebnościami.

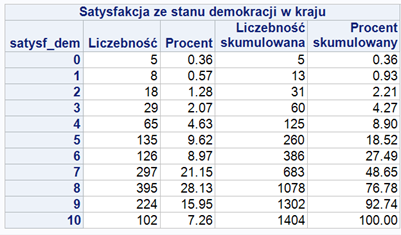
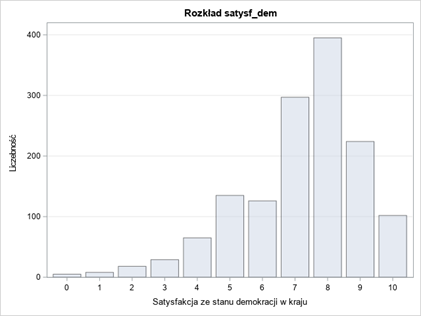


Tabela 38. Liczebność i częstość zmiennej satysf\_dem przed kategoryzacją.



Wykres 28. Histogram dla zmiennej satysf\_dem przed kategoryzacją.

Norwegowie są zdecydowanie zadowoleni ze stanu demokracji w swoim kraju. Z Tabeli nr 38 wynika, iż najczęstszym wskazaniem, była ocena 8 w skali 11 stopniowej - blisko 400 wskazań. Ponad 72% ankietowanych opowiedziało się za tym, że są zadowoleni ze stanu demokracji w kraju.

W związku z tym, że dla kategorii 0-4 odnotowano zaledwie niespełna 9% wskazań oraz niedostateczną liczbę obserwacji dla badanego zjawiska, postanowiono ponownie skategoryzować zmienną - potwierdza to Tabela nr 39 przedstawiająca kontyngencję zmiennej objaśnianej ze zmienną opisującą zadowolenie ze stanu demokracji. Taki odsetek odpowiedzi uznano za dostateczną frakcję dla 1 kategorii, dodanie do niej kolejnych obserwacji zniekształciłoby wyjściowy rozkład.



Tabela 39. Tabela kontyngencji dla zmiennych poziom\_szczescia i satysf\_dem przed kategoryzacją.

Zespół projektowy zdecydował się na ponowną kategoryzację zmiennej na 3 kategorie oznaczające małą (kategoria wyjściowa 0-4), średnią (5-7) oraz dużą (8-10) satysfakcję ze stanu demokracji w kraju. Przeprowadzenie tego procesu zapewniło dostateczną liczbę obserwacji dla wszystkich stworzonych kategorii, co pozwala na wykorzystanie zmiennej do modelu.

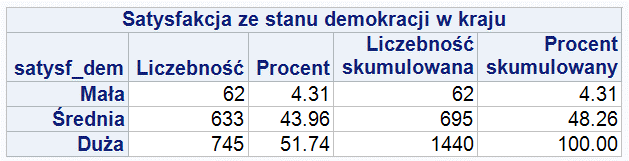
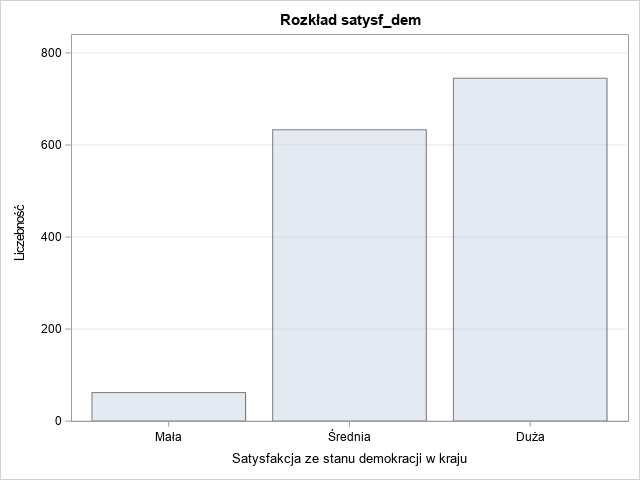


Tabela 40. Liczebność i częstość zmiennej satysf\_dem po kategoryzacji.



Wykres 29. Histogram dla zmiennej satysf\_dem po kategoryzacji.



Tabela 41. Tablica kontyngencji dla zmiennych poziom\_szczescia i satysf\_dem po kategoryzacji.

**Zmienna: przyw\_kraj**

Zespół projektowy wybrał zmienną przywiązanie do kraju jako zmienną, która może statystycznie istotnie wpływać na poziom subiektywnie odczuwanego szczęścia przez jednostkę. Wartości takie jak tożsamość narodowa, czy patriotyzm wydają się być nieodzownie związane ze szczęściem jednostki. Człowiek jako istota społeczna ma potrzebę istnieć nie tylko, jako odrębna, niezależna jednostka, ale również potrzebuje być częścią większej grupy, z którą mógłby się identyfikować. Zatem przywiązanie do kultury, historii czy tradycji danego kraju może istotnie wpływać na poczucie szczęścia jednostki. Powyższe przypuszczenia grupy projektowej znajdują uzasadnienie również w literaturze.

Teoria społecznej tożsamości (Social Identity Theory- SIT) zaproponowana przez Henriego Tajfela w 1981 r. zakłada, że bycie członkiem grupy jest kluczową składową tożsamości i poczucia własnej wartości. Teoria ta stała się punktem wyjścia i najważniejszą przesłanką dla grupy projektowej do włączenia do modelu zmiennej ‘przywiązanie do kraju’. Ponadto międzynarodowe badanie z 2011 r. raportuje, że satysfakcja narodowa, a dokładniej satysfakcja z przeszłości, teraźniejszości i przyszłej sytuacji własnej narodowości jest pozytywnie związana ze szczęściem jednostki[[28]](#footnote-28).

W analizowanym badaniu wartości zmiennej przywiązanie do kraju (przyw\_kraj) zostały zebrane w wyniku odpowiedzi respondentów na pytanie: Jak emocjonalnie jesteś przywiązany do kraju?

Po usunięciu braków danych oraz wartości wskazujących na odmowę odpowiedzi lub nieznajomość odpowiedzi zmienna przyjmowała następujące wartości. Zmienną mierzono na skali jedenastostopniowej, gdzie 0 oznacza, że respondent nie jest w ogóle emocjonalnie przywiązany do kraju, natomiast wartość 10 oznacza, że respondent czuje się bardzo emocjonalnie związany z krajem. Wykres nr 30 przedstawia rozkład zmiennej w badanej próbie Norwegów.

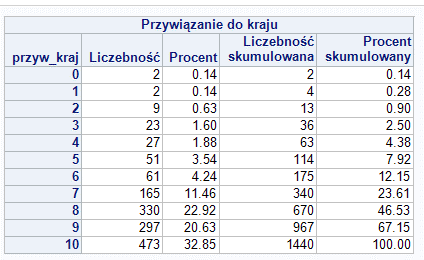
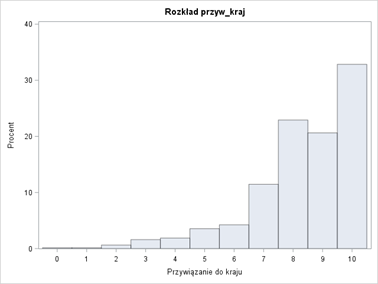


Tabela 42. Liczebność i częstość zmiennej przyw\_kraj przed kategoryzacją.



Wykres 30. Histogram dla zmiennej przyw\_kraj przed kategoryzacją.

Najliczniejszą grupą respondentów są osoby, które są bardzo emocjonalnie przywiązane do kraju (wskazanie na 10) i stanowią oni ok. 33% badanych. Najmniej liczne grupy to te ze wskazaniami odpowiednio 0,1,2,3 (brak lub małe przywiązanie do kraju) i stanowią one w sumie 2,5% respondentów. W następnym etapie dokonano analizy tablicy kontyngencji w celu zbadania łącznych rozkładów zmiennej przywiązanie do kraju oraz zmiennej poziom szczęścia. Tablica kontyngencji została przedstawiona w Tabeli nr 43.

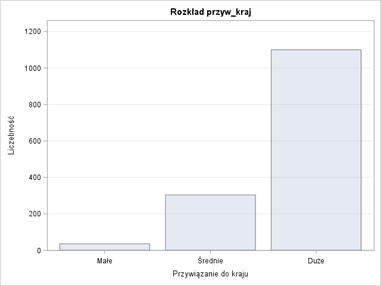


Tabela 43. Tablica kontyngencji dla zmiennej poziom\_szczescia i przyw\_kraj przed kategoryzacją.

Ze względu na małą liczbę obserwacji w niektórych komórkach zdecydowano się na dokonanie rekategoryzacji zmiennej. Operacja ta pozwoli na zmniejszenie liczby kategorii i uproszczenie analizy. Zdecydowano się utworzyć następujące nowe kategorie zmiennej:

* 1 - mały stopień emocjonalnego przywiązania do kraju
* 2 - średni stopień emocjonalnego przywiązania do kraju
* 3 - wysoki stopień emocjonalnego przywiązania do kraju

Nowa kategoria ‘1’ obejmuje wcześniejsze kategorie mniejsze lub równe 3, nowa kategoria ‘2’ obejmuje wcześniejsze kategorie należące do zbioru [4,7], natomiast nowa kategoria ‘3’ obejmuje wcześniejsze kategorie większe lub równe 8. Rozkład zmiennej przywiązanie do kraju po ponownej kategoryzacji przedstawia Wykres nr 31 oraz Tabela nr 44.



Wykres 31. Histogram dla zmiennej przyw\_kraj po rekategoryzacji.

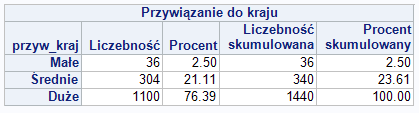


Tabela 44. Liczebność i częstość zmiennej przyw\_kraj.

Analizując rozkład odpowiedzi dotyczących emocjonalnego przywiązania Norwegów do kraju dostrzeżono, że tak, jak w pierwotnym rozkładzie najliczniejszą grupę stanowią osoby o wysokim przywiązaniu do kraju (76,39%), natomiast najmniej liczną grupę stanowią osoby o małym przywiązaniu do kraju (2,5%). Sporządzono również tablicę kontyngencji dla rekategoryzowanej zmiennej.

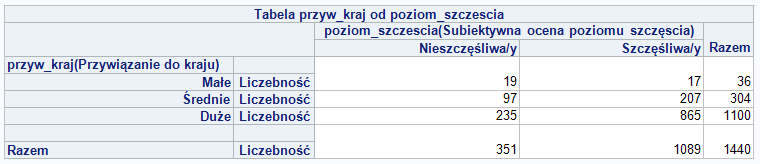


Tabela 45. Tablica kontyngencji dla zmiennej poziom\_szczescia i przyw\_kraj po kategoryzacji.

**Zmienna: doch\_gosp**

Zmienna *doch\_gosp* przedstawia dochód netto gospodarstwa domowego. W badaniach zaobserwowano, że przyrost dochodu inaczej oddziałuje na poziom szczęścia niż inne czynniki. Ludzie o niskim statusie społecznym dużo bardziej cieszą się z podwyższenia swojego statusu społecznego niż osoby zamożne. Dodatkowo ludzie z niezaspokojonymi podstawowymi potrzebami deklarują najniższy poziom szczęścia. Z jednej strony poczucie szczęścia spada, gdy człowiek podporządkowuje swoje szczęście dobrom materialnym i bogaceniu się[[29]](#footnote-29).

Z drugiej strony na podstawie artykułu J. Hartoga i H. Oosterbeeka pt.: „*Health, wealth, and happiness: why pursue a higher education?*”, w którym ukazano pozytywną korelację pomiędzy szczęściem a dochodem w wyniku analiz zbioru danych dotyczących Holandii, można stwierdzić, że dochód jest zmienną istotną i wpływającą na poziom deklarowanego szczęścia[[30]](#footnote-30).

Zmienna *doch\_gosp* początkowo była podzielona na 10 kategorii, które zostały przedstawione poniżej. W nawiasach umieszczono przedziały miesięcznych zarobków wyrażonych w norweskich koronach, które zostały podane w kwestionariuszu:

* 1 – J (Poniżej 20 900 NOK)
* 2 – R (20 901 - 29 700 NOK)
* 3 – C (29 701 - 37 300 NOK)
* 4 – M (37 301 - 44 700 NOK)
* 5 – F (44 701 - 52 800 NOK)
* 6 – S (52 801 - 60 700 NOK)
* 7 – K (60 701 - 69 200 NOK)
* 8 – P (69 201 - 79 600 NOK)
* 9 – D (79 601 - 96 800 NOK)
* 10 - H (96 801 NOK lub więcej)

Początkowe liczebności oraz częstości przedstawia Tabela nr 46. Mediana to kategoria 5, 197 respondentów, czyli 13,68% próby zadeklarowało przedział zarobków od 44 701 do 52 800 norweskich koron. Jednak każda z kategorii osiągnęła liczebności powyżej 100, a najniższe wartości przyjmują pierwsze oraz ostatnie kategorie. Możemy zatem powiedzieć, że zmienna *doch\_gosp* ma rozkład symetryczny.

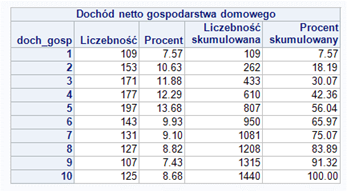
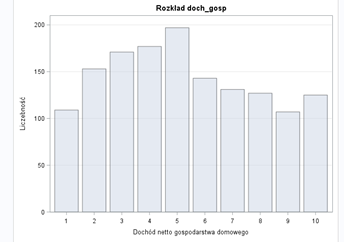


Tabela 46. Liczebność i częstość zmiennej doch\_gosp przed kategoryzacją.



Wykres 32. Histogram dla zmiennej doch\_gosp przed kategoryzacją.

Tabela nr 47 przedstawia tablicę kontyngencji dla tej zmiennej. Wyższe liczebności są zauważalne dla osób szczęśliwych. Jednak rozkład zmiennej *doch\_gosp* pomiędzy kategoriami zarówno dla osób szczęśliwych i nieszczęśliwych jest równomierny.

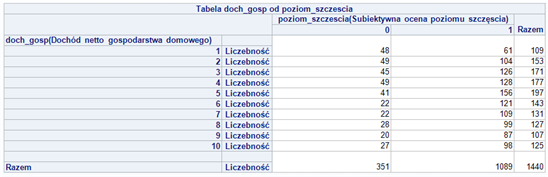
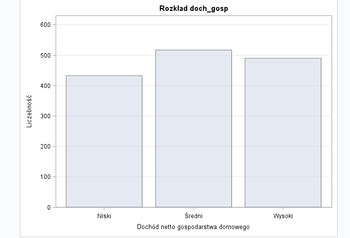


Tabela 47. Tablica kontyngencji dla zmiennych poziom\_szczescia oraz doch\_gosp przed kategoryzacją.

Zdecydowano podzielić zmienną na trzy nowe kategorie określające dochód jako: 1 – niski (uwzględniający kategorie 1, 2 i 3), 2 – średni (zawierający kategorie 4, 5 i 6), 3 – wysoki (dla kategorii 7, 8, 9 i 10). Po kategoryzacji, osoby o niskim dochodzie stanowią 30,07% próby, a osoby o średnim i wysokim, odpowiednio 35,90% i 34,03%.



Tabela 48. Liczebność i częstość dla zmiennej doch\_gosp po rekategoryzacji.



Wykres 33. Histogram dla zmiennej doch\_gosp po rekategoryzacji.

Tabela nr 49 przedstawia tablicę kontyngencji dla zmiennej *doch\_gosp* w zależności o zmiennej *poziom\_szczescia* dla nowych kategorii.

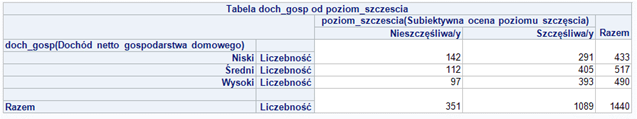


Tabela 49. Tablica kontyngencji dla zmiennych poziom\_szczescia i doch\_gosp po rekategoryzacji.

# Analiza współliniowości

Współliniowość pojawia się, gdy jedna lub więcej zmiennych niezależnych w modelu mogą być w przybliżeniu determinowane przez inne zmienne niezależne. Kiedy pojawia się współliniowość, wyestymowane współczynniki regresji dopasowanego modelu mogą być wysoce niemiarodajne. Jeśli dwie lub więcej zmiennych są ze sobą wysoko skorelowane, uzyskanie dobrego oszacowania ich wpływu na zmienną zależną jest utrudnione. Chociaż współliniowość nie obciąża współczynników, to sprawia, że są one bardziej niestabilne.

Oprócz tego błędy standardowe mogą być większe, a zmienne, które w rzeczywistości mają słaby wpływ na zmienną objaśnianą pojedynczo, mogą wykazywać dość silny wpływ jako grupa. Odpowiadające testy mają mniejszą moc.

W sytuacji skorelowania zmiennych niezależnych zmiany w jednej zmiennej są związane ze zmianami w innej zmiennej. Im silniejsza jest korelacja, tym trudniejsza jest zmiana wartości jednej zmiennej bez zmiany drugiej. Trudniejszą staje się estymacja niezależnej relacji pomiędzy każdą ze zmiennych niezależnych, a zmienną zależną, ponieważ wartości zmiennych niezależnych zmieniają się jednocześnie. Z tego względu, każda metoda modelowania powinna sprawdzać możliwą współliniowość na różnych etapach procesu doboru zmiennych.

Potrzeba redukcji współliniowości zmiennych zależy od stopnia współliniowości i celu dla jakiego model regresji został zbudowany. Im większy stopień współliniowości, tym bardziej utrudnione jest przygotowanie odpowiedniego modelu. Umiarkowana współliniowość zmiennych nie wymaga, aby problem ten został rozwiązany. Ponadto współliniowość oddziałuje tylko na konkretne zmienne niezależne, które są skorelowane. Z tego względu, jeśli współliniowość nie występuje dla zmiennych niezależnych, które są zmiennymi głównego zainteresowania, to rozwiązywanie problemu współliniowości nie jest konieczne. Zakładając, że model zawiera zmienne głównego zainteresowania i zmienne kontrolne, jeśli wysoka współliniowość istnieje tylko dla zmiennych kontrolnych i nie dotyczy zmiennych głównego zainteresowania, to można dokonywać interpretacji bez problemów.

Kolejnym istotnym punktem jest to, że współliniowość oddziałuje na współczynniki i p-value, ale nie wpływa na oszacowania predykcji, czy precyzję oszacowań predykcyjnych. Jeśli celem modelu jest prognozowanie, to nie ma potrzeby redukcji silnej współliniowości zmiennych.

W niniejszej pracy wykorzystano następujące metody identyfikacji współliniowości:

1. Tolerancja zmiennej (TOL)

2. Wskaźnik inflacji wariancji (VIF)

## Tolerancja zmiennej

Tolerancja zmiennej jest często stosowaną miarą współliniowości. Tolerancję definiuje się jako wielkość zmienności wybranej zmiennej niezależnej, która nie jest objaśniana przez inne zmienne niezależne. Obliczana jest za pomocą Wzoru 1:

*Wzór 1. Tolerancja zmiennej.*

gdzie R^2\*i jest współczynnikiem determinacji wielokrotnej, który wskazuje w jakim stopniu wybrana zmienna niezależna jest wyjaśniana przez wszystkie inne zmienne niezależne w modelu regresji. Wraz ze spadkiem wartości tolerancji, zmienna jest coraz bardziej przewidywana przez inne zmienne niezależne. Wskazuje to na współliniowość zmiennych. Popularnie przyjmowanym progiem jest tolerancja na poziomie 0.1. Niższe wartości tolerancji wskazują na wysoki poziom korelacji danej zmiennej niezależnej z innymi zmiennymi niezależnymi. Jednakże na przykład w przypadku małej wielkości próby można przyjąć bardziej restrykcyjne założenia. Tolerancja na poziomie 0 oznacza, że rozważana zmienna niezależna jest doskonałą kombinacją liniową pozostałych zmiennych niezależnych ujętych w modelu. Tolerancja na poziomie 1 oznacza natomiast, że rozważana zmienna niezależna jest całkowicie niezależna od innych zmiennych niezależnych uwzględnionych w modelu. W niniejszej pracy przyjmuje się akceptowalny poziom tolerancji leżący powyżej wartości 0,1. W celu identyfikacji współliniowości zmiennych obliczono poziomy tolerancji zmiennych dla modelu regresji logistycznej, który za zmienną objaśnianą przyjmuje zmienną *poziom\_szczescia*, a zmiennymi objaśniającymi jest 13 zmiennych opisanych w poprzedniej części pracy. Analiza Tabeli nr 50 pozwala stwierdzić, że dla każdej zmiennej przyjętej w modelu poziom tolerancji jest akceptowalny.

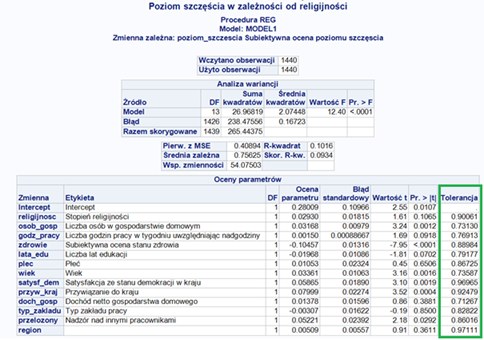


Tabela 50. Wartości tolerancji.

## Wskaźnik inflacji wariancji

W celu analizy współliniowości wykorzystano w pracy wskaźnik inflacji wariancji (ang. variance inflation factor - VIF). Jest to wskaźnik efektu, jaki inne zmienne niezależne mają na błąd standardowy oszacowań współczynników regresji. Czynnik inflacji wariancji jest bezpośrednio związany z wartością tolerancji, obliczany jest jako odwrotność wartości tolerancji. VIF bierze swoją nazwę z faktu, że pierwiastek kwadratowy z wartości VIF odzwierciedla stopień, w jakim błąd standardowy został zwiększony z powodu współliniowości. Wysokie wartości VIF wskazują również wysoki stopień współliniowości między zmiennymi. Do obliczenia wartości VIF wykorzystuje się Wzór 2:

*Wzór 2. Wskaźnik inflacji wariancji.*

W celu oszacowania wskaźnika inflacji wariancji zbudowano model regresji logistycznej, który za zmienną objaśnianą przyjmuje zmienną *poziom\_szczescia*, a zmiennymi objaśniającymi jest 13 zmiennych opisanych w poprzedniej części pracy. Za akceptowalny poziom inflacji wariancji (nadęcia wariancji) przyjęto poziom 10 i mniej.

Analiza Tabeli nr 51 pozwala wykryć brak niedokładnej współliniowości zmiennych, ponieważ żadne oszacowanie wartości współczynnika inflacji wariancji nie przekracza nawet wartości 1,5. Największe nadęcie wariancji (na poziomie 1,4) występuje dla zmiennej *doch\_gosp* (dochód netto gospodarstwa domowego). Oznacza to, że błąd standardowy zmiennej *doch\_gosp* jest (pierw. 1,4) 1,18 razy większy ze względu na występowanie słabej współliniowości. W takim przypadku w dalszej analizie uwzględniono wszystkie zmienne.

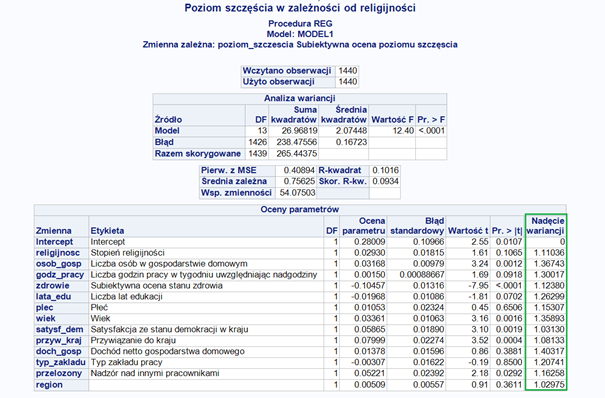


Tabela 51. Wartości inflacji wariancji.

# Weryfikacja modelu regresji logistycznej

W oparciu o wszystkie wcześniej wyszczególnione zmienne oraz ich efekty krzyżowe 2 stopnia zdecydowano się skonstruować model, którego podstawowe statystyki estymacyjne ilustruje Tabela nr 52.

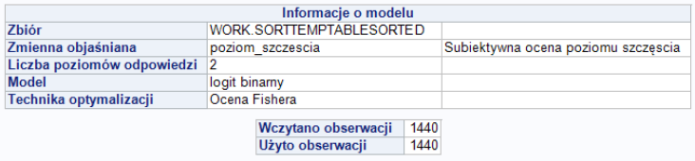


Tabela 52. Informacje o modelu regresji logistycznej.

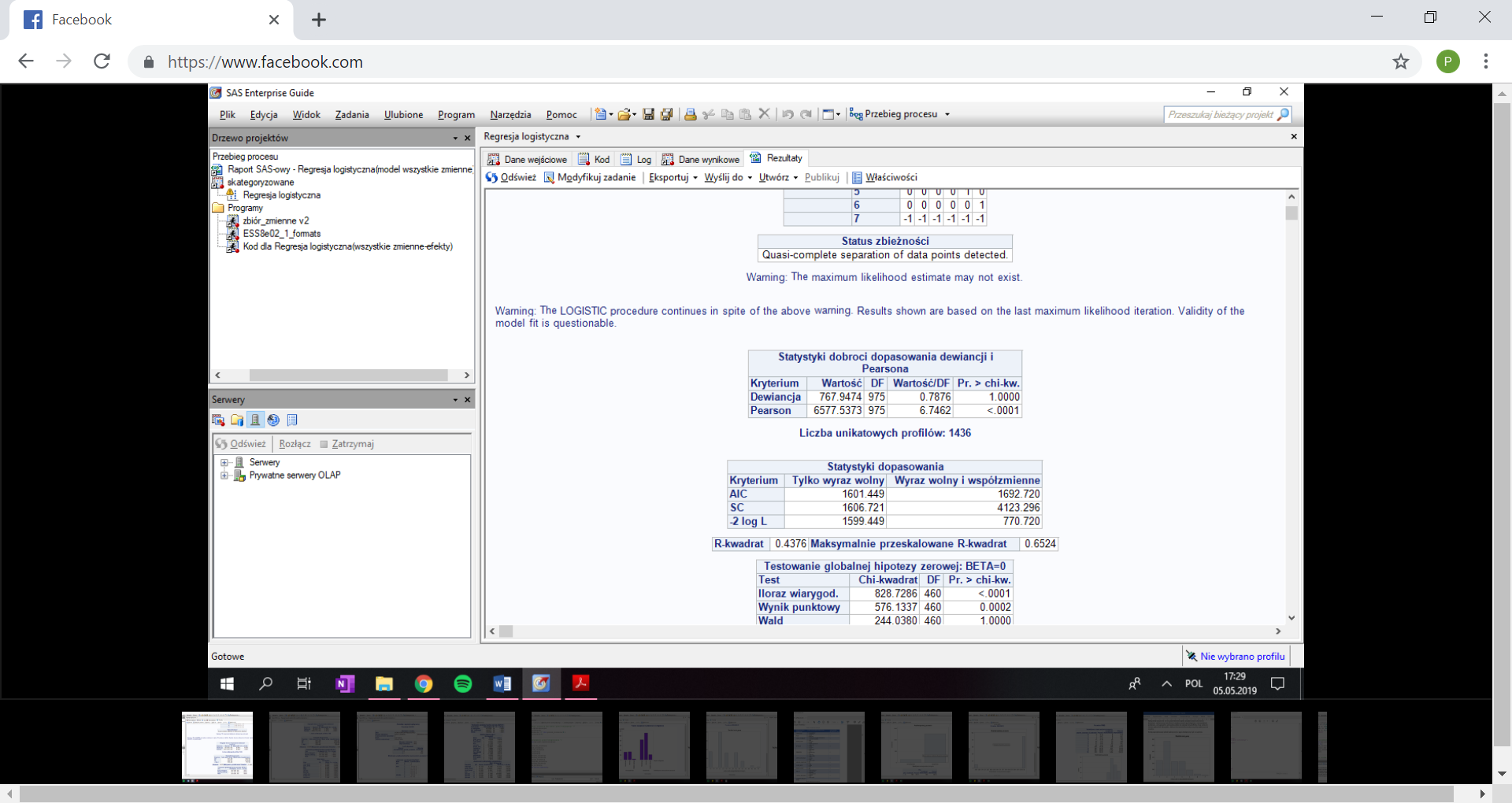


Tabela 53. Status zbieżności modelu regresji logistycznej.

Zmienną objaśnianą jest *poziom\_szczęścia*, która przyjmuje dwie wartości: 0 - osoba nieszczęśliwa, 1 - osoba szczęśliwa. Model został utworzony w oparciu o logit binarny używając wszystkich dostępnych 1440 obserwacji. Zjawiskiem modelowanym było przyjęcie przez zmienną wartość 1, a więc prawdopodobieństwo bycia szczęśliwym. W modelu nie osiągnięto kryteriów zbieżności - została zidentyfikowana quasi-kompletna separacja punktów danych.

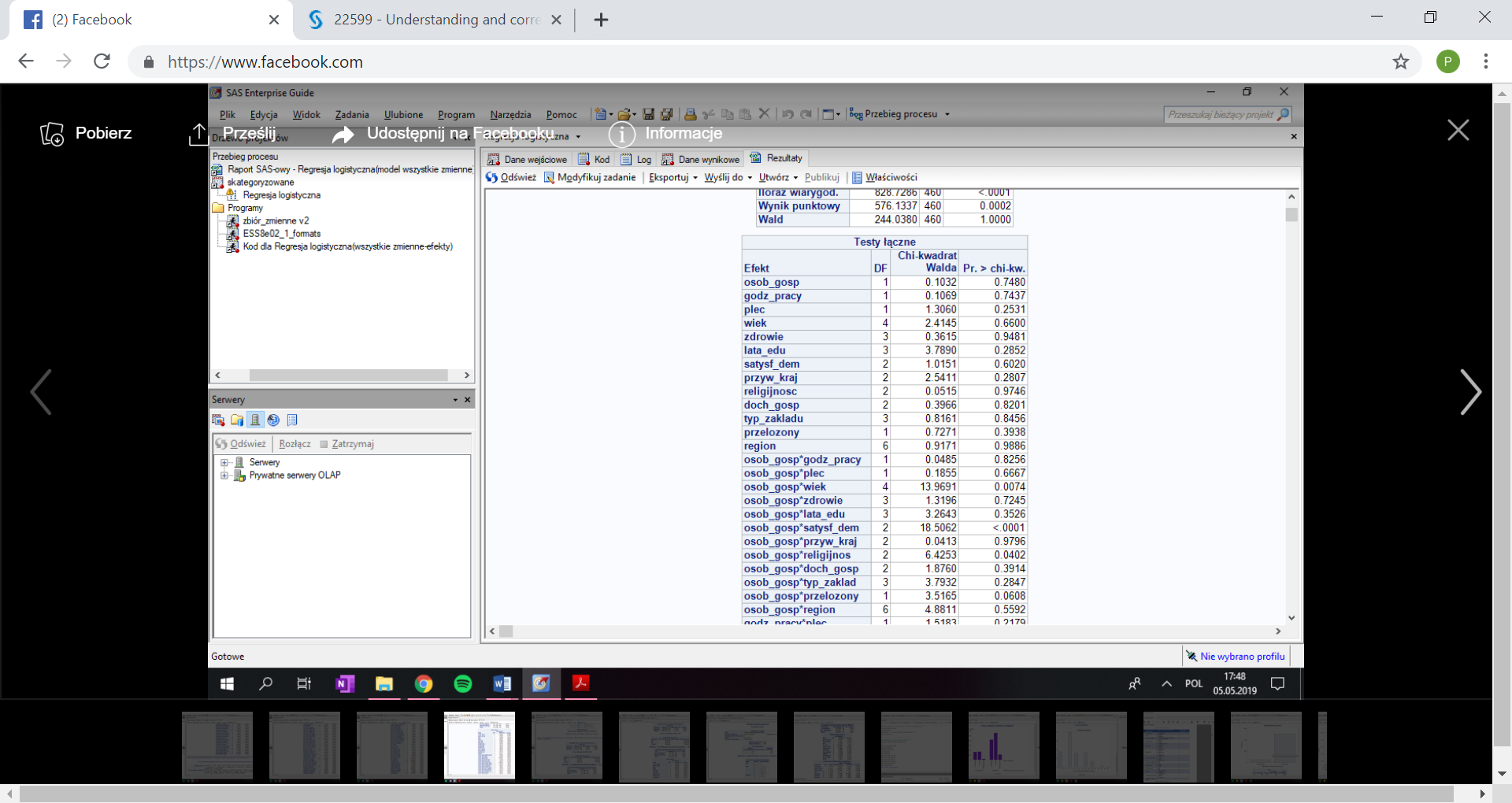


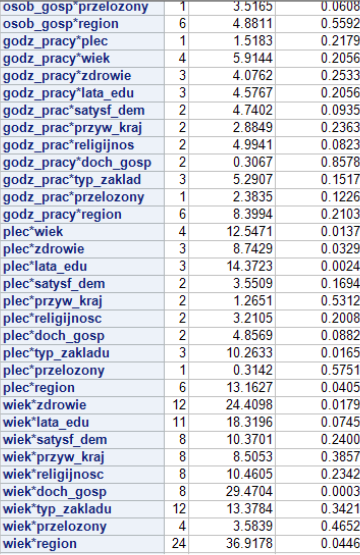
Tabela 54. Statystyki dobroci dopasowania modelu regresji logistycznej.

W Tabeli Nr 54 zostały zaprezentowane podstawowe statystyki estymacyjne modelu. Rozpatrując kryterium dobroci dewiancji i Pearsona dochodzimy do sprzecznych wniosków. P-value dla kryterium dewiancji wskazuje na brak podstaw do odrzucenia hipotezy zerowej mówiącej o dobrym dopasowaniu modelu do danych, natomiast dla kryterium reszt Pearsona mamy sytuację odwrotną.

Analizując kryteria informacyjne Akaikego, Schwarza oraz statystykę ilorazu wiarygodności powinniśmy się kierować minimalizacją ich wartości. Rozpatrując pierwsze dwa kryteria dochodzimy do wniosku, że lepszym modelem byłbym model tylko z wyrazem wolnym, natomiast statystyka ilorazu wiarygodności wskazuje na lepszą jakość modelu ze wszystkimi zmiennymi i ich efektami krzyżowymi drugiego rzędu.

Maksymalnie przeskalowany współczynnik R^2 na poziomie 0,65 wskazuje na dobre dopasowanie modelu do danych. W oparciu o test ilorazu wiarygodności i Walda dochodzimy do sprzecznych wniosków odnośnie łącznej istotności oszacowań współczynników wszystkich zmiennych objaśniających - rezultat pierwszego z nich wskazuje na odrzucenie hipotezy zerowej o braku istotności wszystkich zmiennych, natomiast wynik drugiego z nich wskazuje na brak podstaw do odrzucenia hipotezy zerowej.







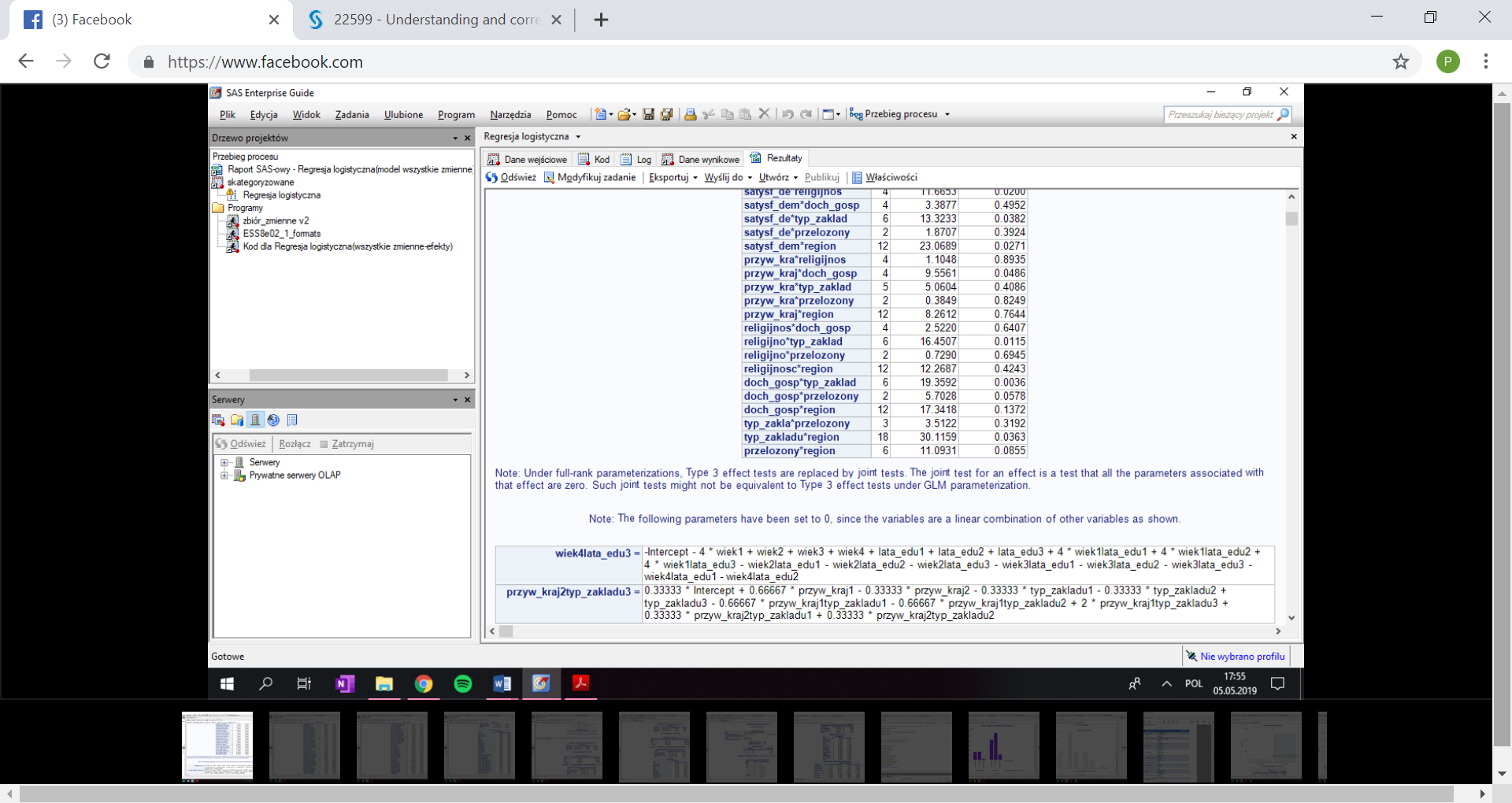


Tabela 55. Statystyki testu Walda w modelu regresji logistycznej.

Większość z efektów w modelu jest nieistotna, tylko dla nielicznych z nich p-value wynosi poniżej przyjętego. Oznacza to, że liczba zmiennych w modelu powinna być znacząco ograniczona. Ponadto z uwagi na to, że została stwierdzona quasi-kompletna separacja punktów danych, należy zastosować jedno z dwóch rozwiązań: albo skategoryzować zmienne ciągłe, albo ograniczyć liczbę efektów w modelu[[31]](#footnote-31). Zespół projektowy zdecydował się na drugie rozwiązanie, wykorzystując metodę selekcji forward i stepwise a następnie po porównaniu statystyk dopasowania wybranie lepszego z nich.

## Model uwzględniający tylko efekty główne

Postanowiono również wygenerować model posiadający wszystkie zmienne użyte jako efekty główne, bez efektów krzyżowych.

Poniżej znajduje się Tabela nr 56, w której zawarte są podstawowe informacje o estymacji modelu.

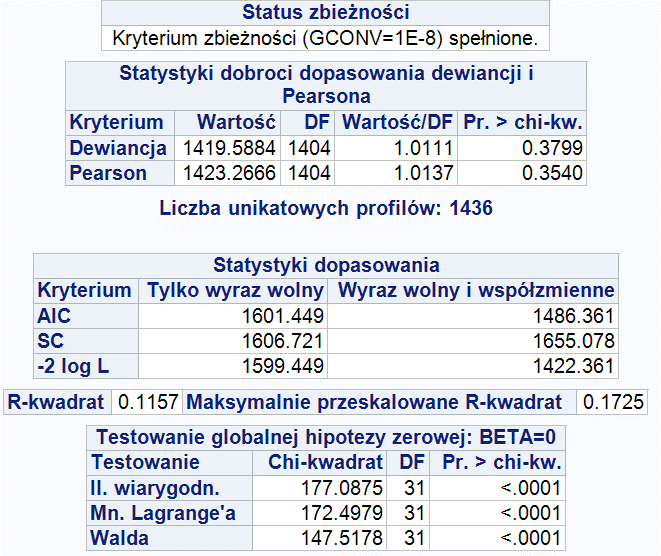


Tabela 56. Statystyki dopasowania modelu z efektami głównymi.

Model osiągnął kryterium zbieżności, a jego wartości statystyk dopasowania poza kryterium Schwarza wskazują na lepsze dopasowanie do danych modelu zawierającego wszystkie włączone zmienne w stosunku do modelu tylko z wyrazem wolnym. Maksymalny przeskalowany R^2 wyniósł 0,1725.

W tym wypadku również postanowiono przetestować hipotezę globalną o łącznej nieistotności współczynnikach przy zmiennych objaśniających. Na podstawie poziomu p-value możemy odrzucić hipotezę zerową mówiącą o łącznej nieistotności zmiennych objaśniających.

W tabeli poniżej przedstawione zostały oceny poszczególnych parametrów.

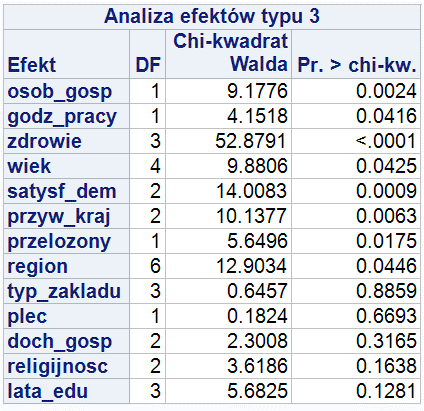


Tabela 57. Ocena istotności efektów głównych.

Wartości p-value wskazują na nieistotność zmiennych: typ\_zakladu, plec, doch\_gosp, religijność, lata\_edu. W związku z tym zasadne byłoby wykluczenie powyższych zmiennych z modelu, jednakże, aby sprawdzić czy powyższe działanie nie zaburza znacząco oceny parametrów zmiennych objaśniających, zespół projektowy postanowił wykluczać po kolei każdą ze zmiennych i badać względną zmianę oceny parametrów modelu. Wyniki opisanej analizy znajdują się w tabeli poniżej.

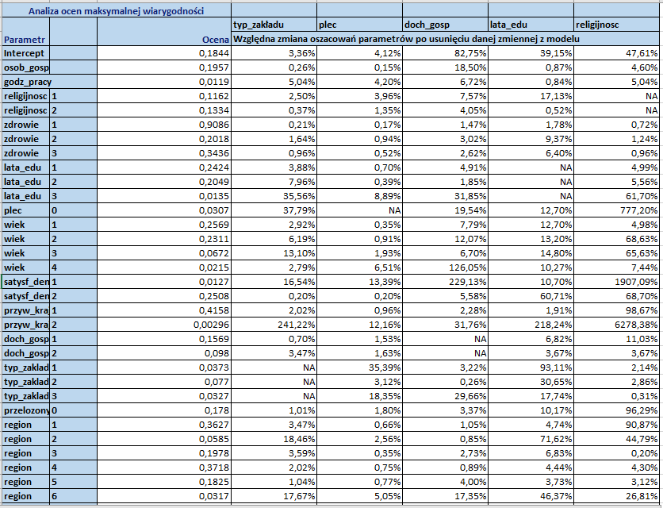


Tabela 58. Względna zmiana ocen parametrów po usunięciu danej zmiennej z modelu.

Analizując względne zmiany można dojść do wniosku, iż w przypadku wykluczenia każdej z potencjalnych zmiennych następują wysokie wahania oszacowań parametrów (powyżej 10%) szczególnie dla zmiennych kontrolnych.

Zespół projektowy zdecydował zatem o pozostawieniu wszystkich zmiennych objaśniających w modelu.

## Modele uzyskane metodą selekcji forward i stepwise

W Tabeli nr 59 znajdują się podstawowe statystyki dopasowania obu uzyskanych modeli.

|  |  |  |
| --- | --- | --- |
| Statystyka/ Kryterium | Metoda forward | Metoda stepwise |
| Max R^2 | 0,1773 | 0,1572 |
| AIC | 1481,20 | 1478,97 |
| S.C. | 1649,92 | 1584,41 |
| - 2 log L | 1417,20 | 1438,97 |

Tabela 59. Statystyki dopasowania modelu regresji logistycznej uzyskanego metodą forward i stepwise.

Z powodu wyższej wartości współczynnika R^2 w przypadku modelu uzyskanego metodą forward (który jest również wyższy od modelu tylko z efektami głównymi), zdecydowano się na wybór właśnie tej metody do estymacji modelu. Ponadto w tej metodzie do modelu została włączona większa liczba zmiennych.

Selekcja forward polega na rozpoczęciu budowy modelu od włączenia tylko wyrazu modelu. Następnie dodawane są zmienne w kolejności od tej, której p-value jest najniższe aż do momentu, w którym żadna ze zmiennych nie ma wartości p-value niższej od przyjętego poziomu istotności 0,05.

W Tabeli nr 60 zaprezentowano podstawowe statystyki dopasowania.

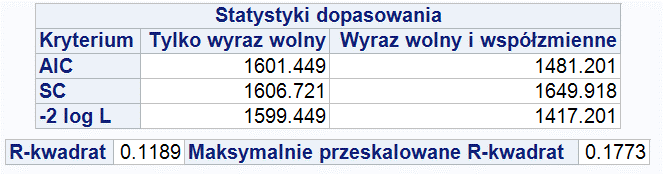


Tabela 60. Statystyki dopasowania modelu regresji logistycznej uzyskanego metodą forward.

Porównując wartości statystyk dla modelu tylko z wyrazem wolnym i modelu ze zmiennymi objaśniającymi dochodzimy do wniosku, że drugi model cechuje się lepszym dopasowaniem modelu do danych.

Postanowiono podobnie jak w przypadku modelu ze wszystkimi efektami zweryfikować globalną hipotezę zerową o łącznej nieistotności wszystkich zmiennych objaśniających.

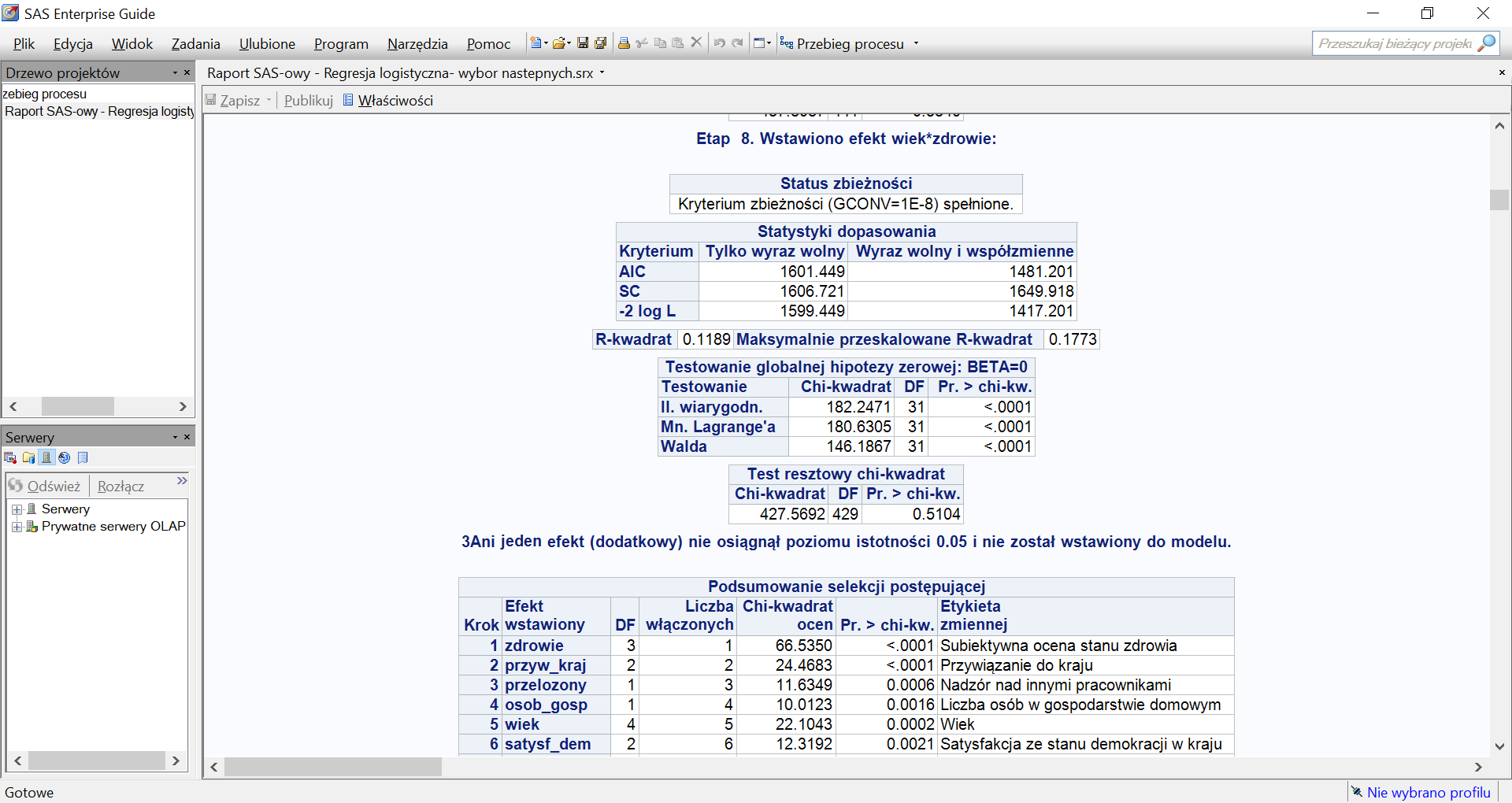


Tabela 61. Statystyki istotności modelu regresji logistycznej.

Wartości p-value dla wszystkich 3 testów wynoszą poniżej 0,05 co oznacza, że możemy odrzucić hipotezę zerową mówiącą o łącznej nieistotności wszystkich zmiennych - co najmniej jedna ze zmiennych jest istotna. Ponadto weryfikacja hipotezy zerowej testu resztowego chi-kwadrat nie pozwala na stwierdzenie, iż tak zredukowana postać jest odpowiednia.

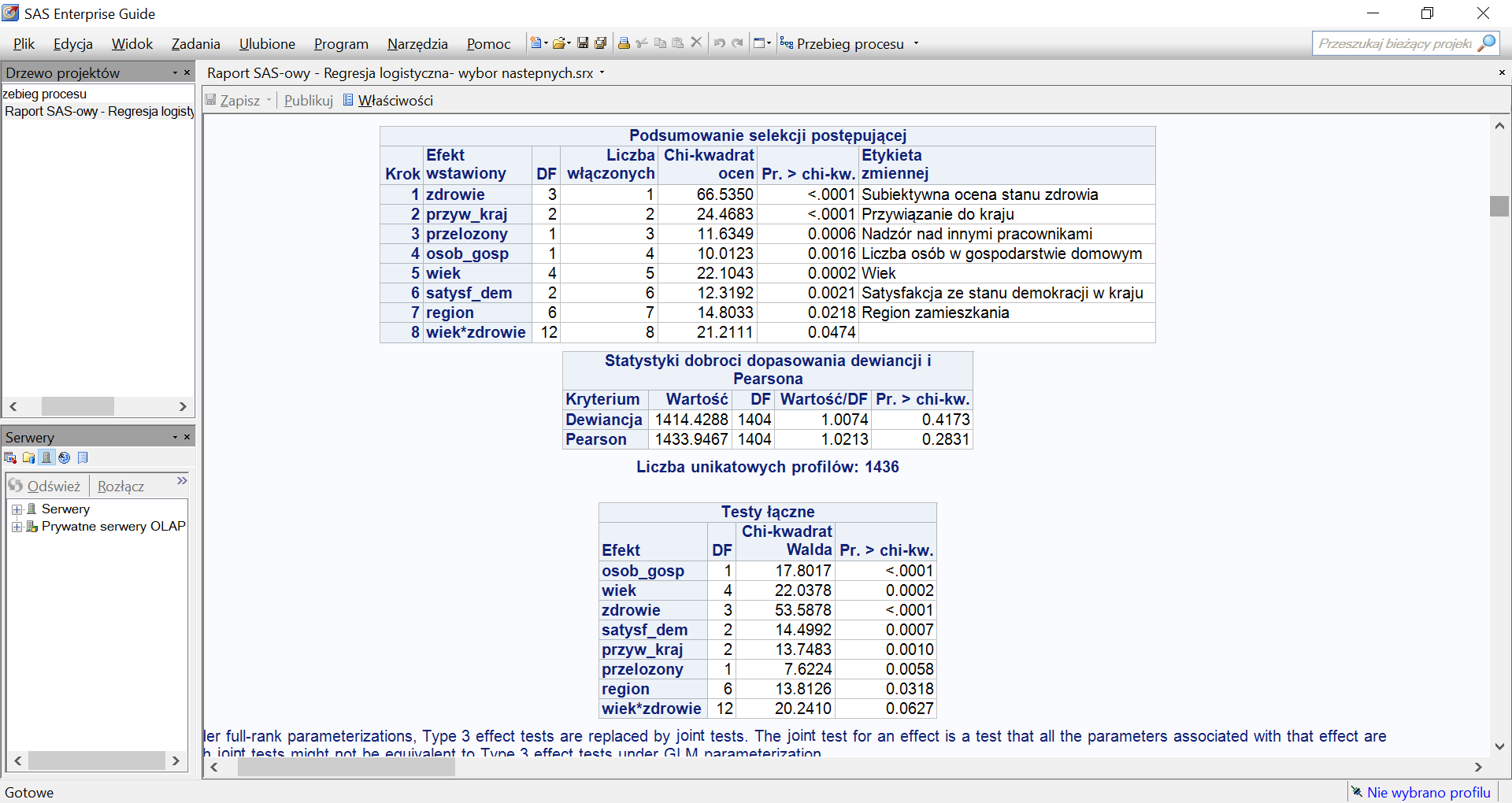
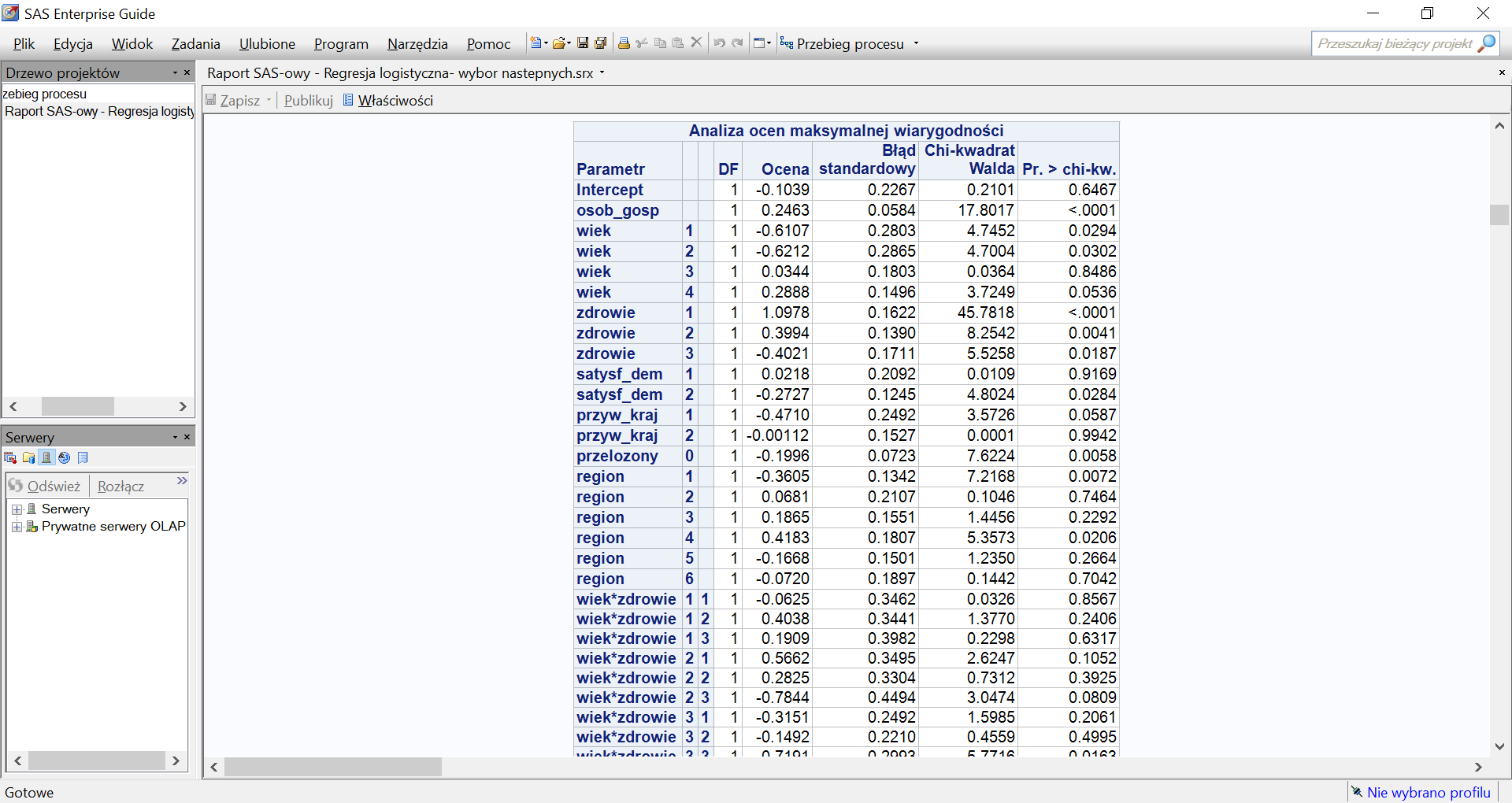


Tabela 62. Podsumowanie postępującej selekcji zmiennych.

W Tabeli nr 62 znajduje się podsumowanie metody forward. Do modelu zostało dodane 8 efektów, z czego 7 głównych i jeden krzyżowy drugiego rzędu - wiek\*zdrowie. Dla wszystkich efektów p-value jest niższe niż 0,05 co potwierdza zasadność włączenia ich do modelu.

## Interpretacja parametrów modelu

Poniższa Tabela nr 63 przedstawia wartości wyestymowanych parametrów dla wszystkich efektów znajdujących się w modelu.



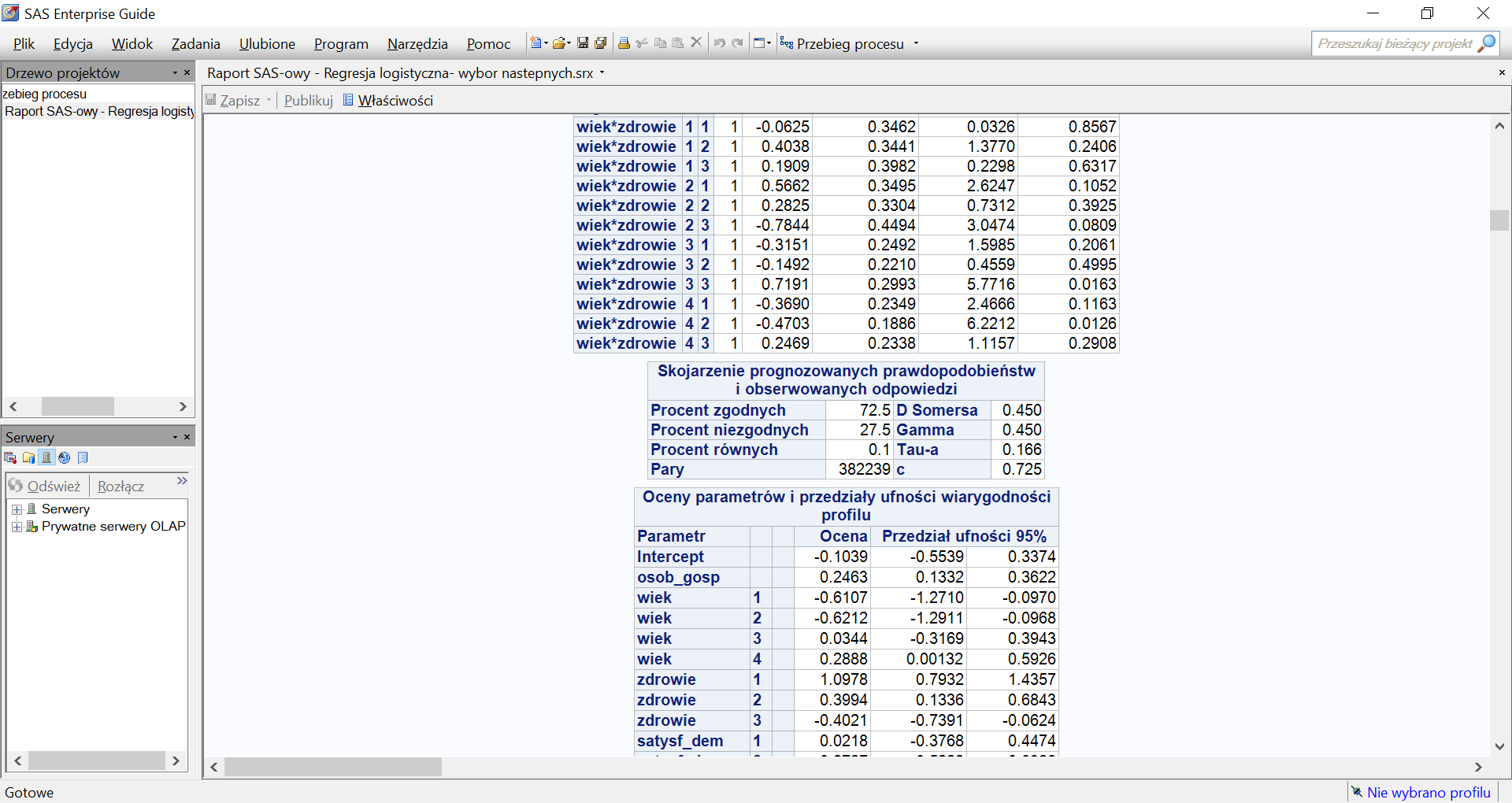


Tabela 63. Analiza ocen maksymalnej wiarygodności.

Przy przyjętym poziomie istotności 0,05 można stwierdzić brak istotności, który występuje przy wyrazie wolnym, dwóch poziomach zmiennej wiek (3 i 4), jednym poziomie zmiennych satysfakcji z demokracji oraz przywiązania do kraju, czterech poziomach zmiennej region oraz 10 poziomach zmiennej przedstawiającej efekt krzyżowy zmiennej wiek i zdrowie. Oznacza to np. że w przypadku, gdy respondent pochodzi z regionu NO02, atrybut ten nie wpływa w sposób statystycznie istotny na poziom szczęścia, który jest zmienną objaśnianą. Pomimo nieistotności niektórych poziomów zmiennych, nie zdecydowano się na ich usunięcie, gdyż przy istotności zmiennych w całości, brak istotności niektórych ich kategorii jest dopuszczalny. Pozostawiono je w celach kontrolnych.

Można by się pokusić o interpretację stojących przy parametrach znaków, jednakże nie jest to możliwe bezpośrednio w tej postaci. Aby tego dokonać należy zinterpretować ilorazy szans, które zostały przedstawione w kolejnej tabeli i na wykresie.

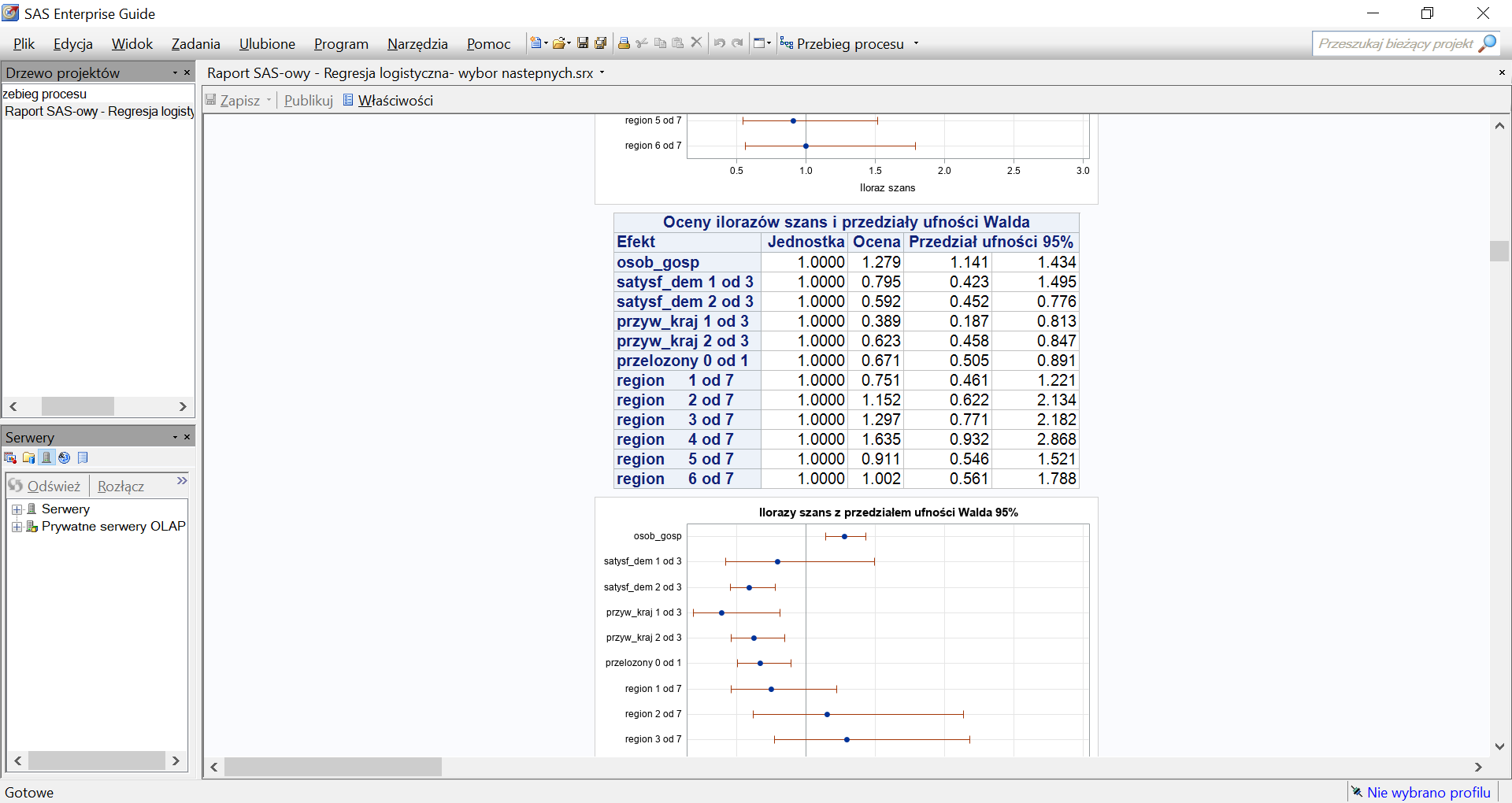


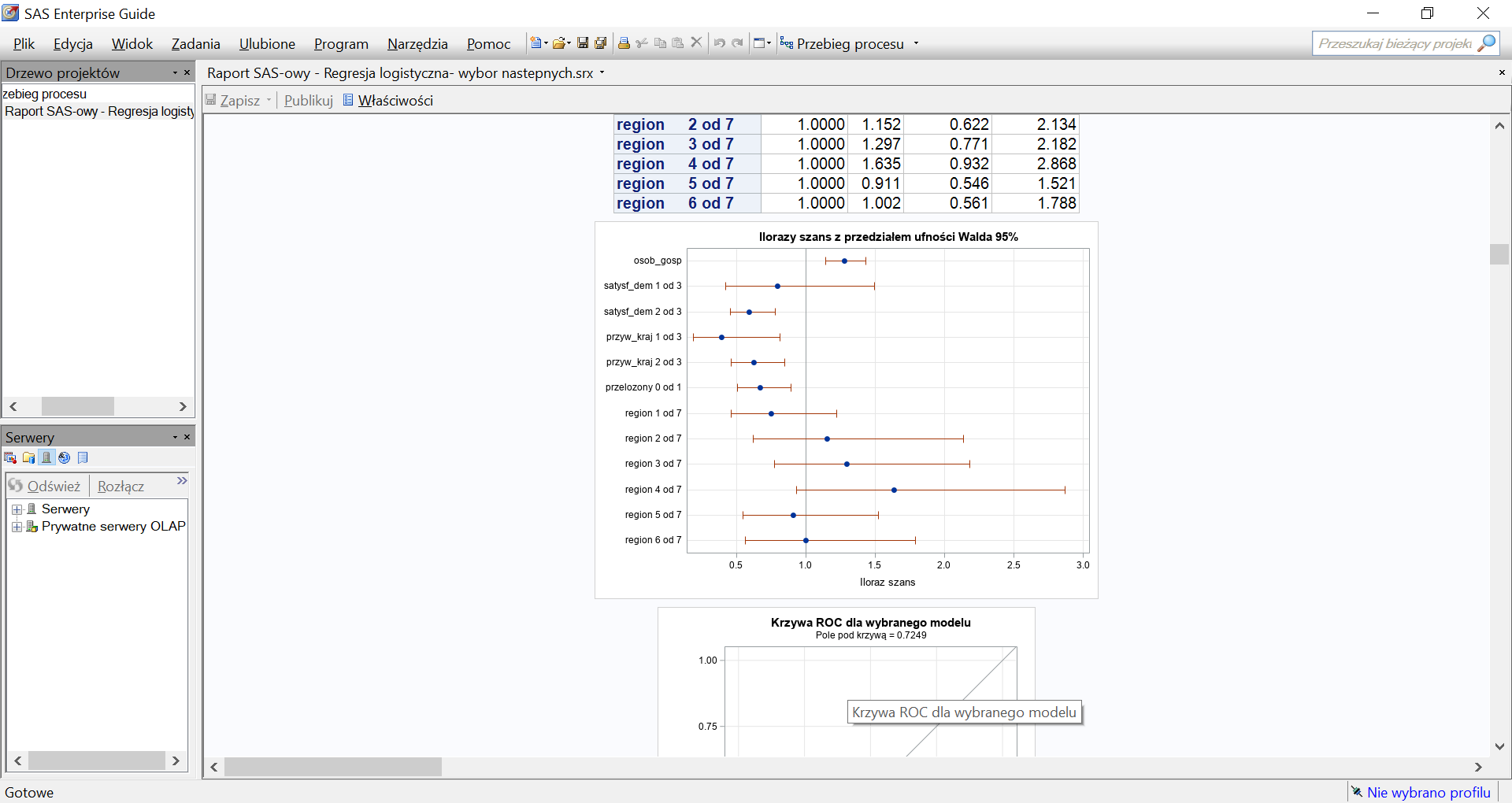
Tabela 64. Oceny ilorazów szans i przedziały ufności Walda.

Dla każdej zmiennej ocena parametru Walda określa jej istotność statystyczną - jeżeli do 95 procentowego przedziału ufności ilorazu szans należy wartość 1, to oznacza, że zmienna jest nieistotna statystyczne.

Interpretacja ilorazu szans wygląda następująco: dla zmiennych binarnych oznacza on o ile procent wzrasta lub maleje szansa na pozytywną ocenę swojego poziomu szczęścia w momencie przejścia danej zmiennej od wartości 1 do wartości 0. Dla zmiennych ciągłych ocena ilorazu szans określa o ile szansa ta wzrasta (maleje) wraz z każdą dodatkową jednostką zmiennej.

Interpretacje ilorazów szans dla poszczególnych zmiennych przedstawiają się następująco:

* Osoby posiadające o 1 osobę w gospodarstwie domowym więcej, mają o 27,9% wyższą szansę na odczuwanie wysokiego poziomu szczęścia (bycia szczęśliwym) przy pozostałych charakterystykach niezmienionych (ceteris paribus).
* Iloraz szans dla respondentów odczuwających małą satysfakcję z demokracji jest statystycznie nieistotny, nie należy go zatem interpretować.
* W porównaniu z respondentami wykazującymi duże zadowolenie ze stanu demokracji w kraju, osoby odczuwające średni poziom satysfakcji ze stanu demokracji mają o 40,8% niższą szansę na odczuwanie szczęścia (ceteris paribus).
* W porównaniu z osobami wykazującymi duże przywiązanie do kraju, osoby, które odczuwają niski poziom przywiązania mają o 61,1% niższą szansę na odczuwanie szczęścia (ceteris paribus).
* W porównaniu z osobami wykazującymi wysoki poziom przywiązania do kraju, osoby, które odczuwają średni poziom przywiązania mają o 31,7% niższą szansę na odczuwanie szczęścia (ceteris paribus).
* Osoby nieposiadające nadzoru nad pracownikami mają o 32,9% niższą szansę na odczuwanie szczęścia w porównaniu z osobami, które ten nadzór posiadają (ceteris paribus).
* Iloraz szans dla osób mieszkających w regionie NO01 jest statystycznie nieistotny, nie należy go zatem interpretować
* Iloraz szans dla osób mieszkających w regionie NO02 jest statystycznie nieistotny, nie należy go zatem interpretować
* Iloraz szans dla osób mieszkających w regionie NO03 jest statystycznie nieistotny, nie należy go zatem interpretować
* Iloraz szans dla osób mieszkających w regionie NO04 jest statystycznie nieistotny, nie należy go zatem interpretować
* Iloraz szans dla osób mieszkających w regionie NO05 jest statystycznie nieistotny, nie należy go zatem interpretować
* Iloraz szans dla osób mieszkających w regionie NO06 jest statystycznie nieistotny, nie należy go zatem interpretować



Wykres 34. Ilorazy szans z przedziałem ufności Walda.

Wykres powyżej przedstawia oceny ilorazu szans wraz z 95 procentowym przedziałem ufności Walda. Z niego w łatwy sposób można odczytać te zmienne dla których oszacowanie ilorazu szans jest statystycznie nieistotne - do przedziału Walda należy 1. Na wykresie widać również, że największe wartości ilorazu szans pod względem odchylenia od 1 uzyskano dla zmiennej sztucznej *przyw\_kraj* 1, zatem to przywiązanie do kraju wpływa w największy sposób na odczuwany poziom szczęścia.

# Ocena jakości modelu

Na ocenę jakości modelu składa się dobroć dopasowania, dyskryminacji i jakość klasyfikacji.

W tabeli poniżej przedstawione są statystyki dobroci dopasowania dewiancji i Pearsona.



Tabela 65. Statystyki dopasowania dewiancji i Pearsona.

Powyższe statystyki służą ocenie dobroci dopasowania modelu do danych. Wartości p-value obu z nich wskazują na brak podstaw do odrzucenia hipotezy zerowej mówiącej o dobroci dopasowania modelu do danych, jednakże ze względu na bardzo duża liczbę profili nie możemy opierać się na interpretacji tychże statystyk.

## Test Hosmera-Lemeshowa

Kolejny etap oceny jakości dopasowania modelu do danych empirycznych to test Hosmera-Lemeshowa. Test ten uznaje się za główną składową diagnostyki modelu regresji logistycznej. Test ten można stosować również, gdy spotykamy problem małej liczby obserwacji w unikatowych profilach. Weryfikuje się hipotezę zerową mówiącą o równości wartości obserwowanych i przewidywanych w arbitralnie wybranych podgrupach. Wystarczająco bliskie wartości obserwowane i przewidywane przez model pozwalają przyjąć założenie o dobroci dopasowania modelu do danych. Statystyka testu Hosmera-Lemeshowa przy założeniu prawdziwości hipotezy zerowej ma asymptotyczny rozkład chi-kwadrat z liczbą stopni swobody równą g-2, gdzie g to liczba grup, na jaką wcześniej podzieliliśmy obserwacje na potrzeby wykonania testu. Wyniki testu Hosmera-Lemeshowa przedstawia Tabela nr 66.



Tabela 66. Statystyka Chi-kwadrat dla testu zgodności Hosmera i Lemeshowa.

Przyjmując poziom istotności równy 0,05 dla zbudowanego modelu regresji logistycznej brak podstaw do odrzucenia hipotezy zerowej, która mówi o tym, że wartości obserwowane i przewidywane przez model są równe. Zatem można wnioskować, że model jest dobrze dopasowany do danych empirycznych.

## Dyskryminacja modelu

Analiza dyskryminacyjna modelu obejmuje ocenę stopnia, w jakim model prawidłowo rozróżnia obserwacje w dwóch grupach- zajścia i braku zajścia analizowanego zjawiska. Weryfikacja odbywa się poprzez wyznaczenie par zgodnych i niezgodnych w oparciu o wyestymowane wartości prawdopodobieństwa. Zdolności dyskryminacyjne modelu przedstawia Tabela nr 67. Procent zgodny oznacza, że model poprawnie sklasyfikował 72,5% obserwacji, natomiast procent niezgodny wskazuje, że model sklasyfikował niepoprawnie 27,5% obserwacji. Procent równych oznacza, że pary powiązane (złączone) stanowią 0,1% wszystkich par. Statystyka D-Sommersa przedstawia nadwyżkę par zgodnych nad niezgodnymi w stosunku do wszystkich par i wynosi 45%. Wartość statystyki Gamma równa 0,45 oznacza, że znając zmienne niezależne redukujemy błąd szacunku reagowania o około 45%. Wartość statystyki c określa możliwości prognostyczne zmiennej zależnej i wynosi 0,725. Jest równa polu pod krzywą ROC.

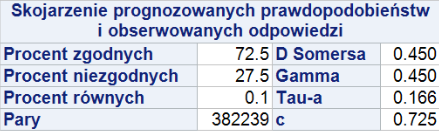


Tabela 67. Ocena jakości klasyfikacji obserwacji.

Ocena jakości klasyfikacji zostanie przeprowadzona w oparciu o wygenerowaną macierz trafności. (Tabela nr 68). Analizując tabelę klasyfikacji wnioskuje się, że przy domyślnym progu odcięcia na poziomie 0,5 model poprawnie sklasyfikował jedynki dla 1050 obserwacji, natomiast poprawnie sklasyfikował zera dla 55 obserwacji. Niepoprawne klasyfikacje przedstawiają się następująco:

-296 obserwacji zostało sklasyfikowanych jako jedynki, a w rzeczywistości są zerami,

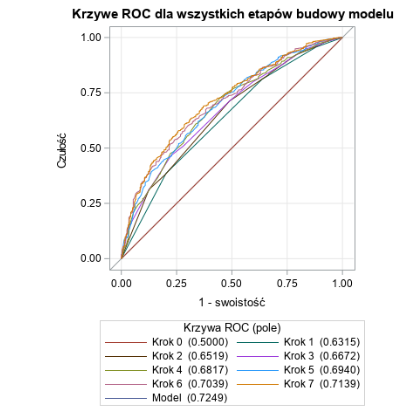
-39 obserwacji zostało sklasyfikowanych jako zera, a w rzeczywistości są jedynkami.

Trafność modelu, czyli stosunek poprawnie zakwalifikowanych obserwacji do wszystkich obserwacji, wyniosła 76,7%. Odnosząc tą wartość do modelu o zerowych wartościach prognostycznych (randomowego), który poprawnie klasyfikuje około 50% obserwacji, można powiedzieć, że otrzymana trafność jest wyraźnie wyższa. Kolejną miarą przedstawioną w Tabeli nr 68 jest czułość, która mówi o procencie obserwacji zaklasyfikowanych jako 1, które w rzeczywistości było jedynkami z wszystkich obserwacji sklasyfikowanych jako 1. Czułość naszego modelu wynosi 96,4%. Z kolei specyficzność to procent obserwacji zaklasyfikowanych jako 0 i w rzeczywistości będących zerami z wszystkich obserwacji sklasyfikowanych jako 0. Specyficzność naszego modelu wynosi 15,7%. Wnioskuje się, że przy progu odcięcia 0,5 model lepiej prognozuje jedynki. Zmieniając wartości punktu odcięcia można wpływać na wartości wskaźników czułości i specyficzności. Możliwe jest zwiększenie wartości czułości kosztem wartości specyficzności i odwrotnie. Przy ustalaniu progu odcięcia powinniśmy zastanowić się, czy bardziej zależy nam na przewidywaniu 0 czy 1. W kolejnym etapie omówiono krzywą ROC, która prezentuje wszystkie możliwe kombinacje czułości i specyficzności dla różnych progów odcięcia.



Tabela 68. Ocena klasyfikacji obserwacji jako poprawne i niepoprawne

## Krzywa ROC

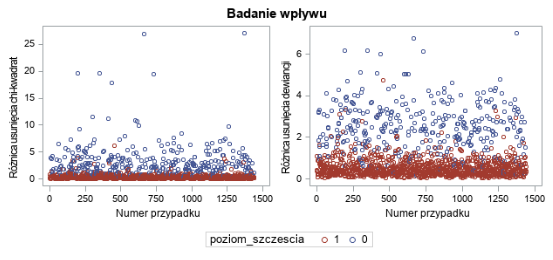


Wykres 35. Krzywa ROC dla wszystkich etapów budowy modelu.

Wartość pod krzywą ROC modelu finalnego wynosi 0,7249. Model jest zatem klasyfikatorem dobrej jakości. Krzywa ROC przedstawia zależność pomiędzy procentem prawidłowych prognoz, że badany Norweg czuje się szczęśliwy od odsetka prawidłowych prognoz, że badany Norweg czuje się nieszczęśliwy. Dobra jakość wyestymowanego modelu jest potwierdzona w relatywnie dużej odległości krzywej ROC właściwego modelu od krzywej 45 stopni dla modelu randomowego.

## Analiza reszt, miary wpływu

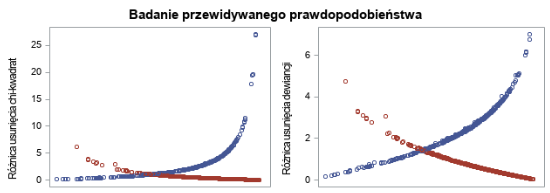
W tej części pracy przeanalizowano reszty oraz wybrane miary wpływu. Analiza reszt umożliwia identyfikację obserwacji, które zostały błędnie dopasowane w modelu. Natomiast miary wpływu informują o zmianie wartości konkretnej statystyki w wyniku usunięcia danej obserwacji. Używa się ich w celu identyfikacji jednostek odstających, silnie oddziałujących na oszacowania regresji.



Wykres 36. Badanie wpływu.

Wykres nr 36 przedstawia wpływ usunięcia poszczególnych obserwacji na zmianę wartości statystyk chi-kwadrat Pearsona oraz dewiancji. Obserwacje, dla których zaobserwowano najwyższe wartości różnic można uznać za odstające od reszty. Większy wpływ na zmianę wartości statystyki chi-Pearsona oraz dewiancji mają jednostki, dla których poziom szczęścia wynosi 0 (kolor niebieski). Na wykresie widzimy większe rozproszenie „niebieskich obserwacji”.

Następnie przedstawiono wykresy zależności zmiany dewiancji oraz wartości chi-kwadrat Pearsona od prognozowanego prawdopodobieństwa (Wykres nr 37). Wykresy pozwalają wykryć obserwacje odstające, odseparowane od innych punktów, mające duży wpływ na model. Analiza wykresów wskazuje, że obserwacje w prawym górnym rogu oraz lewym dolnym rogu mogą zostać uznane za wpływowe. Wykres ma typową budowę. Krzywa rosnąca odpowiada obserwacją, dla których zmienna celu przyjmuje wartość 0. Natomiast krzywa malejąca odpowiada obserwacjom ze zmienną zależną równą 1.



Wykres 37. Badanie przewidywanego prawdopodobieństwa.

# Podsumowanie i wnioski

Celem raportu przygotowanego przez zespół projektowy była analiza czynników determinujących poczucie szczęścia obywateli Norwegii oraz weryfikacji podstawionych hipotez badawczych. Do osiągnięcia postawionych celów wykorzystano model regresji logistycznej wykonany w środowisku SAS za pomocą narzędzia SAS Enterprise Guide.

Pierwszym zadaniem wykonanym przez zespół projektowy było wybranie zmiennych wykorzystanych następnie w modelu regresji logistycznej. W tym celu wykonany został przegląd literatury, w celu zidentyfikowania potencjalnych istotnych zmiennych. Dodatkowo zespół projektowy kierował się własną intuicją i część wybranych zmiennych została uwzględniona z powodu przypuszczenia o potencjalnych interesujących wynikach w kontekście Norwegii

Następnie została przeprowadzona eksploracja zbioru danych, po czym zbiór został oczyszczony z błędnych i niepełnych odpowiedzi, przeprocesowany i przygotowany do modelowania poziomu szczęścia.

W dalszym etapie zespół projektowy opisał każdą ze zmiennych wykorzystanych w modelu. Sprawdzono ich rozkłady oraz odpowiednio je przekodowano. Wygenerowane zostały tablice przedstawiające częstości i liczebności poszczególnych wartości lub kategorii przyjmowanych przez zmienne. Wygenerowano także histogramy w celu lepszego zobrazowania rozkładów wykorzystanych zmiennych. Dodano także tablice kontyngencji w celu analizy i wskazania wpływu zmiennych objaśniających na poczucie szczęścia.

Kolejnym krokiem było przeprowadzenie analizy współliniowości w celu zidentyfikowania potencjalnych silnych korelacji między zmiennymi lub kombinacji liniowych, które mogłyby zakłócić budowę modelu regresji logistycznej i wpłynąć negatywnie na precyzję oszacowania parametrów. W tym celu wykorzystano dwie miary: tolerancja zmiennej (TOL) oraz wskaźnik inflacji wariancji (VIF). Na podstawie wartości obydwu statystyk można stwierdzić, że pomiędzy zmiennymi wykorzystanymi w modelu nie występuje współliniowość.

Ostatnim etapem było zbudowanie i ocena jakości modelu regresji logistycznej. W początkowej fazie budowy modelu uwzględniono wszystkie wybrane wcześniej 13 zmiennych oraz ich efekty krzyżowe drugiego stopnia. W celu oceny jakości modelu wykorzystano kryteria dobroci dewiancji i Pearsona oraz kryteria informacyjne. Na podstawie wartości testu ilorazu wiarygodności i testu Walda zweryfikowano istotność zmiennych włączonych do modelu. Dodatkowo wygenerowano model uwzględniający tylko efekty główne i krokowo sprawdzano oceny parametrów przy usuwaniu z modelu poszczególnej zmiennej. Zdecydowano się na pozostawienie wszystkich zmiennych w modelu.

Następnie za pomocą metod forward (postępującej) i stepwise (postępującej) stworzono i porównano dwa modele. Na podstawie analizy wartości kryteriów dobroci wariancji i Pearsona, testu ilorazu wiarygodności oraz kryteriów informacyjnych wybrano model cechujący się większą precyzją oceny wartości parametrów – model wykonany metodą postępującą.

Z 13 początkowych zmiennych model został zredukowany do 8 istotnych statystycznie parametrów – 7 głównych oraz 1 krzyżowego. Po wyborze końcowego modelu, wykonano interpretację jego parametrów. Zinterpretowano ilorazy szans oraz oceniono jakość dopasowania modelu. W tym celu sprawdzono wartość testu Hosmera-Lemeshowa, wykorzystano analizę reszt i miary wpływu. Przeprowadzona została analiza dyskryminacyjną, sprawdzono także wartość pola pod krzywą ROC.

W początkowej części raportu w rozdziale zespół projektowe postawił dwie hipotezy dotyczące badanego zjawiska – poziomu szczęścia wśród obywateli Norwegii.

Według pierwszej postawionej hipotezy wysoki poziom religijności zwiększa szansę na poczucie wysokiego poziomu szczęścia. Wyniki końcowego wyestymowanego modelu wskazują na nieistotność statystyczną wykorzystanej w modelu zmiennej *religijność*. Oznacza to, że w Norwegii osoby deklarujące wysoki poziom religijności nie mają statystycznie większych szans na bycie szczęśliwymi w stosunku do osób nie deklarujących wysokiego poziomu religijności przy założeniu, że osoby nie różnią się pod względem wartości pozostałych parametrów uwzględnionych w modelu (*ceteris paribus*).

Według drugiej postawionej hipotezy duża liczba godzin w tygodniu spędzonych na pracy zmniejsza szansę na poczucie wysokiego poziomu szczęścia. Wyniki końcowego wyestymowanego modelu wskazują na nieistotność statystyczną włączonej w początkowej fazie do zmiennej *godz\_pracy*. Oznacza to, że w Norwegii osoby deklarujące większą od przeciętnej liczbę godzin pracy w tygodniu nie mają statystycznie mniejszych szans na bycie szczęśliwymi w stosunku do osób deklarujących mniejszą od przeciętnej liczbę godzin pracy w tygodniu przy założeniu, że osoby nie różnią się pod względem wartości pozostałych parametrów uwzględnionych w modelu (*ceteris paribus*).

Nie potwierdziły się zatem obydwa przypuszczenia postawione przez zespół projektowy. Okazuje się, że w społeczeństwie Norwegii dużo ważniejszymi czynnikami wpływającymi na poczucie szczęścia są między innymi stan zdrowia, przywiązanie do kraju, nadzór nad innymi pracownikami, duża liczba osób w gospodarstwie domowym, wiek, satysfakcja ze stanu demokracji w kraju i region zamieszkania.

# Bibliografia

A. Krause, “*Happiness and Work”*, IZA, 2014

A. Strużyńska- Kułajowicz ‘*’Czy władza szczęście daje? Władza a nastrój, emocje i satysfakcja z życia*’’, Wyższa Szkoła Psychologii Społecznej w Sopocie, Sopot 2010

B. Stevenson, J. Wolfers, “The Paradox of Declining Female Happiness”, 2009

C. Batz, L. Tay, “*Gender Differences in Subjective Well-Being*”, 2017

Centre for Bhutan Studies & GNH Research, *“A Compass Towards a Just and Harmonious Society”,* Thimpiu 2015

D. G. Blanchflower, A. J. Oswald, Well-being over time in Britain and the USA, 2004, Journal of Public Economics 88, 1359-1386.

D. G. Blanchflower, A. J. Oswald, Is well-being U-shaped over the life cycle?, 2008 Social Science and Medicine 66, 1733-1749.

J. Hartoga, H. Oosterbeek, „*Health, wealth, and happiness: why pursue a higher eduation?*”, Economics of Education Review, Vol. 17, No. 3, Amsterdam 1998

J. F. Helliwell, R. Layard, J. D. Sachs, “*World happiness report*”, 2019

J. F. Terelak, A. Borkowska, *“Satysfakcja z pracy jako źródło dobrostanu psychicznego u polskich emigrantów konsumpcyjnych w Kanadzie”*, Warszawa 2007.

H. Tajfel, „*Human groups and social categories”*, Cambridge University Press 1981,

M. Morrison, L. Tay, E. Diener, „*Subjective well-being and national satisfaction: Findings from a worldwide survey”*, Psychological Science 2011

[S. Doughty.](https://www.dailymail.co.uk/home/search.html?s=&authornamef=Steve+Doughty+Social+Affairs+Correspondent+For+The+Daily+Mail)  „*A big family 'is the key to a happy life: Two thirds who live in household with three or more children consider themselves to be satisfied with their lives*“.

T. A. Wright, K. K. Benett, T. Dun, “*Life and Job Satisfaction*”, Nevada 1989.

T. Kanasz, “*Uwarunkowania szczęścia*”, Wydawnictwo Akademii Pedagogiki Specjalnej, Warszawa 2015

https://pdf.helion.pl/khygge/khygge.pdf (data odsłony: 03.05.2019 r.)

https://www.norwegofil.pl/ludzie-i-tradycje/hygge-i-kos-norweska-metoda-na-szczescie   
(data odsłony: 03.05.2019 r.)

https://www.dailymail.co.uk/news/article-3106612/A-big-family-key-happy-life-Two-thirds-live-household-three-children-consider-satisfied-lives.html (data odsłony: 01.05.2019 r.)

https://medium.com/@andrewmerle/this-is-how-many-hours-you-should-really-be-working-ff1e8a6ad958 (data odsłony: 5.05.2019 r.)

https://kirken.no/nb-NO/church-of-norway/about/basics-and-statistics/ (data odsłony: 04.05.2019 r.).

http://worldpopulationreview.com/countries/norway-population/, (data odsłony: 01.05.2019 r.).

https://www.theatlantic.com/health/archive/2017/03/seasonal-affective-disorder-mosaic/519495/, (data odsłony: 01.05.2019 r.).

http://support.sas.com/kb/22/599.html (data odsłony: 05.05.2019 r.)

# Spis tabel

[Tabela 1. Zestawienie zmiennych wykorzystanych w modelu. 11](#_Toc8217040)

[Tabela 2. Zestawienie braków danych. 12](#_Toc8217041)

[Tabela 3. Zestawienie wartości przyjmowanych przez zmienne wykorzystane w modelu. 13](#_Toc8217042)

[Tabela 4. Liczebność i częstość zmiennej poziom\_szczescia. 15](#_Toc8217043)

[Tabela 5. Liczebność i częstość zmiennej poziom\_szczescia po dychotomizacji. 16](#_Toc8217044)

[Tabela 6. Liczebność i częstość zmiennej osob\_gosp. 17](#_Toc8217045)

[Tabela 7. Statystyki dla zmiennej osob\_gosp. 18](#_Toc8217046)

[Tabela 8. Statystyki dla zmiennej godz\_pracy. 20](#_Toc8217047)

[Tabela 9. Liczebność i częstość zmiennej godz\_pracy. 21](#_Toc8217048)

[Tabela 10. Liczebność i częstość zmiennej plec. 24](#_Toc8217049)

[Tabela 11. Tablica kontyngencji dla zmiennej plec. 25](#_Toc8217050)

[Tabela 12. Statystyki opisowe dla zmiennej wiek. 26](#_Toc8217051)

[Tabela 13. Liczebność i częstość zmiennej wiek. 27](#_Toc8217052)

[Tabela 14. Liczebność i częstość zmiennej wiek po kategoryzacji. 29](#_Toc8217053)

[Tabela 15. Tabela kontyngencji dla zmiennych poziom\_szczescia i wiek. 30](#_Toc8217054)

[Tabela 16. Liczebność i częstość zmiennej przelozony przed kategoryzacją. 31](#_Toc8217055)

[Tabela 17. Tabela kontyngencji dla zmiennych poziom\_szczescia i przelozony przed kategoryzacją. 32](#_Toc8217056)

[Tabela 18. Liczebność i częstość zmiennej przelozony po kategoryzacji. 32](#_Toc8217057)

[Tabela 19. Tablica kontyngencji dla zmiennych poziom\_szczescia i przelozony po kategoryzacji. 33](#_Toc8217058)

[Tabela 20. Liczebność i częstość zmiennej typ\_zakladu przed kategoryzacją. 34](#_Toc8217059)

[Tabela 21. Tablica kontyngencji dla zmiennych poziom\_szczescia i typ\_zakladu. 35](#_Toc8217060)

[Tabela 22. Liczebność i częstość zmiennej typ\_zakladu po rekategoryzacji. 36](#_Toc8217061)

[Tabela 23. Tablica kontyngencji dla zmiennych poziom\_szczescia i typ\_zakladu. 36](#_Toc8217062)

[Tabela 24. Liczebność i częstość zmiennej lata\_edu. 37](#_Toc8217063)

[Tabela 25. Statystyki opisowe dla zmienna lata\_edu. 38](#_Toc8217064)

[Tabela 26. Liczebność i częstość zmiennej lata\_edu po kategoryzacji. 39](#_Toc8217065)

[Tabela 27. Tablica kontyngencji dla zmiennej lata\_edu po kategoryzacji. 39](#_Toc8217066)

[Tabela 28. Liczebność i częstość zmiennej religijność przed kategoryzacją. 41](#_Toc8217067)

[Tabela 29. Tablica kontyngencji dla zmiennej poziom\_szczescia i religijnosc przed kategoryzacją. 42](#_Toc8217068)

[Tabela 30. Liczebność i częstość zmiennej religijnosc po kategoryzacji. 42](#_Toc8217069)

[Tabela 31. Tablica kontyngencji dla zmiennej poziom\_szczescia i religijnosc po kategoryzacji. 43](#_Toc8217070)

[Tabela 32. Liczebność i częstość zmiennej region 45](#_Toc8217071)

[Tabela 33. Tablica kontyngencji dla zmiennej poziom\_szczescia i region. 45](#_Toc8217072)

[Tabela 34. Liczebność i częstość zmiennej zdrowie przed ponowną kategoryzacją. 47](#_Toc8217073)

[Tabela 35. Tablica kontyngencji zmiennych poziom\_szczescia i zdrowie przed ponowną kategoryzacją. 47](#_Toc8217074)

[Tabela 36. Liczebność i częstość zmiennej zdrowie po ponownej kategoryzacji. 48](#_Toc8217075)

[Tabela 37. Tablica kontyngencji dla zmiennych poziom\_szczescia i zdrowie po ponownej kategoryzacji. 49](#_Toc8217076)

[Tabela 38. Liczebność i częstość zmiennej satysf\_dem przed kategoryzacją. 50](#_Toc8217077)

[Tabela 39. Tabela kontyngencji dla zmiennych poziom\_szczescia i satysf\_dem przed kategoryzacją. 51](#_Toc8217078)

[Tabela 40. Liczebność i częstość zmiennej satysf\_dem po kategoryzacji. 51](#_Toc8217079)

[Tabela 41. Tablica kontyngencji dla zmiennych poziom\_szczescia i satysf\_dem po kategoryzacji. 52](#_Toc8217080)

[Tabela 42. Liczebność i częstość zmiennej przyw\_kraj przed kategoryzacją. 53](#_Toc8217081)

[Tabela 43. Tablica kontyngencji dla zmiennej poziom\_szczescia i przyw\_kraj przed kategoryzacją. 54](#_Toc8217082)

[Tabela 44. Liczebność i częstość zmiennej przyw\_kraj. 55](#_Toc8217083)

[Tabela 45. Tablica kontyngencji dla zmiennej poziom\_szczescia i przyw\_kraj po kategoryzacji. 55](#_Toc8217084)

[Tabela 46. Liczebność i częstość zmiennej doch\_gosp przed kategoryzacją. 56](#_Toc8217085)

[Tabela 47. Tablica kontyngencji dla zmiennych poziom\_szczescia oraz doch\_gosp przed kategoryzacją. 57](#_Toc8217086)

[Tabela 48. Liczebność i częstość dla zmiennej doch\_gosp po rekategoryzacji. 57](#_Toc8217087)

[Tabela 49. Tablica kontyngencji dla zmiennych poziom\_szczescia i doch\_gosp po rekategoryzacji. 58](#_Toc8217088)

[Tabela 50. Wartości tolerancji. 61](#_Toc8217089)

[Tabela 51. Wartości inflacji wariancji. 62](#_Toc8217090)

[Tabela 52. Informacje o modelu regresji logistycznej. 63](#_Toc8217091)

[Tabela 53. Status zbieżności modelu regresji logistycznej. 63](#_Toc8217092)

[Tabela 54. Statystyki dobroci dopasowania modelu regresji logistycznej. 63](#_Toc8217093)

[Tabela 55. Statystyki testu Walda w modelu regresji logistycznej. 66](#_Toc8217094)

[Tabela 56. Statystyki dopasowania modelu z efektami głównymi. 66](#_Toc8217095)

[Tabela 57. Ocena istotności efektów głównych. 67](#_Toc8217096)

[Tabela 58. Względna zmiana ocen parametrów po usunięciu danej zmiennej z modelu. 67](#_Toc8217097)

[Tabela 59. Statystyki dopasowania modelu regresji logistycznej uzyskanego metodą forward i stepwise. 68](#_Toc8217098)

[Tabela 60. Statystyki dopasowania modelu regresji logistycznej uzyskanego metodą forward. 68](#_Toc8217099)

[Tabela 61. Statystyki istotności modelu regresji logistycznej. 69](#_Toc8217100)

[Tabela 62. Podsumowanie postępującej selekcji zmiennych. 69](#_Toc8217101)

[Tabela 63. Analiza ocen maksymalnej wiarygodności. 70](#_Toc8217102)

[Tabela 64. Oceny ilorazów szans i przedziały ufności Walda. 71](#_Toc8217103)

[Tabela 65. Statystyki dopasowania dewiancji i Pearsona. 74](#_Toc8217104)

[Tabela 66. Statystyka Chi-kwadrat dla testu zgodności Hosmera i Lemeshowa. 74](#_Toc8217105)

[Tabela 67. Ocena jakości klasyfikacji obserwacji. 75](#_Toc8217106)

[Tabela 68. Ocena klasyfikacji obserwacji jako poprawne i niepoprawne 76](#_Toc8217107)

# Spis wykresów

[Wykres 1. Histogram zmiennej poziom\_szczescia. 15](#_Toc8217002)

[Wykres 2. Histogram zmiennej poziom\_szczescia po dychotomizacji. 16](#_Toc8217003)

[Wykres 3. Histogram dla zmiennej osob\_gosp. 18](#_Toc8217004)

[Wykres 4. Wykres pudełkowy dla zmiennej osob\_gosp. 18](#_Toc8217005)

[Wykres 5. Histogram dla zmiennej osob\_gosp przy skategoryzowanej zmiennej poziom\_szcescia. 19](#_Toc8217006)

[Wykres 6. Histogram dla zmiennej godz\_pracy. 22](#_Toc8217007)

[Wykres 7. Wykres pudełkowy dla zmiennej godz\_pracy. 22](#_Toc8217008)

[Wykres 8. Histogram dla zmiennej godz\_pracy przy skategoryzowanej zmiennej poziom\_szcescia. 23](#_Toc8217009)

[Wykres 9. Histogram dla zmiennej plec. 24](#_Toc8217010)

[Wykres 10. Histogram dla zmiennej wiek. 28](#_Toc8217011)

[Wykres 11. Wykres pudełkowy dla zmiennej wiek. 28](#_Toc8217012)

[Wykres 12. Histogram dla zmiennej wiek. 29](#_Toc8217013)

[Wykres 13. Histogram dla zmiennej wiek przy skategoryzowanej zmiennej poziom\_szcescia. 30](#_Toc8217014)

[Wykres 14. Histogram dla zmiennej przelozony przed kategoryzacją. 31](#_Toc8217015)

[Wykres 15. Histogram dla zmiennej przelozony. 32](#_Toc8217016)

[Wykres 16. Histogram dla zmiennej typ\_zakladu. 34](#_Toc8217017)

[Wykres 17. Rozkład zmiennej typ\_zakladu po rekategoryzacji. 35](#_Toc8217018)

[Wykres 18. Histogram dla zmiennej lata\_edu. 37](#_Toc8217019)

[Wykres 19. Wykres pudełkowy dla zmiennej lata\_edu. 38](#_Toc8217020)

[Wykres 20. Histogram dla zmiennej lata\_edu. 39](#_Toc8217021)

[Wykres 21. Histogram dla zmiennej lata\_edu przy skategoryzowanej zmiennej poziom\_szcescia. 40](#_Toc8217022)

[Wykres 22. Histogram dla zmiennej religijność po kategoryzacji. 41](#_Toc8217023)

[Wykres 23. Histogram dla zmiennej religijnosc po kategoryzacji. 42](#_Toc8217024)

[Wykres 24. Histogram dla zmiennej religijnosc przy skategoryzowanej zmiennej poziom\_szcescia. 43](#_Toc8217025)

[Wykres 25. Histogram dla zmiennej region 45](#_Toc8217026)

[Wykres 26. Histogram dla zmiennej zdrowie przed ponowną kategoryzacją. 47](#_Toc8217027)

[Wykres 27. Histogram dla zmiennej zdrowie po ponownej kategoryzacji. 48](#_Toc8217028)

[Wykres 28. Histogram dla zmiennej satysf\_dem przed kategoryzacją. 50](#_Toc8217029)

[Wykres 29. Histogram dla zmiennej satysf\_dem po kategoryzacji. 51](#_Toc8217030)

[Wykres 30. Histogram dla zmiennej przyw\_kraj przed kategoryzacją. 53](#_Toc8217031)

[Wykres 31. Histogram dla zmiennej przyw\_kraj po rekategoryzacji. 54](#_Toc8217032)

[Wykres 32. Histogram dla zmiennej doch\_gosp przed kategoryzacją. 57](#_Toc8217033)

[Wykres 33. Histogram dla zmiennej doch\_gosp po rekategoryzacji. 58](#_Toc8217034)

[Wykres 34. Ilorazy szans z przedziałem ufności Walda. 72](#_Toc8217035)

[Wykres 35. Krzywa ROC dla wszystkich etapów budowy modelu. 76](#_Toc8217036)

[Wykres 36. Badanie wpływu. 77](#_Toc8217037)

[Wykres 37. Badanie przewidywanego prawdopodobieństwa. 78](#_Toc8217038)

# Spis rysunków

[Rysunek 1. Podział administracyjny Norwegii na regiony. 44](#_Toc8217039)

# Wykorzystany kod

/\*SEMESTR LETNI 2018/2019 - REGRESJA LOGISTYCZNA Z WYKORZYSTANIEM NARZĘDZI SAS\*/

/\*223480-1528\*/

/\*Lista autorów:\*/

/\*Michał Hajdan 68501\*/

/\*Anna Kozak 68570\*/

/\*Marcin Mandziej 68373\*/

/\*Eryk Mazuś 68835\*/

/\*Edyta Rybicka 67670\*/

/\*Sylwia Semeniuk 67978\*/

/\*Justyna Stanek 64235\*/

/\*Paweł Tafliński 67500\*/

/\*Projekt zaliczeniowy I – Model regresji logistycznej –

modelowanie poziomu szczęścia na przykładzie Norwegii\*/

/\*Należy utworzyć folder i umieścić w nim zbiór o nazwie ESS8NO,

a następnie wprowadzić poniżej odpowiednią ścieżkę\*/

libname projekt 'D:\Magisterka\Semestr II\Regresja logistyczna SAS\Nasz projekt';

/\*Wybór ziennych, zmiana nazw i etykiet\*/

**data** projekt.norwegia

(rename=(happy=poziom\_szczescia GNDR=plec AGEA=wiek HEALTH=zdrowie eduyrs=lata\_edu

HHMMB=osob\_gosp stfdem=satysf\_dem atchctr=przyw\_kraj rlgdgr=religijnosc

HINCTNTA=doch\_gosp WKHTOT=godz\_pracy TPORGWK=typ\_zakladu jbspv=przelozony)) ;

label

HAPPY='Subiektywna ocena poziomu szczęscia'

GNDR='Płeć'

AGEA='Wiek'

HEALTH='Subiektywna ocena stanu zdrowia'

EDUYRS='Liczba lat edukacji'

HHMMB='Liczba osób w gospodarstwie domowym'

STFDEM='Satysfakcja ze stanu demokracji w kraju'

ATCHCTR='Przywiązanie do kraju'

RLGDGR='Stopień religijności'

HINCTNTA='Dochód netto gospodarstwa domowego'

WKHTOT='Liczba godzin pracy w tygodniu uwzględniając nadgodziny'

TPORGWK='Typ zakładu pracy'

JBSPV='Nadzór nad innymi pracownikami'

REGION='Region'

;

keep happy GNDR AGEA HEALTH eduyrs HHMMB stfdem atchctr rlgdgr

HINCTNTA WKHTOT TPORGWK jbspv region;

set projekt.ESS8NO;

**run**;

/\*Tablice, liczebności przed usunięciem odpowiedzi typu nie wiem, braku odpowiedzi lub odmowy\*/

**proc** **freq** data=projekt.norwegia order=internal;

ods graphics on;

tables poziom\_szczescia plec wiek zdrowie lata\_edu

osob\_gosp satysf\_dem przyw\_kraj religijnosc doch\_gosp

godz\_pracy typ\_zakladu przelozony region /

plots=freqplot;

**run**;

/\*Usunięcie odpowiedzi nie wiem, braku odpowiedzi lub odmowy\*/

**data** projekt.norwegia2;

set projekt.norwegia;

if poziom\_szczescia in (**77**,**88**,**99**)

or plec in (**9**)

or wiek in (**999**)

or zdrowie in (**7**,**8**,**9**)

or lata\_edu in (**77**,**88**,**99**)

or osob\_gosp in (**77**,**88**,**99**)

or satysf\_dem in (**77**,**88**,**99**)

or przyw\_kraj in (**77**,**88**,**99**)

or religijnosc in (**77**,**88**,**99**)

or doch\_gosp in (**77**,**88**,**99**)

or godz\_pracy in (**777**,**888**,**999**)

or typ\_zakladu in (**77**,**88**,**99**)

or przelozony in (**7**,**8**,**9**)

or region='99999'

then delete;

**run**;

/\*Tablice i liczebności- sprawdzenie po usunięciu odpowiedzi nie wiem, braku odpowiedzi lub odmowy\*/

**proc** **freq** data=projekt.norwegia2 order=internal;

ods graphics on;

tables poziom\_szczescia plec wiek zdrowie lata\_edu

osob\_gosp satysf\_dem przyw\_kraj religijnosc doch\_gosp

godz\_pracy typ\_zakladu przelozony region /

plots=freqplot;

**run**;

/\*Sprawdzenie braków danych\*/

**proc** **means** data=projekt.norwegia2 nmiss;

**run**;

/\*Zmienne ciągłe\*/

/\*Statystyki opisowe\*/

**proc** **means** data=projekt.norwegia2

N min mean max Q1 median Q3 std skewness;

var osob\_gosp ;

**run**;

**proc** **means** data=projekt.norwegia2

N min mean max Q1 median Q3 std skewness;

var godz\_pracy;

**run**;

**proc** **means** data=projekt.norwegia2

N min mean max Q1 median Q3 std skewness;

var lata\_edu;

**run**;

**proc** **means** data=projekt.norwegia2

N min mean max Q1 median Q3 std skewness;

var wiek;

**run**;

/\*Wykres pudełkowy\*/

**proc** **sgplot** data=projekt.norwegia2;

vbox osob\_gosp;

**run**;

**proc** **sgplot** data=projekt.norwegia2;

vbox godz\_pracy ;

**run**;

**proc** **sgplot** data=projekt.norwegia2;

vbox lata\_edu ;

**run**;

**proc** **sgplot** data=projekt.norwegia2;

vbox wiek ;

**run**;

/\*Dychotomizacja szczescia dla tablic kontyngencji przed kategoryzacja \*/

**data** projekt.dychotomizacja;

set projekt.norwegia2;

if poziom\_szczescia <=**7** then poziom\_szczescia=**0** ;

else poziom\_szczescia=**1**;

**run**;

/\*Tablice kontyngencji przed kategoryzacja \*/

**proc** **freq** data=projekt.dychotomizacja;

tables

religijnosc\*poziom\_szczescia

zdrowie\*poziom\_szczescia

lata\_edu\*poziom\_szczescia

plec\*poziom\_szczescia

wiek\*poziom\_szczescia

satysf\_dem\*poziom\_szczescia

przyw\_kraj\*poziom\_szczescia

doch\_gosp\*poziom\_szczescia

typ\_zakladu\*poziom\_szczescia

przelozony\*poziom\_szczescia

region\*poziom\_szczescia

/ nocol nocum norow nopercent;

**run**;

/\*Kategoryzacja zmiennych\*/

**data** projekt.skategoryzowane;

set projekt.norwegia2;

if zdrowie in (**4**,**5**) then zdrowie=**4**;

if poziom\_szczescia <=**7** then poziom\_szczescia=**0** ;

else poziom\_szczescia=**1**;

if plec=**1** then plec=**0** ;

if plec=**2** then plec=**1**;

if wiek <=**22** then wiek=**1**;

if wiek>**22** and wiek<=**30** then wiek=**2**;

if wiek>**30** and wiek<=**45** then wiek=**3**;

if wiek>**45** and wiek<=**66** then wiek=**4**;

if wiek>**66** then wiek=**5**;

if lata\_edu<= **12** then lata\_edu= **1**;

if lata\_edu in (**13**,**14**) then lata\_edu=**2** ;

if lata\_edu in (**15**,**16**,**17**) then lata\_edu=**3** ;

if lata\_edu =>**18** then lata\_edu=**4**;

if satysf\_dem<=**3** then satysf\_dem=**1**;

if satysf\_dem in (**4**,**5**,**6**,**7**) then satysf\_dem=**2**;

if satysf\_dem =>**8** then satysf\_dem=**3**;

if przyw\_kraj <=**3** then przyw\_kraj=**1** ;

if przyw\_kraj in (**4**,**5**,**6**,**7**) then przyw\_kraj=**2**;

if przyw\_kraj =>**8** then przyw\_kraj=**3** ;

if religijnosc <=**2** then religijnosc=**1**;

if religijnosc in (**3**,**4**,**5**,**6**) then religijnosc=**2**;

if religijnosc >**6** then religijnosc=**3**;

if doch\_gosp <=**3** then doch\_gosp=**1**;

if doch\_gosp in (**4**,**5**,**6**) then doch\_gosp=**2**;

if doch\_gosp >**6** then doch\_gosp=**3**;

if godz\_pracy=**666** then godz\_pracy=**0** ;

if typ\_zakladu in(**1**,**2**,**3**) then typ\_zakladu=**1**;

if typ\_zakladu=**4** then typ\_zakladu=**2**;

if typ\_zakladu=**5** then typ\_zakladu=**3**;

if typ\_zakladu in(**6**,**66**) then typ\_zakladu=**4**;

if przelozony=**1** then przelozony=**1**;

if przelozony=**2** then przelozony=**0**;

if przelozony=**6** then przelozony=**0**;

if region='NO01' then region=**1**;

if region='NO02' then region=**2**;

if region='NO03' then region=**3**;

if region='NO04' then region=**4**;

if region='NO05' then region=**5**;

if region='NO06' then region=**6**;

if region='NO07' then region=**7**;

**run**;

/\*Zmiana typu zmiennej region ze znakowej na numeryczną\*/

**data** projekt.skategoryzowane;

set projekt.skategoryzowane;

region\_num=input(region,**8.**);

drop region;;

**run**;

**data** projekt.skategoryzowane;

set projekt.skategoryzowane;

rename region\_num=region;

**run**;

**data** projekt.skategoryzowane;

set projekt.skategoryzowane;

label region='Region zamieszkania';

**run**;

/\*Formaty\*/

**PROC** **FORMAT** lib=work;

value poziom\_szczescia **0** = 'Nieszczęśliwa/y' **1** = 'Szczęśliwa/y' ;

value plec **1**= 'Kobiety' **0**= 'Mezczyźni' ;

value wiek **1**='22 lat i mniej' **2**='23-30 lat' **3**='31-45 lat' **4**='46-66 lat' **5**='67 lat i więcej';

value zdrowie **1** = 'Bardzo dobre' **2** = 'Dobre' **3** = 'Umiarkowane' **4** = 'Złe' ;

value lata\_edu **1**= '12 lat i mniej' **2** = '13-14 lat' **3** = '15-17 lat' **4** = '18 lat i więcej';

value satysf\_dem **1** = 'Mała' **2** = 'Średnia' **3** = 'Duża';

value przyw\_kraj **1** = 'Małe' **2** = 'Średnie' **3** = 'Duże';

value religijnosc **1**='Niewierząca/y' **2**='Wierząca/y' **3**='Bardzo wierząca/y';

value doch\_gosp **1** = 'Niski' **2** = 'Średni' **3** = 'Wysoki';

value typ\_zakladu **1** = 'Sektor państwowy'

**2** = 'Sektor prywatny' **3**='Samozatrudnienie' **4**='Inne' ;

value przelozony **0**='Nie' **1**='Tak';

value region **1**='NO01' **2**='NO02' **3**='NO03' **4**='NO04'

**5**='NO05' **6**='NO06' **7**='NO07';

**run**;

/\*Po kategoryzaji z odpowiednim formatem \*/

**proc** **freq** data=projekt.skategoryzowane order=internal;

ods graphics on;

format poziom\_szczescia poziom\_szczescia.

plec plec. zdrowie zdrowie. lata\_edu lata\_edu. wiek wiek.

satysf\_dem satysf\_dem. przyw\_kraj przyw\_kraj.

religijnosc religijnosc. doch\_gosp doch\_gosp.

typ\_zakladu typ\_zakladu. przelozony przelozony. region region.

;

tables poziom\_szczescia plec wiek zdrowie lata\_edu

osob\_gosp satysf\_dem przyw\_kraj religijnosc doch\_gosp

godz\_pracy typ\_zakladu przelozony region /

plots=freqplot;

**run**;

/\*Tablice kontygencji po kategoryzacji\*/

**proc** **freq** data=projekt.skategoryzowane;

format poziom\_szczescia poziom\_szczescia.

plec plec. zdrowie zdrowie. lata\_edu lata\_edu. wiek wiek.

satysf\_dem satysf\_dem. przyw\_kraj przyw\_kraj.

religijnosc religijnosc. doch\_gosp doch\_gosp.

typ\_zakladu typ\_zakladu. przelozony przelozony. region region. ;

tables

religijnosc\*poziom\_szczescia

zdrowie\*poziom\_szczescia

lata\_edu\*poziom\_szczescia

plec\*poziom\_szczescia

wiek\*poziom\_szczescia

satysf\_dem\*poziom\_szczescia

przyw\_kraj\*poziom\_szczescia

doch\_gosp\*poziom\_szczescia

typ\_zakladu\*poziom\_szczescia

przelozony\*poziom\_szczescia

region\*poziom\_szczescia

/ nocol nocum norow nopercent;

**run**;

/\*Zmienne ciągłe oraz dla zm. z hipotez, tablice kontyngencji po kategoryzacji - histogramy\*/

title 'Poziom szczęścia w zależności od religijności';

**proc** **gchart** data=projekt.skategoryzowane;

format poziom\_szczescia poziom\_szczescia. religijnosc religijnosc.;

vbar religijnosc/ group=poziom\_szczescia midpoints=**1** **2** **3** ;

pattern color=blueviolet;

**run**;

title 'Poziom szczęścia w zależności od liczby godzin pracy';

**proc** **gchart** data=projekt.skategoryzowane;

format poziom\_szczescia poziom\_szczescia. godz\_pracy;

vbar godz\_pracy / group=poziom\_szczescia ;

pattern color=blueviolet;

**run**;

title 'Poziom szczęścia w zależności od wieku';

**proc** **gchart** data=projekt.skategoryzowane;

format poziom\_szczescia poziom\_szczescia. wiek wiek.;

vbar wiek / group=poziom\_szczescia ;

pattern color=blueviolet;

**run**;

title 'Poziom szczęścia w zależności od liczby osób w gospodarstwie domowych';

**proc** **gchart** data=projekt.skategoryzowane;

format poziom\_szczescia poziom\_szczescia. osob\_gosp;

vbar osob\_gosp / group=poziom\_szczescia ;

pattern color=blueviolet;

**run**;

title 'Poziom szczęścia w zależności od liczby lat edukacji';

**proc** **gchart** data=projekt.skategoryzowane;

format poziom\_szczescia poziom\_szczescia. lata\_edu lata\_edu.;

vbar lata\_edu / group=poziom\_szczescia ;

pattern color=blueviolet;

**run**;

/\*Badanie współliniowości\*/

**proc** **corr** data=projekt.skategoryzowane out=projekt.pearson; var poziom\_szczescia religijnosc

osob\_gosp

godz\_pracy

zdrowie

lata\_edu

plec

wiek

satysf\_dem

przyw\_kraj

doch\_gosp

typ\_zakladu

przelozony

region

;

**run**;

**proc** **reg** data=projekt.skategoryzowane ; model poziom\_szczescia=

religijnosc

osob\_gosp

godz\_pracy

zdrowie

lata\_edu

plec

wiek

satysf\_dem

przyw\_kraj

doch\_gosp

typ\_zakladu

przelozony

region

/

vif tol collin;

**run**;

/\*Reg logistyczna\*/

/\* -------------------------------------------------------------------

Kod wygenerowany przez zadanie SAS-a

Wygenerowany dnia: niedziela, 5 maja 2019 o godz. 21:09:11

Przez zadanie: Regresja logistyczna

Dane wejściowe: Local:PROJEKT.SKATEGORYZOWANE

Serwer: Local

------------------------------------------------------------------- \*/

ODS GRAPHICS ON;

%***\_eg\_conditional\_dropds***(WORK.PREDLogRegPredictions,

WORK.SORTTempTableSorted);

/\* -------------------------------------------------------------------

Sortowanie zbioru Local:PROJEKT.SKATEGORYZOWANE

------------------------------------------------------------------- \*/

**PROC** **SQL**;

CREATE VIEW WORK.SORTTempTableSorted AS

SELECT \*

FROM PROJEKT.SKATEGORYZOWANE

;

**QUIT**;

TITLE;

TITLE1 "Rezultaty regresji logistycznej";

FOOTNOTE;

FOOTNOTE1 "Wygenerowane przez System SAS (&\_SASSERVERNAME, &SYSSCPL) dnia %TRIM(%QSYSFUNC(DATE(), NLDATE20.)) o godz. %TRIM(%SYSFUNC(TIME(), NLTIMAP20.))";

**PROC** **LOGISTIC** DATA=WORK.SORTTempTableSorted

PLOTS(ONLY)=ALL

;

CLASS plec (PARAM=EFFECT) wiek (PARAM=EFFECT) zdrowie (PARAM=EFFECT) lata\_edu (PARAM=EFFECT) satysf\_dem (PARAM=EFFECT) przyw\_kraj (PARAM=EFFECT) religijnosc (PARAM=EFFECT) doch\_gosp (PARAM=EFFECT) typ\_zakladu (PARAM=EFFECT)

przelozony (PARAM=EFFECT) region (PARAM=EFFECT);

MODEL poziom\_szczescia (Event = '1')=osob\_gosp godz\_pracy plec wiek zdrowie lata\_edu satysf\_dem przyw\_kraj religijnosc doch\_gosp typ\_zakladu przelozony region osob\_gosp\*godz\_pracy osob\_gosp\*plec osob\_gosp\*wiek osob\_gosp\*zdrowie osob\_gosp\*lata\_edu osob\_gosp\*satysf\_dem osob\_gosp\*przyw\_kraj osob\_gosp\*religijnosc osob\_gosp\*doch\_gosp osob\_gosp\*typ\_zakladu osob\_gosp\*przelozony osob\_gosp\*region godz\_pracy\*plec godz\_pracy\*wiek godz\_pracy\*zdrowie godz\_pracy\*lata\_edu godz\_pracy\*satysf\_dem godz\_pracy\*przyw\_kraj godz\_pracy\*religijnosc godz\_pracy\*doch\_gosp godz\_pracy\*typ\_zakladu godz\_pracy\*przelozony godz\_pracy\*region plec\*wiek plec\*zdrowie plec\*lata\_edu plec\*satysf\_dem plec\*przyw\_kraj plec\*religijnosc plec\*doch\_gosp plec\*typ\_zakladu plec\*przelozony plec\*region wiek\*zdrowie wiek\*lata\_edu wiek\*satysf\_dem wiek\*przyw\_kraj wiek\*religijnosc wiek\*doch\_gosp wiek\*typ\_zakladu wiek\*przelozony wiek\*region zdrowie\*lata\_edu zdrowie\*satysf\_dem zdrowie\*przyw\_kraj zdrowie\*religijnosc zdrowie\*doch\_gosp zdrowie\*typ\_zakladu zdrowie\*przelozony zdrowie\*region lata\_edu\*satysf\_dem lata\_edu\*przyw\_kraj lata\_edu\*religijnosc lata\_edu\*doch\_gosp lata\_edu\*typ\_zakladu lata\_edu\*przelozony lata\_edu\*region satysf\_dem\*przyw\_kraj satysf\_dem\*religijnosc satysf\_dem\*doch\_gosp satysf\_dem\*typ\_zakladu satysf\_dem\*przelozony satysf\_dem\*region przyw\_kraj\*religijnosc przyw\_kraj\*doch\_gosp przyw\_kraj\*typ\_zakladu przyw\_kraj\*przelozony przyw\_kraj\*region religijnosc\*doch\_gosp religijnosc\*typ\_zakladu religijnosc\*przelozony religijnosc\*region doch\_gosp\*typ\_zakladu doch\_gosp\*przelozony doch\_gosp\*region typ\_zakladu\*przelozony typ\_zakladu\*region przelozony\*region /

SELECTION=NONE

CORRB

COVB

INFLUENCE

LACKFIT

AGGREGATE SCALE=NONE

RSQUARE

LINK=LOGIT

CLPARM=BOTH

CLODDS=BOTH

ALPHA=**0.05**

;

OUTPUT OUT=WORK.PREDLogRegPredictions(LABEL="Statystyki i prognozy regresji logistycznej dla PROJEKT.SKATEGORYZOWANE")

PREDPROBS=INDIVIDUAL;

**RUN**;

**QUIT**;

/\* -------------------------------------------------------------------

Koniec kodu zadania

------------------------------------------------------------------- \*/

**RUN**; **QUIT**;

%***\_eg\_conditional\_dropds***(WORK.SORTTempTableSorted);

TITLE; FOOTNOTE;

ODS GRAPHICS OFF;

/\*Model tylko z efektami głównymi\*/

/\* -------------------------------------------------------------------

Kod wygenerowany przez zadanie SAS-a

Wygenerowany dnia: środa, 8 maja 2019 o godz. 15:44:10

Przez zadanie: Regresja logistyczna

Dane wejściowe: Local:PROJEKT.SKATEGORYZOWANE

Serwer: Local

------------------------------------------------------------------- \*/

ODS GRAPHICS ON;

%***\_eg\_conditional\_dropds***(WORK.SORTTempTableSorted,

WORK.TMP2TempTablePredictions);

/\* -------------------------------------------------------------------

Sortowanie zbioru Local:PROJEKT.SKATEGORYZOWANE

------------------------------------------------------------------- \*/

**PROC** **SQL**;

CREATE VIEW WORK.SORTTempTableSorted AS

SELECT T.poziom\_szczescia, T.osob\_gosp, T.godz\_pracy, T.plec, T.wiek, T.zdrowie, T.lata\_edu, T.satysf\_dem, T.przyw\_kraj, T.religijnosc, T.doch\_gosp, T.typ\_zakladu, T.przelozony, T.region

FROM PROJEKT.SKATEGORYZOWANE as T

;

**QUIT**;

TITLE;

TITLE1 "Rezultaty regresji logistycznej";

FOOTNOTE;

FOOTNOTE1 "Wygenerowane przez System SAS (&\_SASSERVERNAME, &SYSSCPL) dnia %TRIM(%QSYSFUNC(DATE(), NLDATE20.)) o godz. %TRIM(%SYSFUNC(TIME(), NLTIMAP20.))";

**PROC** **LOGISTIC** DATA=WORK.SORTTempTableSorted

PLOTS(ONLY)=ALL

;

CLASS plec (PARAM=EFFECT) wiek (PARAM=EFFECT) zdrowie (PARAM=EFFECT) lata\_edu (PARAM=EFFECT) satysf\_dem (PARAM=EFFECT) przyw\_kraj (PARAM=EFFECT) religijnosc (PARAM=EFFECT) doch\_gosp (PARAM=EFFECT) typ\_zakladu (PARAM=EFFECT)

przelozony (PARAM=EFFECT) region (PARAM=EFFECT);

MODEL poziom\_szczescia (Event = '1')=osob\_gosp godz\_pracy plec wiek zdrowie lata\_edu satysf\_dem przyw\_kraj religijnosc doch\_gosp typ\_zakladu przelozony region /

SELECTION=NONE

CORRB

COVB

INFLUENCE

LACKFIT

AGGREGATE SCALE=NONE

RSQUARE

CTABLE

LINK=LOGIT

CLPARM=BOTH

CLODDS=BOTH

ALPHA=**0.05**

;

OUTPUT OUT=WORK.TMP2TempTablePredictions

PREDPROBS=INDIVIDUAL;

**RUN**;

**QUIT**;

/\* -------------------------------------------------------------------

Koniec kodu zadania

------------------------------------------------------------------- \*/

**RUN**; **QUIT**;

%***\_eg\_conditional\_dropds***(WORK.SORTTempTableSorted,

WORK.TMP2TempTablePredictions);

TITLE; FOOTNOTE;

ODS GRAPHICS OFF;

/\*Odrzucenie zmiennej typ\_zakladu\*/

/\* -------------------------------------------------------------------

Kod wygenerowany przez zadanie SAS-a

Wygenerowany dnia: środa, 8 maja 2019 o godz. 15:51:12

Przez zadanie: Regresja logistyczna

Dane wejściowe: Local:PROJEKT.SKATEGORYZOWANE

Serwer: Local

------------------------------------------------------------------- \*/

ODS GRAPHICS ON;

%***\_eg\_conditional\_dropds***(WORK.SORTTempTableSorted,

WORK.TMP2TempTablePredictions);

/\* -------------------------------------------------------------------

Sortowanie zbioru Local:PROJEKT.SKATEGORYZOWANE

------------------------------------------------------------------- \*/

**PROC** **SQL**;

CREATE VIEW WORK.SORTTempTableSorted AS

SELECT T.poziom\_szczescia, T.osob\_gosp, T.godz\_pracy, T.plec, T.wiek, T.zdrowie, T.lata\_edu, T.satysf\_dem, T.przyw\_kraj, T.religijnosc, T.doch\_gosp, T.typ\_zakladu, T.przelozony, T.region

FROM PROJEKT.SKATEGORYZOWANE as T

;

**QUIT**;

TITLE;

TITLE1 "Rezultaty regresji logistycznej";

FOOTNOTE;

FOOTNOTE1 "Wygenerowane przez System SAS (&\_SASSERVERNAME, &SYSSCPL) dnia %TRIM(%QSYSFUNC(DATE(), NLDATE20.)) o godz. %TRIM(%SYSFUNC(TIME(), NLTIMAP20.))";

**PROC** **LOGISTIC** DATA=WORK.SORTTempTableSorted

PLOTS(ONLY)=ALL

;

CLASS plec (PARAM=EFFECT) wiek (PARAM=EFFECT) zdrowie (PARAM=EFFECT) lata\_edu (PARAM=EFFECT) satysf\_dem (PARAM=EFFECT) przyw\_kraj (PARAM=EFFECT) religijnosc (PARAM=EFFECT) doch\_gosp (PARAM=EFFECT) typ\_zakladu (PARAM=EFFECT)

przelozony (PARAM=EFFECT) region (PARAM=EFFECT);

MODEL poziom\_szczescia (Event = '1')=osob\_gosp godz\_pracy plec wiek zdrowie lata\_edu satysf\_dem przyw\_kraj religijnosc doch\_gosp przelozony region /

SELECTION=NONE

CORRB

COVB

INFLUENCE

LACKFIT

AGGREGATE SCALE=NONE

RSQUARE

CTABLE

LINK=LOGIT

CLPARM=BOTH

CLODDS=BOTH

ALPHA=**0.05**

;

OUTPUT OUT=WORK.TMP2TempTablePredictions

PREDPROBS=INDIVIDUAL;

**RUN**;

**QUIT**;

/\* -------------------------------------------------------------------

Koniec kodu zadania

------------------------------------------------------------------- \*/

**RUN**; **QUIT**;

%***\_eg\_conditional\_dropds***(WORK.SORTTempTableSorted,

WORK.TMP2TempTablePredictions);

TITLE; FOOTNOTE;

ODS GRAPHICS OFF;

/\*Odrzucenie zmiennej plec\*/

/\* -------------------------------------------------------------------

Kod wygenerowany przez zadanie SAS-a

Wygenerowany dnia: środa, 8 maja 2019 o godz. 15:53:43

Przez zadanie: Regresja logistyczna

Dane wejściowe: Local:PROJEKT.SKATEGORYZOWANE

Serwer: Local

------------------------------------------------------------------- \*/

ODS GRAPHICS ON;

%***\_eg\_conditional\_dropds***(WORK.SORTTempTableSorted,

WORK.TMP2TempTablePredictions);

/\* -------------------------------------------------------------------

Sortowanie zbioru Local:PROJEKT.SKATEGORYZOWANE

------------------------------------------------------------------- \*/

**PROC** **SQL**;

CREATE VIEW WORK.SORTTempTableSorted AS

SELECT T.poziom\_szczescia, T.osob\_gosp, T.godz\_pracy, T.plec, T.wiek, T.zdrowie, T.lata\_edu, T.satysf\_dem, T.przyw\_kraj, T.religijnosc, T.doch\_gosp, T.typ\_zakladu, T.przelozony, T.region

FROM PROJEKT.SKATEGORYZOWANE as T

;

**QUIT**;

TITLE;

TITLE1 "Rezultaty regresji logistycznej";

FOOTNOTE;

FOOTNOTE1 "Wygenerowane przez System SAS (&\_SASSERVERNAME, &SYSSCPL) dnia %TRIM(%QSYSFUNC(DATE(), NLDATE20.)) o godz. %TRIM(%SYSFUNC(TIME(), NLTIMAP20.))";

**PROC** **LOGISTIC** DATA=WORK.SORTTempTableSorted

PLOTS(ONLY)=ALL

;

CLASS plec (PARAM=EFFECT) wiek (PARAM=EFFECT) zdrowie (PARAM=EFFECT) lata\_edu (PARAM=EFFECT) satysf\_dem (PARAM=EFFECT) przyw\_kraj (PARAM=EFFECT) religijnosc (PARAM=EFFECT) doch\_gosp (PARAM=EFFECT) typ\_zakladu (PARAM=EFFECT)

przelozony (PARAM=EFFECT) region (PARAM=EFFECT);

MODEL poziom\_szczescia (Event = '1')=osob\_gosp godz\_pracy wiek zdrowie lata\_edu satysf\_dem przyw\_kraj religijnosc doch\_gosp przelozony region typ\_zakladu /

SELECTION=NONE

CORRB

COVB

INFLUENCE

LACKFIT

AGGREGATE SCALE=NONE

RSQUARE

CTABLE

LINK=LOGIT

CLPARM=BOTH

CLODDS=BOTH

ALPHA=**0.05**

;

OUTPUT OUT=WORK.TMP2TempTablePredictions

PREDPROBS=INDIVIDUAL;

**RUN**;

**QUIT**;

/\* -------------------------------------------------------------------

Koniec kodu zadania

------------------------------------------------------------------- \*/

**RUN**; **QUIT**;

%***\_eg\_conditional\_dropds***(WORK.SORTTempTableSorted,

WORK.TMP2TempTablePredictions);

TITLE; FOOTNOTE;

ODS GRAPHICS OFF;

/\*Odrzucenie zmiennej doch\_gosp\*/

/\* -------------------------------------------------------------------

Kod wygenerowany przez zadanie SAS-a

Wygenerowany dnia: środa, 8 maja 2019 o godz. 15:55:12

Przez zadanie: Regresja logistyczna

Dane wejściowe: Local:PROJEKT.SKATEGORYZOWANE

Serwer: Local

------------------------------------------------------------------- \*/

ODS GRAPHICS ON;

%***\_eg\_conditional\_dropds***(WORK.SORTTempTableSorted,

WORK.TMP2TempTablePredictions);

/\* -------------------------------------------------------------------

Sortowanie zbioru Local:PROJEKT.SKATEGORYZOWANE

------------------------------------------------------------------- \*/

**PROC** **SQL**;

CREATE VIEW WORK.SORTTempTableSorted AS

SELECT T.poziom\_szczescia, T.osob\_gosp, T.godz\_pracy, T.plec, T.wiek, T.zdrowie, T.lata\_edu, T.satysf\_dem, T.przyw\_kraj, T.religijnosc, T.doch\_gosp, T.typ\_zakladu, T.przelozony, T.region

FROM PROJEKT.SKATEGORYZOWANE as T

;

**QUIT**;

TITLE;

TITLE1 "Rezultaty regresji logistycznej";

FOOTNOTE;

FOOTNOTE1 "Wygenerowane przez System SAS (&\_SASSERVERNAME, &SYSSCPL) dnia %TRIM(%QSYSFUNC(DATE(), NLDATE20.)) o godz. %TRIM(%SYSFUNC(TIME(), NLTIMAP20.))";

**PROC** **LOGISTIC** DATA=WORK.SORTTempTableSorted

PLOTS(ONLY)=ALL

;

CLASS plec (PARAM=EFFECT) wiek (PARAM=EFFECT) zdrowie (PARAM=EFFECT) lata\_edu (PARAM=EFFECT) satysf\_dem (PARAM=EFFECT) przyw\_kraj (PARAM=EFFECT) religijnosc (PARAM=EFFECT) doch\_gosp (PARAM=EFFECT) typ\_zakladu (PARAM=EFFECT)

przelozony (PARAM=EFFECT) region (PARAM=EFFECT);

MODEL poziom\_szczescia (Event = '1')=osob\_gosp godz\_pracy wiek zdrowie lata\_edu satysf\_dem przyw\_kraj religijnosc przelozony region typ\_zakladu plec /

SELECTION=NONE

CORRB

COVB

INFLUENCE

LACKFIT

AGGREGATE SCALE=NONE

RSQUARE

CTABLE

LINK=LOGIT

CLPARM=BOTH

CLODDS=BOTH

ALPHA=**0.05**

;

OUTPUT OUT=WORK.TMP2TempTablePredictions

PREDPROBS=INDIVIDUAL;

**RUN**;

**QUIT**;

/\* -------------------------------------------------------------------

Koniec kodu zadania

------------------------------------------------------------------- \*/

**RUN**; **QUIT**;

%***\_eg\_conditional\_dropds***(WORK.SORTTempTableSorted,

WORK.TMP2TempTablePredictions);

TITLE; FOOTNOTE;

ODS GRAPHICS OFF;

/\*Odrzucenie zmiennej religijnosc\*/

/\* -------------------------------------------------------------------

Kod wygenerowany przez zadanie SAS-a

Wygenerowany dnia: środa, 8 maja 2019 o godz. 15:58:04

Przez zadanie: Regresja logistyczna

Dane wejściowe: Local:PROJEKT.SKATEGORYZOWANE

Serwer: Local

------------------------------------------------------------------- \*/

ODS GRAPHICS ON;

%***\_eg\_conditional\_dropds***(WORK.SORTTempTableSorted,

WORK.TMP2TempTablePredictions);

/\* -------------------------------------------------------------------

Sortowanie zbioru Local:PROJEKT.SKATEGORYZOWANE

------------------------------------------------------------------- \*/

**PROC** **SQL**;

CREATE VIEW WORK.SORTTempTableSorted AS

SELECT T.poziom\_szczescia, T.osob\_gosp, T.godz\_pracy, T.plec, T.wiek, T.zdrowie, T.lata\_edu, T.satysf\_dem, T.przyw\_kraj, T.religijnosc, T.doch\_gosp, T.typ\_zakladu, T.przelozony, T.region

FROM PROJEKT.SKATEGORYZOWANE as T

;

**QUIT**;

TITLE;

TITLE1 "Rezultaty regresji logistycznej";

FOOTNOTE;

FOOTNOTE1 "Wygenerowane przez System SAS (&\_SASSERVERNAME, &SYSSCPL) dnia %TRIM(%QSYSFUNC(DATE(), NLDATE20.)) o godz. %TRIM(%SYSFUNC(TIME(), NLTIMAP20.))";

**PROC** **LOGISTIC** DATA=WORK.SORTTempTableSorted

PLOTS(ONLY)=ALL

;

CLASS plec (PARAM=EFFECT) wiek (PARAM=EFFECT) zdrowie (PARAM=EFFECT) lata\_edu (PARAM=EFFECT) satysf\_dem (PARAM=EFFECT) przyw\_kraj (PARAM=EFFECT) religijnosc (PARAM=EFFECT) doch\_gosp (PARAM=EFFECT) typ\_zakladu (PARAM=EFFECT)

przelozony (PARAM=EFFECT) region (PARAM=EFFECT);

MODEL poziom\_szczescia (Event = '1')=osob\_gosp godz\_pracy wiek zdrowie lata\_edu satysf\_dem przyw\_kraj przelozony region typ\_zakladu plec doch\_gosp /

SELECTION=NONE

CORRB

COVB

INFLUENCE

LACKFIT

AGGREGATE SCALE=NONE

RSQUARE

CTABLE

LINK=LOGIT

CLPARM=BOTH

CLODDS=BOTH

ALPHA=**0.05**

;

OUTPUT OUT=WORK.TMP2TempTablePredictions

PREDPROBS=INDIVIDUAL;

**RUN**;

**QUIT**;

/\* -------------------------------------------------------------------

Koniec kodu zadania

------------------------------------------------------------------- \*/

**RUN**; **QUIT**;

%***\_eg\_conditional\_dropds***(WORK.SORTTempTableSorted,

WORK.TMP2TempTablePredictions);

TITLE; FOOTNOTE;

ODS GRAPHICS OFF;

/\*Odrzucenie zmiennej lata\_edu\*/

/\* -------------------------------------------------------------------

Kod wygenerowany przez zadanie SAS-a

Wygenerowany dnia: środa, 8 maja 2019 o godz. 15:59:53

Przez zadanie: Regresja logistyczna

Dane wejściowe: Local:PROJEKT.SKATEGORYZOWANE

Serwer: Local

------------------------------------------------------------------- \*/

ODS GRAPHICS ON;

%***\_eg\_conditional\_dropds***(WORK.SORTTempTableSorted,

WORK.TMP2TempTablePredictions);

/\* -------------------------------------------------------------------

Sortowanie zbioru Local:PROJEKT.SKATEGORYZOWANE

------------------------------------------------------------------- \*/

**PROC** **SQL**;

CREATE VIEW WORK.SORTTempTableSorted AS

SELECT T.poziom\_szczescia, T.osob\_gosp, T.godz\_pracy, T.plec, T.wiek, T.zdrowie, T.lata\_edu, T.satysf\_dem, T.przyw\_kraj, T.religijnosc, T.doch\_gosp, T.typ\_zakladu, T.przelozony, T.region

FROM PROJEKT.SKATEGORYZOWANE as T

;

**QUIT**;

TITLE;

TITLE1 "Rezultaty regresji logistycznej";

FOOTNOTE;

FOOTNOTE1 "Wygenerowane przez System SAS (&\_SASSERVERNAME, &SYSSCPL) dnia %TRIM(%QSYSFUNC(DATE(), NLDATE20.)) o godz. %TRIM(%SYSFUNC(TIME(), NLTIMAP20.))";

**PROC** **LOGISTIC** DATA=WORK.SORTTempTableSorted

PLOTS(ONLY)=ALL

;

CLASS plec (PARAM=EFFECT) wiek (PARAM=EFFECT) zdrowie (PARAM=EFFECT) lata\_edu (PARAM=EFFECT) satysf\_dem (PARAM=EFFECT) przyw\_kraj (PARAM=EFFECT) religijnosc (PARAM=EFFECT) doch\_gosp (PARAM=EFFECT) typ\_zakladu (PARAM=EFFECT)

przelozony (PARAM=EFFECT) region (PARAM=EFFECT);

MODEL poziom\_szczescia (Event = '1')=osob\_gosp godz\_pracy wiek zdrowie satysf\_dem przyw\_kraj przelozony region typ\_zakladu plec doch\_gosp religijnosc /

SELECTION=NONE

CORRB

COVB

INFLUENCE

LACKFIT

AGGREGATE SCALE=NONE

RSQUARE

CTABLE

LINK=LOGIT

CLPARM=BOTH

CLODDS=BOTH

ALPHA=**0.05**

;

OUTPUT OUT=WORK.TMP2TempTablePredictions

PREDPROBS=INDIVIDUAL;

**RUN**;

**QUIT**;

/\* -------------------------------------------------------------------

Koniec kodu zadania

------------------------------------------------------------------- \*/

**RUN**; **QUIT**;

%***\_eg\_conditional\_dropds***(WORK.SORTTempTableSorted,

WORK.TMP2TempTablePredictions);

TITLE; FOOTNOTE;

ODS GRAPHICS OFF;

/\*Selekcja forward\*/

/\* -------------------------------------------------------------------

Kod wygenerowany przez zadanie SAS-a

Wygenerowany dnia: niedziela, 5 maja 2019 o godz. 22:17:51

Przez zadanie: Regresja logistyczna

Dane wejściowe: Local:PROJEKT.SKATEGORYZOWANE

Serwer: Local

------------------------------------------------------------------- \*/

ODS GRAPHICS ON;

%***\_eg\_conditional\_dropds***(WORK.PREDLogRegPredictions,

WORK.SORTTempTableSorted);

/\* -------------------------------------------------------------------

Sortowanie zbioru Local:PROJEKT.SKATEGORYZOWANE

------------------------------------------------------------------- \*/

**PROC** **SQL**;

CREATE VIEW WORK.SORTTempTableSorted AS

SELECT \*

FROM PROJEKT.SKATEGORYZOWANE

;

**QUIT**;

TITLE;

TITLE1 "Rezultaty regresji logistycznej";

FOOTNOTE;

FOOTNOTE1 "Wygenerowane przez System SAS (&\_SASSERVERNAME, &SYSSCPL) dnia %TRIM(%QSYSFUNC(DATE(), NLDATE20.)) o godz. %TRIM(%SYSFUNC(TIME(), NLTIMAP20.))";

**PROC** **LOGISTIC** DATA=WORK.SORTTempTableSorted

PLOTS(ONLY)=ALL

;

CLASS plec (PARAM=EFFECT) wiek (PARAM=EFFECT) zdrowie (PARAM=EFFECT) lata\_edu (PARAM=EFFECT) satysf\_dem (PARAM=EFFECT) przyw\_kraj (PARAM=EFFECT) religijnosc (PARAM=EFFECT) doch\_gosp (PARAM=EFFECT) typ\_zakladu (PARAM=EFFECT)

przelozony (PARAM=EFFECT) region (PARAM=EFFECT);

MODEL poziom\_szczescia (Event = '1')=osob\_gosp godz\_pracy plec wiek zdrowie lata\_edu satysf\_dem przyw\_kraj religijnosc doch\_gosp typ\_zakladu przelozony region osob\_gosp\*godz\_pracy osob\_gosp\*plec osob\_gosp\*wiek osob\_gosp\*zdrowie osob\_gosp\*lata\_edu osob\_gosp\*satysf\_dem osob\_gosp\*przyw\_kraj osob\_gosp\*religijnosc osob\_gosp\*doch\_gosp osob\_gosp\*typ\_zakladu osob\_gosp\*przelozony osob\_gosp\*region godz\_pracy\*plec godz\_pracy\*wiek godz\_pracy\*zdrowie godz\_pracy\*lata\_edu godz\_pracy\*satysf\_dem godz\_pracy\*przyw\_kraj godz\_pracy\*religijnosc godz\_pracy\*doch\_gosp godz\_pracy\*typ\_zakladu godz\_pracy\*przelozony godz\_pracy\*region plec\*wiek plec\*zdrowie plec\*lata\_edu plec\*satysf\_dem plec\*przyw\_kraj plec\*religijnosc plec\*doch\_gosp plec\*typ\_zakladu plec\*przelozony plec\*region wiek\*zdrowie wiek\*lata\_edu wiek\*satysf\_dem wiek\*przyw\_kraj wiek\*religijnosc wiek\*doch\_gosp wiek\*typ\_zakladu wiek\*przelozony wiek\*region zdrowie\*lata\_edu zdrowie\*satysf\_dem zdrowie\*przyw\_kraj zdrowie\*religijnosc zdrowie\*doch\_gosp zdrowie\*typ\_zakladu zdrowie\*przelozony zdrowie\*region lata\_edu\*satysf\_dem lata\_edu\*przyw\_kraj lata\_edu\*religijnosc lata\_edu\*doch\_gosp lata\_edu\*typ\_zakladu lata\_edu\*przelozony lata\_edu\*region satysf\_dem\*przyw\_kraj satysf\_dem\*religijnosc satysf\_dem\*doch\_gosp satysf\_dem\*typ\_zakladu satysf\_dem\*przelozony satysf\_dem\*region przyw\_kraj\*religijnosc przyw\_kraj\*doch\_gosp przyw\_kraj\*typ\_zakladu przyw\_kraj\*przelozony przyw\_kraj\*region religijnosc\*doch\_gosp religijnosc\*typ\_zakladu religijnosc\*przelozony religijnosc\*region doch\_gosp\*typ\_zakladu doch\_gosp\*przelozony doch\_gosp\*region typ\_zakladu\*przelozony typ\_zakladu\*region przelozony\*region /

SELECTION=FORWARD

SLE=**0.05**

INCLUDE=**0**

CORRB

COVB

INFLUENCE

LACKFIT

AGGREGATE SCALE=NONE

RSQUARE

CTABLE

LINK=LOGIT

CLPARM=BOTH

CLODDS=BOTH

ALPHA=**0.05**

;

OUTPUT OUT=WORK.PREDLogRegPredictions(LABEL="Statystyki i prognozy regresji logistycznej dla PROJEKT.SKATEGORYZOWANE")

PREDPROBS=INDIVIDUAL;

**RUN**;

**QUIT**;

/\* -------------------------------------------------------------------

Koniec kodu zadania

------------------------------------------------------------------- \*/

**RUN**; **QUIT**;

%***\_eg\_conditional\_dropds***(WORK.SORTTempTableSorted);

TITLE; FOOTNOTE;

ODS GRAPHICS OFF;

/\*Selekcja stepwise\*/

/\* -------------------------------------------------------------------

Kod wygenerowany przez zadanie SAS-a

Wygenerowany dnia: poniedziałek, 6 maja 2019 o godz. 00:52:55

Przez zadanie: Regresja logistyczna (2)

Dane wejściowe: Local:PROJEKT.SKATEGORYZOWANE

Serwer: Local

------------------------------------------------------------------- \*/

ODS GRAPHICS ON;

%***\_eg\_conditional\_dropds***(WORK.PREDLOGREGPREDICTIONS\_0000,

WORK.SORTTempTableSorted);

/\* -------------------------------------------------------------------

Sortowanie zbioru Local:PROJEKT.SKATEGORYZOWANE

------------------------------------------------------------------- \*/

**PROC** **SQL**;

CREATE VIEW WORK.SORTTempTableSorted AS

SELECT \*

FROM PROJEKT.SKATEGORYZOWANE

;

**QUIT**;

TITLE;

TITLE1 "Rezultaty regresji logistycznej";

FOOTNOTE;

FOOTNOTE1 "Wygenerowane przez System SAS (&\_SASSERVERNAME, &SYSSCPL) dnia %TRIM(%QSYSFUNC(DATE(), NLDATE20.)) o godz. %TRIM(%SYSFUNC(TIME(), NLTIMAP20.))";

**PROC** **LOGISTIC** DATA=WORK.SORTTempTableSorted

PLOTS(ONLY)=ALL

;

CLASS plec (PARAM=EFFECT) wiek (PARAM=EFFECT) zdrowie (PARAM=EFFECT) lata\_edu (PARAM=EFFECT) satysf\_dem (PARAM=EFFECT) przyw\_kraj (PARAM=EFFECT) religijnosc (PARAM=EFFECT) doch\_gosp (PARAM=EFFECT) typ\_zakladu (PARAM=EFFECT)

przelozony (PARAM=EFFECT) region (PARAM=EFFECT);

MODEL poziom\_szczescia (Event = '1')=osob\_gosp godz\_pracy plec wiek zdrowie lata\_edu satysf\_dem przyw\_kraj religijnosc doch\_gosp typ\_zakladu przelozony region osob\_gosp\*godz\_pracy osob\_gosp\*plec osob\_gosp\*wiek osob\_gosp\*zdrowie osob\_gosp\*lata\_edu osob\_gosp\*satysf\_dem osob\_gosp\*przyw\_kraj osob\_gosp\*religijnosc osob\_gosp\*doch\_gosp osob\_gosp\*typ\_zakladu osob\_gosp\*przelozony osob\_gosp\*region godz\_pracy\*plec godz\_pracy\*wiek godz\_pracy\*zdrowie godz\_pracy\*lata\_edu godz\_pracy\*satysf\_dem godz\_pracy\*przyw\_kraj godz\_pracy\*religijnosc godz\_pracy\*doch\_gosp godz\_pracy\*typ\_zakladu godz\_pracy\*przelozony godz\_pracy\*region plec\*wiek plec\*zdrowie plec\*lata\_edu plec\*satysf\_dem plec\*przyw\_kraj plec\*religijnosc plec\*doch\_gosp plec\*typ\_zakladu plec\*przelozony plec\*region wiek\*zdrowie wiek\*lata\_edu wiek\*satysf\_dem wiek\*przyw\_kraj wiek\*religijnosc wiek\*doch\_gosp wiek\*typ\_zakladu wiek\*przelozony wiek\*region zdrowie\*lata\_edu zdrowie\*satysf\_dem zdrowie\*przyw\_kraj zdrowie\*religijnosc zdrowie\*doch\_gosp zdrowie\*typ\_zakladu zdrowie\*przelozony zdrowie\*region lata\_edu\*satysf\_dem lata\_edu\*przyw\_kraj lata\_edu\*religijnosc lata\_edu\*doch\_gosp lata\_edu\*typ\_zakladu lata\_edu\*przelozony lata\_edu\*region satysf\_dem\*przyw\_kraj satysf\_dem\*religijnosc satysf\_dem\*doch\_gosp satysf\_dem\*typ\_zakladu satysf\_dem\*przelozony satysf\_dem\*region przyw\_kraj\*religijnosc przyw\_kraj\*doch\_gosp przyw\_kraj\*typ\_zakladu przyw\_kraj\*przelozony przyw\_kraj\*region religijnosc\*doch\_gosp religijnosc\*typ\_zakladu religijnosc\*przelozony religijnosc\*region doch\_gosp\*typ\_zakladu doch\_gosp\*przelozony doch\_gosp\*region typ\_zakladu\*przelozony typ\_zakladu\*region przelozony\*region /

SELECTION=STEPWISE

SLE=**0.05**

SLS=**0.05**

INCLUDE=**0**

CORRB

COVB

INFLUENCE

LACKFIT

AGGREGATE SCALE=NONE

RSQUARE

CTABLE

LINK=LOGIT

CLPARM=BOTH

CLODDS=BOTH

ALPHA=**0.05**

;

OUTPUT OUT=WORK.PREDLOGREGPREDICTIONS\_0000(LABEL="Statystyki i prognozy regresji logistycznej dla PROJEKT.SKATEGORYZOWANE")

PREDPROBS=INDIVIDUAL;

**RUN**;

**QUIT**;

/\* -------------------------------------------------------------------

Koniec kodu zadania

------------------------------------------------------------------- \*/

**RUN**; **QUIT**;

%***\_eg\_conditional\_dropds***(WORK.SORTTempTableSorted);

TITLE; FOOTNOTE;

ODS GRAPHICS OFF;

1. T. Kanasz, “*Uwarunkowania szczęścia*”, Wydawnictwo Akademii Pedagogiki Specjalnej, Warszawa 2015, s. 66-67. [↑](#footnote-ref-1)
2. J. F. Helliwell, R. Layard, J. D. Sachs, “*World happiness report*”, 2019, s. 94, 101. [↑](#footnote-ref-2)
3. Centre for Bhutan Studies & GNH Research, *“A Compass Towards a Just and Harmonious Society”,* Thimpiu 2015, s. 37, 169, 209. [↑](#footnote-ref-3)
4. T. A. Wright, K. K. Benett, T. Dun, “*Life and Job Satisfaction*”, Nevada 1989. [↑](#footnote-ref-4)
5. J. F. Terelak, A. Borkowska, *“Satysfakcja z pracy jako źródło dobrostanu psychicznego u polskich emigrantów konsumpcyjnych w Kanadzie”*, Warszawa 2007. [↑](#footnote-ref-5)
6. J. Hartog, H. Oosterbeek, *“Health, Wealth and Happiness: Why Pursue a Higher Education?”*, Amsterdam 1989. [↑](#footnote-ref-6)
7. https://pdf.helion.pl/khygge/khygge.pdf (data odsłony: 03.05.2019 r.) [↑](#footnote-ref-7)
8. https://www.norwegofil.pl/ludzie-i-tradycje/hygge-i-kos-norweska-metoda-na-szczescie   
   (data odsłony: 03.05.2019 r.) [↑](#footnote-ref-8)
9. T. Kanasz „*Uwarunkowania szczęścia*”, Warszawa 2015, s.9 [↑](#footnote-ref-9)
10. [S. Doughty.](https://www.dailymail.co.uk/home/search.html?s=&authornamef=Steve+Doughty+Social+Affairs+Correspondent+For+The+Daily+Mail)  „*A big family 'is the key to a happy life: Two thirds who live in household with three or more children consider themselves to be satisfied with their lives*“. [↑](#footnote-ref-10)
11. https://www.dailymail.co.uk/news/article-3106612/A-big-family-key-happy-life-Two-thirds-live-household-three-children-consider-satisfied-lives.html (data odsłony: 01.05.2019 r.) [↑](#footnote-ref-11)
12. https://medium.com/@andrewmerle/this-is-how-many-hours-you-should-really-be-working-ff1e8a6ad958 (data odsłony: 5.05.2019 r.) [↑](#footnote-ref-12)
13. C. Batz, L. Tay, “*Gender Differences in Subjective Well-Being*”, 2017. [↑](#footnote-ref-13)
14. B. Stevenson, J. Wolfers, “The Paradox of Declining Female Happiness”, 2009. [↑](#footnote-ref-14)
15. D. G. Blanchflower, A. J. Oswald, “*Well-being over time in Britain and the USA”*, Journal of Public Economics, 2004, s. 88, 1359-1386. [↑](#footnote-ref-15)
16. D. G. Blanchflower, A. J. Oswald, “*Is well-being U-shaped over the life cycle?”*, Social Science and Medicine, 2008, s. 66, 1733-1749. [↑](#footnote-ref-16)
17. A. Strużyńska- Kułajowicz „*Czy władza szczęście daje? Władza a nastrój, emocje i satysfakcja z życia*”, Wyższa Szkoła Psychologii Społecznej w Sopocie, Sopot 2010, s. 126, 132. [↑](#footnote-ref-17)
18. A. Krause, “*Happiness and Work”*, IZA, 2014. [↑](#footnote-ref-18)
19. J. Hartoga, H. Oosterbeek, „*Health, wealth, and happiness: why pursue a higher education?*”, Economics of Education Review, Vol. 17, No. 3, 1998, s. 11. [↑](#footnote-ref-19)
20. T. Kanasz „*Uwarunkowania szczęścia*”, Warszawa 2015, s. 70. [↑](#footnote-ref-20)
21. T. Kanasz „*Uwarunkowania szczęścia*”, Warszawa 2015, s. 67, 112, 142. [↑](#footnote-ref-21)
22. https://kirken.no/nb-NO/church-of-norway/about/basics-and-statistics/ (data odsłony: 04.05.2019 r.). [↑](#footnote-ref-22)
23. <http://worldpopulationreview.com/countries/norway-population/>, (data odsłony: 01.05.2019 r.). [↑](#footnote-ref-23)
24. https://www.theatlantic.com/health/archive/2017/03/seasonal-affective-disorder-mosaic/519495/  
       (data odsłony: 01.05.2019 r.). [↑](#footnote-ref-24)
25. J. F. Helliwell, R. Layard, J. D. Sachs, “*World Happiness Report*”, 2019. [↑](#footnote-ref-25)
26. T. Kanasz „*Uwarunkowania szczęścia*”, Warszawa 2015, s. 24. [↑](#footnote-ref-26)
27. J. F. Helliwell, R. Layard, J. D. Sachs „*World Happiness Report 2019*”, s. 42. [↑](#footnote-ref-27)
28. H. Tajfel, „*Human groups and social categories”*, Cambridge University Press, 1981.

    M. Morrison, L. Tay, E. Diener, „*Subjective well-being and national satisfaction: Findings from a worldwide survey”*, Psychological Science, 2011, s. 166-171. [↑](#footnote-ref-28)
29. T, Kanasz, „*Uwarunkowania szczęścia*”, Warszawa 2015, s. 72. [↑](#footnote-ref-29)
30. J. Hartoga, H. Oosterbeek, „*Health, wealth, and happiness: why pursue a higher education?*”, Economics of Education Review, Vol. 17, No. 3, 1998, s. 4. [↑](#footnote-ref-30)
31. http://support.sas.com/kb/22/599.html (data odsłony: 05.05.2019 r.) [↑](#footnote-ref-31)