

Regresja logistyczna z wykorzystaniem narzędzi SAS - kraj Niemcy

Autorzy

Bartosz Ciołek

Małgorzata Smolińska

Marcin Stołowski

Michał Foryt

Andrzej Borowiec

Mateusz Mozgawa

Krystian Iciek

Michalina Pawłowska

Maj 2019

[**1. Wstęp**](#_vhuzzip5uma7) **4**

[**2. Metodologia**](#_52pvpax6kkd1) **6**

[**3. Dane**](#_bg8uxoic98kp) **6**

[**4. Opis narzędzi oraz statystyk**](#_c9k8vs9xmxsw) **8**

[**5. Hipotezy badawcze**](#_qzijxcc3chwa) **11**

[**6. Korelaty demograficzne i psychologiczne - objaśnienie**](#_qqvhs7bvpl7l) **12**

[**7. Opis wybranych zmiennych**](#_g63l1lwcwitc) **17**

[**8. Kategoryzacja zmiennych**](#_12pzsgi6qt3p) **55**

[**9. Analiza**](#_2s8302iheool) **71**

[9.1 Światopogląd](#_55lfmlb73o2r) 71

[9.2 Środowisko](#_5mwemxr35uv2) 75

[9.3 Porównanie obu modeli](#_7j0epzs99bxl) 85

[**10. Analiza ostatecznego modelu regresji**](#_fnec3byercg4) **87**

[10.1 Wynik testu zgodności Hosmera i Lemeshowa](#_2hvq5nyqe3zt) 87

[10.2 Statystyka dobroci dopasowania dewiancji i Pearsona](#_6ie4nxc4xu9b) 88

[10.3 Interpretacja ilorazów szans i weryfikacja postawionych hipotez](#_cwbogr4nwyl9) 88

[10.4 Wyniki a postawione hipotezy badawcze](#_ci6ovln36teu) 91

[10.5 Krzywa ROC](#_kty44ncrrh5o) 92

[10.6 Badanie wpływu](#_k00lhv7f2xae) 93

[**11. Podsumowanie**](#_t9l31xsxkobd) **94**

[**12. Bibliografia**](#_geyoxyu7l0gh) **95**

[**13. Kody SAS**](#_1zy97zroqhm4) **96**

# 

# 1. Wstęp

“Małżeństwo zwykło być początkiem związku. W dzisiejszych czasach jest to jego finał.” – twierdzi Helen Fisher, światowej sławy profesor antropologii i ekspert od miłości romantycznej i od ponad dziesięciu lat studiującej ewolucję ludzkich emocji.

Celem projektu jest analiza dotycząca identyfikacji czynników narażających pary małżeńskie na rozwód poprzez rozpoznanie grup osób podlegających podwyższonymu ryzyku rozpadu małżeństwa. Realizując tak sformułowany cel, wykorzystano metody regresji logistycznej, analizy i syntezy połączonej   
z logicznym wnioskowaniem.

Rozwód jest działaniem prawnym między małżonkami w celu rozwiązania związku małżeńskiego. Można to określić jako rozwiązanie małżeństwa i zasadniczo jest to postępowanie prawne, które kończy małżeństwo przed śmiercią któregokolwiek z małżonków. Z jednej strony rozwód jest bardzo subiektywnym przeżyciem dla jednostek, aczkolwiek można na niego również spojrzeć od strony zmian demograficznych oraz trendów społecznych. Trend rozwodów ma dalszy wpływ na współczynnik dzietności oraz dla dobrostanu dzieci pochodzących z takiego związku oraz dobrostanu partnerów tworzących związek, dlatego też jest coraz ważniejszym przedmiotem zainteresowania dla demografów.

Według raportów Eurostat, statystyki dotyczące małżeństwa i rozwodu bazowane na najnowszych danych z czerwca 2018, dostępnych dla wszystkich państw członkowskich UE w 2015 r. W UE-28 miało miejsce około 2,2 mln małżeństw i 946 tys. Rozwodów. Liczby te mogą być wyrażone jako 4,3 małżeństwa na każde 1000 osób (innymi słowy surowy współczynnik małżeństwa) i 1,9 rozwodu na każde 1000 osób (innymi słowy surowy współczynnik rozwodów).

Od 1965 r. Surowy wskaźnik małżeństwa w UE-28 spadł o blisko 50% w wartościach względnych (z 7,8 na 1 000 osób w 1965 r. Do 4,3 w 2015 r.). W tym samym czasie wskaźnik rozwodów wzrósł ponad dwukrotnie, z 0,8 na 1000 osób w 1965 r. Do 1,9 w 2015 r. Część tego wzrostu może wynikać z faktu, że w kilku państwach członkowskich UE rozwód został zalegalizowany w tym okresie (na przykład we Włoszech, Hiszpanii, Irlandii i na Malcie).

Istnieje wiele różnych sposobów określania rozpadu małżeństwa w literaturze przedmiotu. Współautorzy Kitson, Babri i Roach wyodrębniają trzy wymiary przeprowadzenia badań nad rozwodami [1]:

a) poziom makrospołeczny - poszukiwanie zjawisk i procesów społecznych, głównie w obszarze zmian strukturalnych (industrializacja, urbanizacja); przykładowym procesem jest różnorodność zmian   
w rodzinie, osłabiająca jej stabilność;

b) korelaty demograficzne i psychologiczne - poszukiwanie danych demograficznych   
i psychologicznych, które osłabiają stabilność małżeństwa; przykład takich czynników to ciążą przedmałżeńską, zbyt młodym wiekiem małżeństwa, rozwodem w rodzinie, niskim statusem społeczno-ekonomicznym itp.

c) przyczyny rozpadu związku małżeńskiego z perspektywy osób rozwiedzionych, takie jak alkoholizm, zdrada, brak zgodności charakterów itp.

Śledząc literaturę polską w zakresie tychże badań reprezentuje polski demograf Edward Rosset [2]. Pośród czynników determinujących rozwody wyróżnia on zasadniczo zgodne z propozycjami uprzednio wymienionych amerykańskich autorów, czyli przesłanki rozwodu (prognozy, czynniki zwiększające prawdopodobieństwo rozwodu, np. małżeństwa wymuszone, zbyt młody wiek zawarcia związku), czynniki rozwodu (industrializacja, urbanizacja, aktywizacja zawodowa kobiet, tolerancja opinii publicznej itp.) przyczyny rozwodu (alkoholizm, brak wierności itp.).

Podczas gdy byli małżonkowie podają za przyczyny rozpadu związku bezpośrednie winy partnera lub partnerki, w niniejszej pracy zostaną rozważone strukturalno-demograficznych korelaty rozwodu, ponieważ decyzje są uwarunkowane nie tylko przez czynniki indywidualne, takie jak wyznawane wartości rodzinne czy emocjonalne, ale także należy zwrócić uwagę na determinanty ekonomiczne, takie jak zarobki, czy też etniczne wprowadzając mniej lub bardziej trudne pod względem oceny społecznej (tolerancja), podjęcie tematu rozwiązania przysięgi małżeńskiej.

Występowanie rozwodów wpływa na cały system socjo-demograficzny, dlatego warto zwrócić uwagę na jego skutki.

Szybkim zmianom instytucjonalnym towarzyszą zmiany behawioralne, również w obszarach życia rodzinnego. W rezultacie zmiany systemu stymulującego wczesne małżeństwo i macierzyństwo na taki, który zachęca do opóźniania decyzji dotyczących założenia rodziny, zmieniając się zachowania demograficzne [3]. W Polsce po latach względnie stabilnych wzorców powstawania rodziny i stopniowego spadku płodności teoretyczny współczynnik płodności zmniejszył się z 2,05 dziecka na kobietę w 1989 r. do 1,23 w 2003 r. (I pozostaje niski - w 2009 r. Wyniósł 1,40), małżeństwo było opóźnione i związek stał się mniej trwały [4]. Osłabieniu tendencji do zawierania małżeństw korelowało rosnące występowanie zjawiska konkubinatu [5]. Malejące znaczenie małżeństwa jest również widoczne w rosnącej liczbie rozwodów po 2000 r. Ogólny przekrojowy wskaźnik rozwodów podwoił się z 0,16 w 1999 r. Do 0,30   
w 2006 r. I 0,28 w 2008 r.

Opisywane zmiany społeczne prowadzą nawet do zmian z tradycji poprzez zmiany instytucjonalne tak, że obecnie zmienia się wręcz formalne związki, np. rejestrowane związki partnerskie. Okazuje się, że nie wpływa to liniowo na dzietność. Terytoria, w których małżeństwo nie jest priorytetem i gdzie odnotowywana jest wysoka liczba rozwodów, mają na ogół wyższą dzietność. Natomiast kraje, w których dominuje tradycyjne podejście do rodziny, bywa iż mają kłopoty z demografią w państwie. Najwięcej dzieci rodzi się w liberalnej Francji oraz w Szwecji. Takie rozumowanie jest proponowane w opracowaniu „Uwarunkowania zachowań prokreacyjnych” prof. Ireny Kotowskiej z SGH [6].

# 2. Metodologia

Dalsze opracowanie składa się z opisu zastosowanych statystyk oraz narzędzi, następnie postawienia hipotezy oraz wyjaśnienia zmiennych objaśniających, następnie wprowadzenia tabel kontyngencji, przeprowadzenie analizy statystycznej z wykorzystaniem narzędzi SAS oraz analizy końcowej   
z wyciągnięciem wniosków.

W części opisowej dla zmiennych korelatów rozwodów społeczno-demograficzne rozpatrywanych w literaturze zostały wyselekcjonowane 24 zmiennych, na podstawie rozważań literatury oraz własnej intuicji. Kolejno zmienne będą śledzone i na ich podstawie modele będą dopasowywane. Na tej podstawie zostało sformułowana hipoteza dotycząca czynników wpływających na rozpad związku małżeńskiego.

W kolejnej części nastąpi prezentacja danych oraz dopracowanie modelu wykorzystany w analizach empirycznych. Wyniki estymacji zawarto w części piątej.

Zakończenie zawiera analizę otrzymanych wyników oraz postawienie wniosków z nich wynikających.

# 3. Dane

Projekt ten został oparty na danych Europejskiego Sondażu Społecznego z rundy 8 z 2016 roku.

Dane wykorzystane w analizie to badania pochodzące z Europejskiego Sondażu Społecznego (ang. European Social Survey – ESS). Ankietowani odpowiadają na pytania dotyczące ich zachowania, przekonań, postaw społecznych.

Zestaw danych ESS opiera się na badaniach bezpośrednich (PAPI) oraz wywiadach komputerowych (CAPI). Próbka ankietowanych obejmuje ponad 30 krajów, a wymaganiem jest wiek powyżej 15 lat. Warto wskazać, że badanie jest prowadzone na jednostkach, a nie gospodarstwach domowych.

Dane zostały załączone wraz z projektem, lecz są one także dostępne pod adresem: https://www.europeansocialsurvey.org/download.html?file=ESS8DE&c=DE&y=2016

# 4. Opis narzędzi oraz statystyk

Regresja logistyczna jest jedną z odmian analizy regresji. Jej cechą główną jest zjawisko, iż zmienna objaśniana jest zmienną dychotomiczną. Oznacza to, że w sposób naturalny lub w ramach klasteryzacji zmienna ta przyjmuje dwie wartości najczęściej jest to 0 i 1. Zwykle wartości jakie przyjmuje zmienna objaśniana wskazuje na wystąpienie danego zjawiska, lub jego brak. Regresja logistyczna zatem pozwala na obliczenie prawdopodobieństwa sukcesu tego zdarzenia.

Przeprowadzone badanie składa się z czterech podstawowych etapów:

- zebranie i przygotowanie danych

- zredukowanie liczby analizowanych zmiennych

- wybór modelu

- zweryfikowanie z założeniami;

- stwierdzenie ważności modelu.

Badanie powstało przy wykorzystaniu programu SAS. Wykorzystane w projekcie dane zostały poddane procedurom takim jak FREQ, CORR. W celu budowy modelu regresji logistycznej zastosowana została procedura PROC LOGISTIC.

Do pełnej analizy najważniejszych elementów wykorzystano między innymi statystyki dopasowania dewiancji i Pearsona, statystyki dopasowania, test dobroci dopasowania Hosmera i Lemeshowa.

Statystyki dewiancji i dobroci dopasowania Pearsona umożliwiają poznanie odpowiedzi na pytanie postawione w hipotezie zerowej zgodnie z którą model jest dobrze dopasowany do danych. Metod oszacowania dobroci modelu jest wiele. Przedmiotem weryfikacji jest sprawdzenie, czy model jest dobrze dopasowany do danych (co wiąże się z częstotliwością „pomyłek” modelu), czy model ma dużą pojemność informacyjną i czy dobrze opisuje dane zjawisko.

Ważnym elementem badania jest weryfikacja istotności statystycznej poszczególnych parametrów. W tym celu zastosowany został test Walda. Do głównych zastosowań tego testu należy między innymi badanie istotności modelu, które polega na testowaniu wszystkich parametrów oprócz stałych, badanie łącznej istotności zmiennych oraz badanie istotności pojedynczego parametru. Wynik statystyki pozwala odpowiedzieć na pytanie czy warto jest oszacować model z wybranymi zmiennymi objaśniającymi. Nieistotność jakiejś zmiennej sugeruje, że zmienna nie wnosi dużo do analizy.

Dla prawidłowego badania zmiennych konieczna jest selekcja zmiennych. Zabieg ten pozwoli na wybór najbardziej właściwego podzbioru zmiennych. W tym celu wykorzystywana jest zazwyczaj automatyczna selekcja zmiennych, która dostępna jest w pakietach statystycznych. Selekcja ta pozwoli na ograniczenie licznego zbioru przypuszczalnych zmiennych objaśniających, eliminacje zmiennych, które tylko pozornie wpływają na zmienną zależną lub też eliminację współliniowości zmiennych niezależnych.

Wybór metody pozwala określić to w jaki sposób zmienne niezależne będą wprowadzane do analizy. Przy wykorzystaniu różnych metod możliwe jest skonstruowanie wielu modeli regresji dla jednego zbioru zmiennych.

Przy wykorzystaniu metody krokowej na każdym kroku analizy do modelu dołączona jest zmienna niezależna, która nie jest wykorzystana jeszcze w równaniu. Zmienna ta charakteryzuje się najmniejszym prawdopodobieństwem. W następnym kroku zmienne uwzględnione już w równaniu regresji są usuwane jeżeli związane z nimi prawdopodobieństwo staję się dostatecznie duże. Kiedy nie się wykluczyć ani dołączyć żadnych zmiennych procedura kończy się. Metoda selekcji krokowej – stepwise. Jest przykładem jednej z zautomatyzowanych metod selekcji. Drugą metodą jest metoda typu w przód – forward. Pierwsza z nich stanowi ulepszoną wersję drugiej ponieważ pozwala na pozbycie się z modelu zmiennej dodanej w poprzedniej iteracji. Zatem na każdym kroku testowane są zarówno przydatności zmiennych jak i istotności tych zmiennych, które jeszcze nie wyłączono z modelu

Selekcja postępująca ( w przód,forward) jest metodą, w której dobierane zmienne są kolejno wprowadzane do modelu. W pierwszej kolejności pod uwagę brana jest ta zmienna, która jest najsilniej skorelowana ze zmienną zależną. Zostaje ona wprowadzona do modelu tylko wtedy, gdy spełnia kryteria wprowadzenia. Gdy pierwsze zmienna zostanie wprowadzona do modelu następnie rozważana jest ta zmienna, która nie jest jeszcze wprowadzona do modelu a ma kolejno największą wartość współczynnika korelacji cząstkowej ze zmienną zależna. Cała procedura kończy się nie ma już zmiennych, które spełniają kryteria wprowadzanie do modelu.

Eliminacja poprzednich ( wsteczna), jest procedurą, w której następuje dobór zmiennych, wprowadzenie ich do równania a następnie usuwanie ich. Zmienne, które są w najmniejszej korelacji cząstkowej ze zmienną zależna są w pierwszej kolejności rozpatrzone pod względem ich usunięcia. Jeśli zmienne te spełniania określone kryteria zostają usunięte. Po usunięciu pierwszej zmiennej, kolejną braną pod uwagę do usunięcia jest ta zmienna pozostająca w równaniu, która ma najmniejszą korelacją cząstkową ze zmienną zależna. Procedura ulega zakończeniu gdy w równaniu nie występują inne zmienne spełniająca kryteria usunięcia. Ważnym elementem przy wykorzystywaniu metody eliminacji poprzednich jest by pamiętać o potencjalnych efektach zakłócających oraz interakcjach. Za zmienną zakłócającą uznać można taką, której ewentualne usunięcie skutkowałoby znaczną zmianą oszacowań parametrów dla pozostałych zmiennych. Model statystyczny powinien uwzględniać efekty główny i ich działania. Dzięki temu można badać efekt interakcji kilku zmiennych oraz to czy dana zmienna, która nie jest istotna statystycznie nie wchodzi przypadkiem w istotną statystycznie interakcję.

Niezależnie od wykorzystywanej metody aby zmiany zostały wprowadzone do modelu muszą spełniać pewne kryteria tolerancji. Zazwyczaj domyślnym poziomem tolerancji jest wartość 0,0001. W przypadku, gdy zmienna spowodowałaby spadek tolerancji innej wprowadzonej już do modelu zmiennej nie może ona być wówczas wprowadzona do modelu.

Na podstawie testu dobroci dopasowania Hosmera i Lemeshowa można dokonać analizy czy dane są istotne statystycznie czy też nie.

Hosmer i Lemeshow zaproponowali procedurę, która na podstawie wartości testu chi szacuje model pod względem dopasowania i dobroci. Algorytm obliczeniowy zakłada podział obserwacji z próby na podgrupy. Podgrupy te różnią się wyestymowanym prawdopodobieństwem uzyskania wartości kryterialnej zmiennej czyli takiej, która przyjmuje przez nią wartość 1. W kolejnym etapie obliczany jest współczynnik chi.

Test Hosmera i Lemeshowa porównuje wartości oczekiwane na podstawie modelu z wartościami zaobserwowanymi. Test Hosmera i Lemeshowa porównuje wartości oczekiwane na podstawie modelu z wartościami zaobserwowanymi. Test weryfikuje hipotezę o równości wartości obserwowanych i przewidywanych. Jeśli wartości obserwowane i przewidywane są wystarczająco bliskie, wówczas można założyć, model jest dobrze dopasowany do danych. Pożądany jest brak istotności tego testu. Test Hosmera i Lemeshowa jest główną statystyką diagnostyczną modelu regresji logistycznej. Test jest niezbędną statystyką podczas oceny dopasowania modelu do przewidywanych danych. Na podstawie testu Hosmera Lameshowa rozpatrywać można weryfikację modelu, jego przekształcenie lub zastosowanie alternatywnej metody analizy.

Test umożliwia porównanie dwóch modeli. Często wykorzystywany jest do porównania dwóch modeli, gdzie jeden z nich pozbawiony jest wartości nieistotnych statystycznie. Wynik pozwala ocenić rozkład prawdopodobieństw i czy modele są dobrze dopasowane. Dzięki temu można sprawdzić czy model zredukowany (czyli taki, który nie posiada zmiennych nieistotnych statystycznie) znacząco różni się od modelu pełnego.

Obliczeń w ramach projektu dokonano w systemie SAS. W celu budowy modelu regresji logistycznej wykorzystano PROC LOGISTIC. Procedura ta służy do analizy modeli z jakościową zmienną objaśnianą. Procedura LOGISTIC wykorzystywana jest do szacowania modeli regresji logistycznych binarnych, porządkowych oraz wielomianowych. Procedura najczęściej korzysta z metody największej wiarygodności. PROC LOGISTIC oprócz standardowych funkcjonalności szacuje również ilorazy szans oraz udostępnia szeroki wachlarz miar jakości klasyfikacji. PROCEDURA FREQ wykorzystana została do analizy jednowymiarowej i analizy zależności. PROCEDURA CORR pozwala na wyznaczenia współczynnika korelacji.

# 5. Hipotezy badawcze

Celem naszej analizy jest znalezienie odpowiedzi, jakie czynniki determinują, czy para zdecyduje się na rozstanie czy też nie. Ogólnie badane zmienne można podzielić na takie, które dotyczą przeszłości (pochodzenia) respondenta, jak i teraźniejszości, a więc czynników na które badana osoba ma wpływ - jak choćby poglądy, sytuacja materialna albo liczba posiadanych dzieci. W trakcie dyskusji wyklarowały się dwie hipotezy badawcze.

1. **Czy osoby pochodzące z obszaru dawnej Republiki Federalnej Niemiec (RFN) rozwodzą się częściej, niż ludzie zamieszkujący dawną Niemiecką Republikę Demokratyczną (NRD)?**
2. **Czy osoby zadowolone z życia rozwodzą się rzadziej, niż niezadowolone?**

Jak wiadomo, po II Wojnie Światowej Niemcy zostały podzielone na dwie części. Wschód Niemiec znalazł się pod radziecką strefą wpływów jako Niemiecka Republika Demokratyczna. Zachodnią część kontrolowali Alianci - Francja, Wielka Brytania i Stany Zjednoczone. Dopiero w 1990 roku nastąpiło ponowne zjednoczenie Niemiec. W praktyce oznacza to, że dwa pokolenia Niemców żyło w dwóch różnych rzeczywistościach, by następnie znów się połączyć w jeden naród. W naszym badaniu będziemy chcieli sprawdzić, na ile osoby żyjące po stronie “zgniłego zachodu” - jak mówił Gomułka (były I sekretarz KC PPR), chętniej decydują się na rozwód lub rozwiązanie związku partnerskiego, od obywateli terenów pod kontrolą ZSRR.

Druga hipoteza dotyczy zadowolenia z życia. Choć przez wiele lat małżeństwo było związkiem, który wynikał raczej z obopólnych korzyści, niż z romantycznych pobudek, to dzisiaj jest ono wynikiem   
(a przynajmniej powinno być) głębokiej miłości między ukochanymi. W przypadku Niemiec, dotyczy to także par o tej samej płci (30 czerwca 2017 roku Niemcy zalegalizowały małżeństwa homoseksualne). Jak wynika jednak z badań [7], okres zakochania trwa przez pierwsze około 2 lata, dopiero później przeradza się ona w prawdziwą miłość, nieraz do końca życia. Niestety, tylko w 2017 w Niemczech doszło do 153 tysięcy rozwodów [8]. Decyzja o rozstaniu zawsze jest niezwykle trudna, jednak   
w niektórych sytuacjach może ona wyjść obu partnerom na dobre. W naszym badaniu chcielibyśmy sprawdzić, czy osoby zadowolone z życia rozwodzą się rzadziej, niż niezadowolone.

# 6. Korelaty demograficzne i psychologiczne - objaśnienie

Opis badanych zmiennych

1. Zmienna objaśniana - DVRCDEVA
2. Zmienne objaśniające

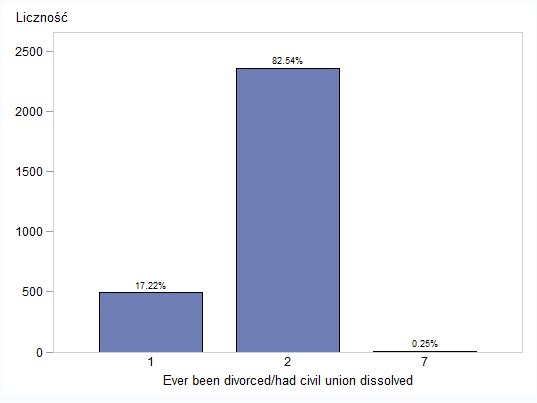
Do realizacji projektu zostało wybranych 25 zmiennych objaśniających. Poniższa tabela prezentuje szczegółowy opis zmiennych

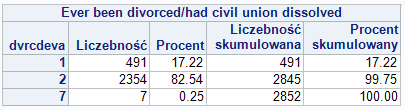
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lp. | Nazwa zmiennej | Typ zmiennej | Opis zmiennej | Wartości zmiennej |
| 1. | agea | Zmienna ciągła | Wiek respondenta | Wartość z zakresu 15-94 lata |
| 2 | anctry1 | Zmienna porządkowa | Pochodzenie | Wartości na podstawie European Standard Classification of Cultural and Ethnic Groups |
| 3 | hhmmb | Zmienna porządkowa | Liczba mieszkańców gospodarstwa domowego włączając badanego | Wartości z od 1 do 10 |
| 4 | rlgatnd | Zmienna porządkowa | Częstość uczestniczenia w praktykach religijnych z pominięciem specjalnych okazji | 1 - każdego dnia  2 - więcej niż raz dziennie  3 - raz na tydzień  4 - co najmniej raz w miesiącu  5 - podczas świąt  6 - rzadziej  7 - nigdy  77 - odmowa  88 - nie wiem  99 - brak odpowiedzi |
| 5 | sclmeet | Zmienna porządkowa | Częstość kontaktów społecznych z przyjaciółmi, rodziną lub znajomymi z pracy | 1 - nigdy  2 - rzadziej niż raz w miesiącu  3 - raz w miesiącu  4 - kilka razy w miesiącu  5 - raz w tygodniu  6 - kilka razy w tygodniu  7 - każdego dnia  77 - odmowa  88 - nie wiem  99 - brak odpowiedzi |
| 6 | rlgdgr | Zmienna porządkowa | Opinia na temat własnej religijności | Wartości z zakresu od 0 - niereligijny do 10 - bardzo religijny  77 - odmowa  88 - nie wiem  99 - brak odpowiedzi |
| 7 | nwspol | Zmienna ciągła | Czas dziennie spędzany na zapoznawaniu się z wiadomościami z polityki w minutach | Wartość z zakresu |
| 8 | eduyrs | Zmienna ciągła | Czas edukacji respondenta w latach | Wartość z zakresu |
| 9 | hlthhmp | Zmienna porządkowa | Utrudnienie w życiu codziennymi wynikające z choroby, niepełnosprawności, choroby umysłowej etc | 1 - duże  2 - do pewnego stopnia  3 - brak  77 - odmowa  88 - nie wiem  99 - brak odpowiedzi |
| 10 | stflife | Zmienna porządkowa | Satysfakcja z życia | Wartości z zakresu 0 - bardzo niezadowolony do 10 - bardzo zadowolony  77 - odmowa  88 - nie wiem  99 - brak odpowiedzi |
| 11 | edctn | Zmienna binarna | Czy przez ostatnie 7 dni responded zajmował się edukacją | 0 - nie  1 - tak |
| 12 | jbspv | Zmienna porządkowa | Pozycja w pracy wymagająca zarządzania innymi pracownikami | 1 - tak  2 - nie  3 - n/a  7 - odmowa  8 - nie wie  9 - brak odpowiedzi |
| 13 | uemp3m | Zmienna porządkowa | Okres przebywania na bezrobociu lub poszukiwania pracy równy lub dłuższy niż 3 miesięcy | 1 - tak  2 - nie  3 - n/a  7 - odmowa  8 - nie wie  9 - brak odpowiedzi |
| 14 | Chldhhe | Zmienna porządkowa | Obecność dziecka w gospodarstwie domowym | 1 - tak  2 - nie  3 - n/a  7 - odmowa  8 - nie wie  9 - brak odpowiedzi |
| 15 | Emplrel | Zmienna porządkowa | Rodzaj stosunku pracy | 1 - zatrudniony  2 - samozatrudniony  3 - firma rodzinna  6 - n/a  7 - odmowa  8 - nie wie  9 - brak odpowiedzi |
| 16 | Bennent | Zmienna porządkowa | Opinia na temat stwierdzenie, że wiele osób niezasłużenie otrzymuje pomoc od państwa | 1 - całkowicie się zgadzam  2 - zgadzam się  3 - ani się zgadzam ani nie zgadzam  4 - nie zgadzam się  5 - całkowicie się nie zgadzam  7 - odmowa  8 - nie wie  9 - brak odpowiedzi |
| 17 | icwhct | Zmienna porządkowa | Posiadanie podstawowej lub stałej liczby godzin | 1 - tak  2 - nie  6 - n/a  9 - brak odpowiedzi |
| 18 | imptrad | Zmienna porządkowa | Subiektywna ocena respondenta podobieństwa do osoby, dla której tradycje i wartości mają dużą wagę | 1 - bardzo podobna  2 - podobna  3 - w pewnym stopniu podobna  4 - trochę podobna  5 - niepodobna  6 - bardzo niepodobna  7 - odmowa  8 - nie wie  9 - brak odpowiedzi |
| 19 | pplfair | Zmienna porządkowa | Ocena zaufania do innych ludzi | Wartości z zakresu 0 - większość osób chce mnie wykorzystać do 10 - większość osób zachowuje się sprawiedliwie |
| 20 | Imprich | Zmienna porządkowa | Jak ważne dla respondenta jest bogactwo, posiadanie pieniędzy i drogich rzeczy | 1 - całkowicie się zgadzam  2 - zgadzam się  3 - średnio się zgadzam  4 - trochę się zgadzam  5 - nie zgadzam się  6 - całkowicie się nie zgadzam  7 - odmowa  8 - nie wie  9 - brak odpowiedzi |
| 21 | Impfree | Zmienna porządkowa | Jak ważna dla respondenta jest niezależność i wolność w podejmowaniu decyzji | 1 - bardzo potrzebna  2 - potrzebna  3 - w pewnym stopniu potrzebna  4 - trochę potrzebna  5 - niepotrzebna  6 - bardzo niepotrzebna  7 - odmowa  8 - nie wie  9 - brak odpowiedzi |
| 22 | hincfel | Zmienna porządkowa | Opinia dotycząca poczucia komfortu / stabilności, w związku z obecnym stanem budżetu domowego. | 1 - Komfortowe życie przy obecnych dochodach  2 - Radzę sobie przy obecnych dochodach  3 - Trudne przy obecnych dochodach  4 - Bardzo trudne przy obecnym poziomie dochodów  7 - Odmowa odpowiedzi  8 - Nie wiem  9 - Brak odpowiedzi |
| 23 | Region | Zmienna porządkowa | Region zamieszkania zadeklarowany przez respondenta | DE1 - Baden-Württember  DE2 - Bayern  DE3 - Berlin  DE4 - Brandenburg  DE5 - Bremen  DE6 - Hamburg  DE7 - Hessen  DE8 - Mecklenburg - Vorpommern  DE9 - Niedersachsen  DEA - Nordrhein-Westfalen  DEB - Rheinland-Pfalz  DEC - Saarland  DED - Sachsen  DEE - Sachsen-Anhalt  DEF - Schleswig-Holstein  DEG - Thüringen |
| 24 | Lnghom1 | Zmienna porządkowa | Język najczęściej używany w domu: pierwszy wymieniony | Zakodowane przy pomocy standardu ISO 639-2 |

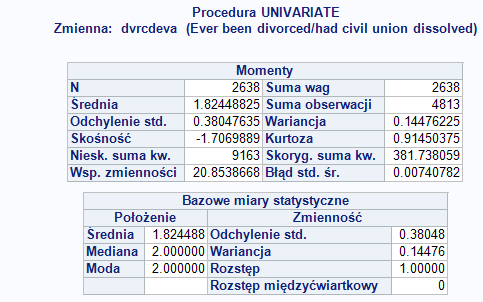
Zmienne te przedstawione są w tabeli formie surowych danych. Aby stworzyć odpowiedni model, należy odpowiednio przetworzyć nasze zmienne - część nich wymaga bowiem kategoryzacji, inne natomiast podlegają procesowi oczyszczenia. Musimy także mieć na uwadze, że niektóre ze zmiennych mogą zostać wykluczone z modelu ze względu na zbyt dużą ilość braku wyników (tzw. ‘missing’) .

# 7. Opis wybranych zmiennych

#### Zmienna objaśniana - dvrcdeva



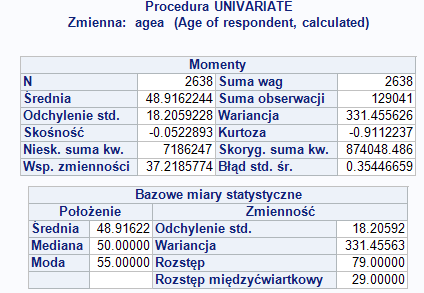




Za zmienna objaśnianą została obrana zmienna dvrcdeva, odpowiadająca na pytanie czy respondent kiedykolwiek był rozwodnikiem. Zgodnie z tabelą - osoby deklarujące rozwód są w mniejszości, jednak stanowią one blisko 20% odpowiedzi. Przy pomocy modelu regresji logistycznej będziemy badać jakie czynniki mogą mieć wpływ na zaistniałe zjawisko.

#### Zmienna objaśniająca - Agea - Wiek respondenta

#### 

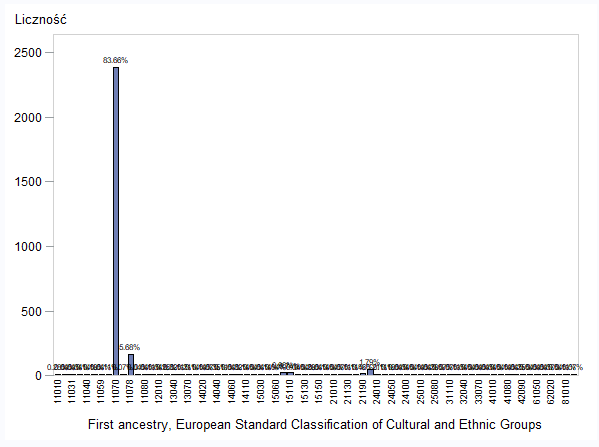


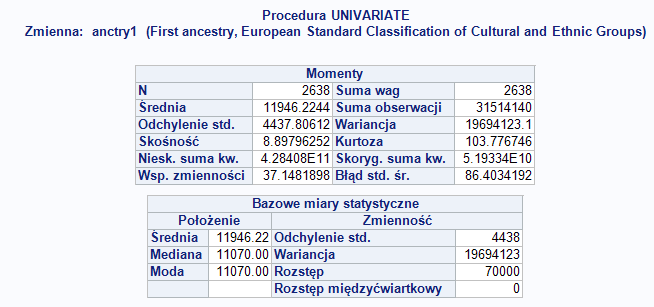
Najwięcej osób, to te które w momencie orzeczenia rozwodu były w wieku 40 – 49 lat (28.1%), 35 – 39 lat (20.8%) oraz 30 – 34 lat (20.0%).

Najczęściej rozwodzą się osoby, które były w momencie zawarcia małżeństwa w wieku 20 – 24 lat (44.6%) oraz 24 – 29 lat (28.8%). [9]

Kurtoza wynosi około -0.911, co świadczy o słabej koncentracji wyników wokół średniej. Średnia wynosi 49 lat, odchylenie standardowe 18, a więc wartości są rozrzucone wokół średniej między 31 a 67 lat, współczynnik asymetrii sugeruje, że rozkład jest zbliżony do rozkładu normalnego (z delikatną przewagą asymetrii lewostronnej).

#### Zmienna objaśniająca - anctry1 - First ancestry

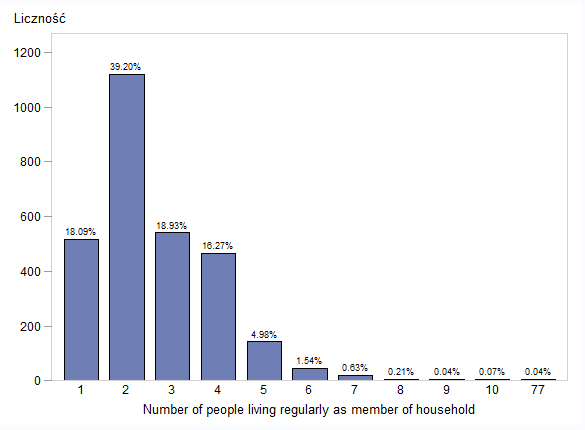


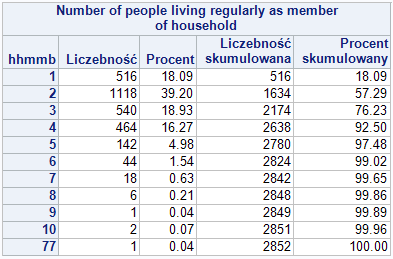


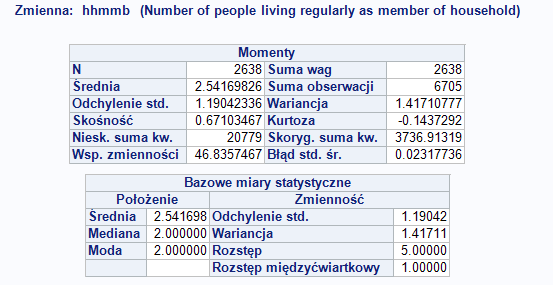
Pierwszą zmienną wybraną do modelu jest zmienna określająca pierwszą z maksymalnie dwóch wskazywanych kultur pochodzenia wskazywanych przez respondenta. Możliwe do wyboru wartości definiowane są przez Europejską Klasyfikację Grup Kulturowych i Etnicznych (ESCEG). Jak stwierdzono z materiałach dodatkowych ESS, to właśnie przynależność do danej grupy etnicznej jest bardzo istotnym prognostykiem, szczególnie w przypadku tradycji i wzorców postępowania, a także stosunku do danych zagadnień. Co warto zauważyć, wartość ta nie ma na celu mierzenia kraju pochodzenia, a raczej pochodzenia kulturowego i etnicznego, które, w coraz bardziej wielonarodowej Europie, ma dziś większe znaczenie i niesie więcej informacji niż kraj pochodzenia rodziców. [10]

Moda i mediana są sobie równe - oznacza to że jednocześnie wartością środkową i najczęstszą jest 11070, a więc osoba pochodzenia Niemieckiego. Wysoka skośność sugeruje, że mamy względnie mocną prawostronną asymetrię. Kurtoza wynosząca prawie 104 świadczy o rozkładzie leptokurtycznym, a więc wartości cechy bardziej skoncentrowane niż przy rozkładzie normalnym.

#### Zmienna objasniająca - Hhmmb - Liczba mieszkańców gospodarstwa domowego





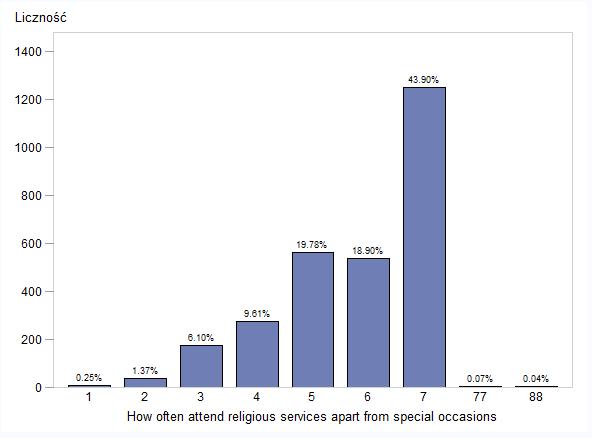


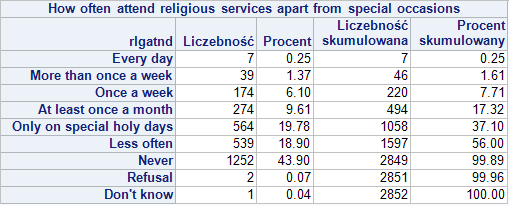
Średnio w gospodarstwie domowym uczestników badania żyją 3 osoby (z racji że jest to wartość całkowita musimy ją zaokrąglić). Dominują wartości 2, odchylenie standardowe 1.19 mówi nam, jak szeroko wartości są rozrzucone wokół średniej. W tym przypadku jest to zakres między 1 a 3 domowników. Współczynnik asymetrii jest delikatnie większy od zera, czyli rozkład charakteryzuje się lekką asymetrią prawostronną. Kurtoza jest ujemna, więc wartości cechy są nieco mniej skoncentrowane niż przy rozkładzie normalnym.

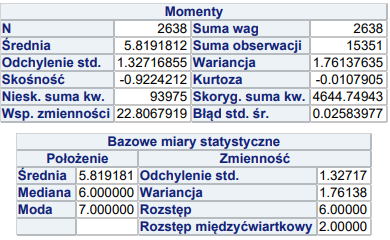
Okazuje się, że ilość osób w gospodarstwie domowym ma wpływ na rozwód.

Odsetek gospodarstw z parą małżeńską mających na utrzymaniu dzieci wykazuje tendencję spadkową, podczas gdy udział gospodarstw kierowanych przez bezdzietne małżeństwa oscyluje wokół takich samych proporcji lub w niektórych latach charakteryzuje się niewielkimi przyrostami. Trendy te tłumaczy się rosnącym wiekiem w momencie pierwszego ślubu i wzrastającym natężeniem rozwodów, co powoduje spadek odsetka ogółu małżeńskich gospodarstw domowych. [11]

#### Zmienna objaśniająca - rlgatnd - Częstość uczestniczenia w praktykach religijnych z pominięciem specjalnych okazji



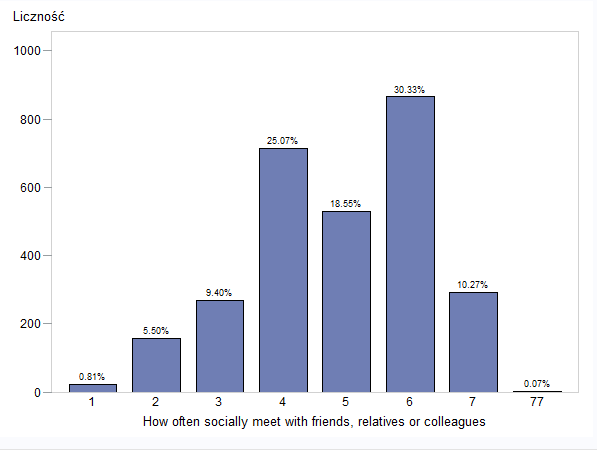


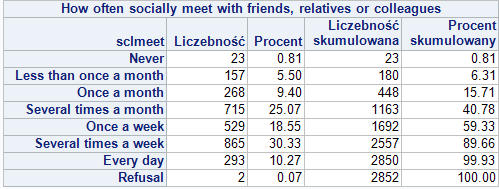


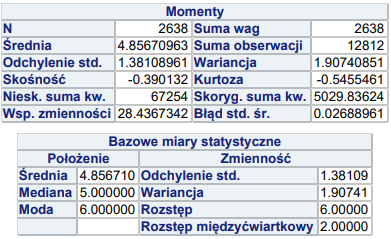
Również w tym wypadku wydaje się, że zmienna opisująca częstość uczestniczenia w praktykach religijnych, a więc pozwalająca wysnuć wniosek nie tylko na temat stosunku do religii, ale również stopnia zaangażowania danego respondenta w praktykach religijnych powinna być w istotny sposób skorelowana z występowaniem rozwodów. Badania mogą wskazywać, że osoby często uczestniczące w obrzędach religijnych, będą mniej skłonne podejmowania decyzji o rozwodzie. [12]

Dane z badania pozwalają zauważyć, że grupa respondentów w większej części nie uczestniczy w obrzędach religijnych (prawie 44%) lub uczestniczy w nich rzadko lub wyłącznie w czasie świąt religijnych (19 i 20%). Mediana wskazuje na rzadkość uczestnictwa w obrzędach religijnych, a moda, że   
z największym prawdopodobieństwem badani mają tendencję do nie uczęszczania w obrzędach religijnych.

#### Zmienna objaśniająca - sclmeet - Częstość kontaktów społecznych z przyjaciółmi, rodziną lub znajomymi z pracy.

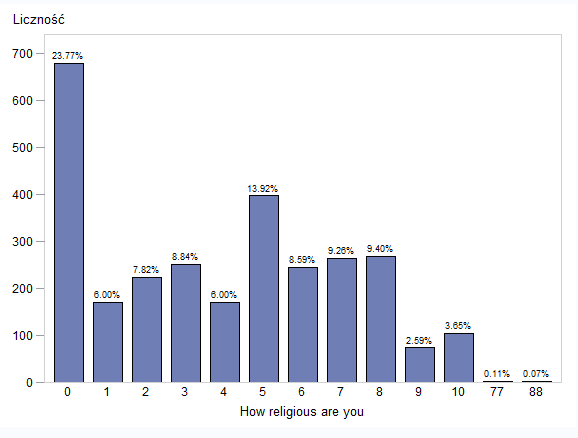


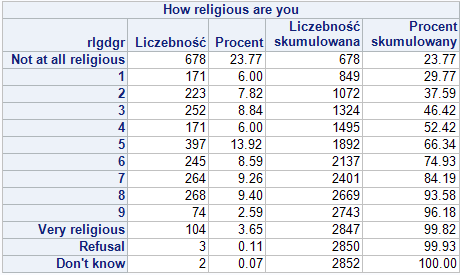




Kolejną zmienną wybraną do badania jest zmienna opisująca intensywność kontaktów społecznych respondentów. Pomimo tego, że korelacja częstości kontaktów społecznych może nie wydawać się oczywista, przykładowe badania przeprowadzone w Holandii potwierdzają [13], że osoby po rozwodzie przejawiają większe zaangażowanie społeczne, częstość ich kontaktów społecznych zwiększa się i co również może przyjmować formę uczestnictwa w zorganizowanych grupach. Co zostało zauważone   
w polskich badaniach [14], kontakty społeczne mogą skutkować zwiększeniem się liczby potencjalnych nowych partnerów, co prowadzi do zmniejszenia kosztów poszukiwania nowego związku i może być istotnym czynnikiem sprzyjającym decyzji o rezygnacji z obecnej relacji. Dane zebrane do badania pozwalają zauważyć, że u większości respondentów częstość kontaktów społecznych można uznać za wysoką (ponad 40% spotyka się z rodziną i znajomymi każdego dnia lub kilka razy w tygodniu, prawie 60% natomiast raz na tydzień lub częściej). Około 35% respondentów spotyka się z rodziną i znajomymi rzadziej niż raz w tygodniu ale częściej niż raz w miesiącu.

#### Zmienna objaśniająca - rlgdgr - opinia na temat własnej religijności

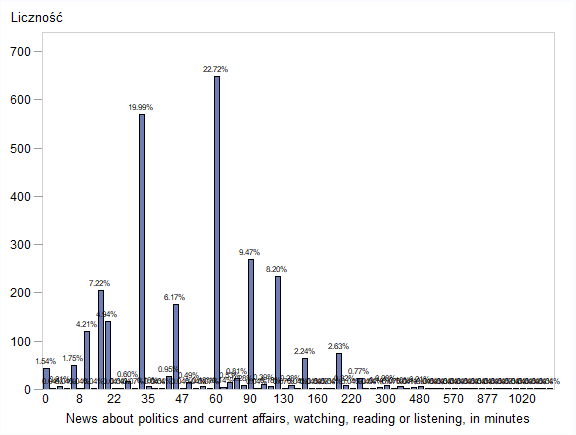






Podobnie jak zmienna rlgatnd, ta zmienna odnosi się do religijności danego respondenta, jednak tym razem, w odróżnieniu od religijności faktycznej, mierzonej częstością uczestniczenia w obrzędach religijnych, opisywana jest deklarowana religijność ankietowanego. Można przypuszczać, że z uwagi na wyraźnie negatywny stosunek wielu religii do kwestii rozwodów [15], mocne przeświadczenie o własnej religijności powinno być ujemnie skorelowane z możliwością zdecydowania się na zakończenie związku. Również badania zdają się potwierdzać, że przynależność religijna i wysoki stopień deklarowanej religijności są czynnikami stabilizującymi małżeństwo [16]. Badając przedstawione dane można zauważyć, że znaczna ilość respondentów określiła się jako osoby całkowicie niereligijne, następnym wyróżniającym stopniem jest określenie religijności na 5 w dziesięciostopniowej skali, które deklaruje niespełna 14% respondentów. Co warto zauważyć, większa część badanej populacji deklaruje religijność ocenioną na 5 i mniej. Jedynie 3,5% respondentów określiło się jako osoby bardzo religijne.

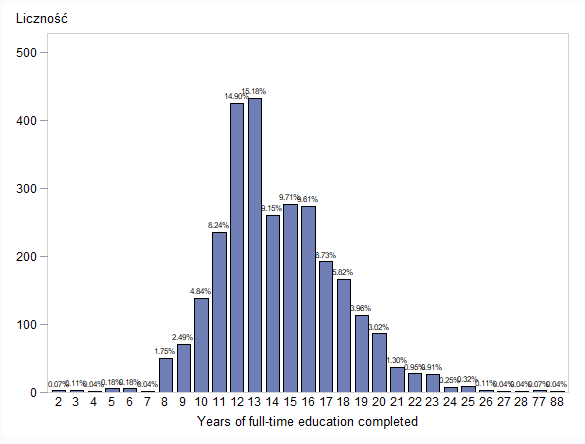
#### Zmienna objaśniająca - nwspol - dzienny czas przeznaczony na wiadomości z polityki





Polityka i rozwód często idą w parze. Takie spory zawsze były ogromnym źródłem zrujnowanych małżeństw. Taki konflikt może zabić związki i doprowadzić do rozwodu. Respondent spędza przeciętnie 67 minut dziennie na czytanie, słuchanie lub oglądanie informacji ze sceny politycznej.

#### Zmienna objaśniająca - eduyrs - czas edukacji respondenta w latach





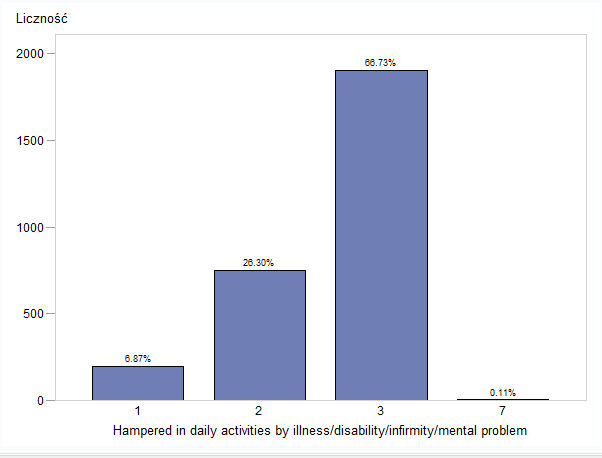


Najwięcej rozwiedzionych mężczyzn legitymowało się wykształceniem zasadniczym zawodowym (35.1%), średnim (34.7%) lub wyższym (19.8%); kobiety zaś najczęściej posiadały wykształcenie średnie (38.5%), wyższe (31.3%) oraz zasadnicze zawodowe (20.8%). [17]

Zespół FAMWELL dokonał ilościowej analizy dotyczącej związku pomiędzy wykształceniem kobiet,   
a ryzykiem rozwodu. Okazuje się, że zgodnie z hipotezą Williama J. Goodego relacja pomiędzy wykształceniem a ryzykiem rozwodu zależy od tego, jak powszechne i jak łatwo dostępne są w danym społeczeństwie rozwody.

Kiedy rozwody są mało rozpowszechnione a związane z nimi koszty legislacyjne, materialne i społeczne są wysokie, kobiety lepiej wykształcone są w stanie lepiej poradzić sobie z tymi trudnościami i rozwodzą się częściej niż kobiety gorzej wykształcone. Różnica ta maleje, gdy rozwody stają się coraz bardziej powszechne i mniej kosztowne. Wówczas również kobiety gorzej wykształcone i dysponujące mniejszymi zasobami mogą pozwolić sobie na rozwód. Gdy rozwody są powszechne i łatwo dostępne, wyższe wykształcenie kobiet zaczyna sprzyjać większej stabilności małżeństw. [18]

#### Zmienna objaśniająca - hlthhmp - utrudnienie w życiu codziennymi wynikające z choroby, niepełnosprawności, choroby umysłowej etc



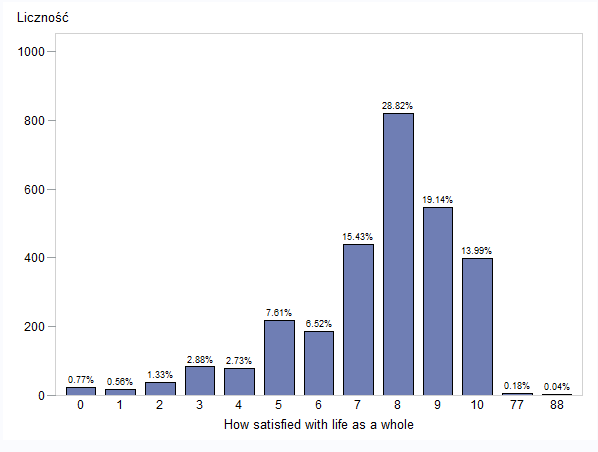


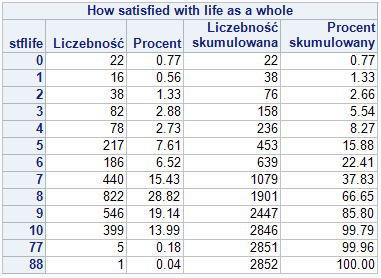


Badanie przedstawia wyniki wywiadów przeprowadzonych z 52 rozwiedzionymi osobami, które brały udział w programie zapobieganiu i poprawy relacji w związku małżeńskich.

Najczęściej zgłaszanymi przyczynami rozwodu był brak zaangażowania, niewierność i konflikt / kłótnia, przemoc domowa i używanie substancji. Uczestnicy obwiniali swoich partnerów bardziej niż samych siebie. Wskazali oni również, że choroba oraz problemy ze zdrowiem mogą być przyczyną rozwodu. [19]

#### Zmienna objaśniająca - stflife - satysfakcja z życia



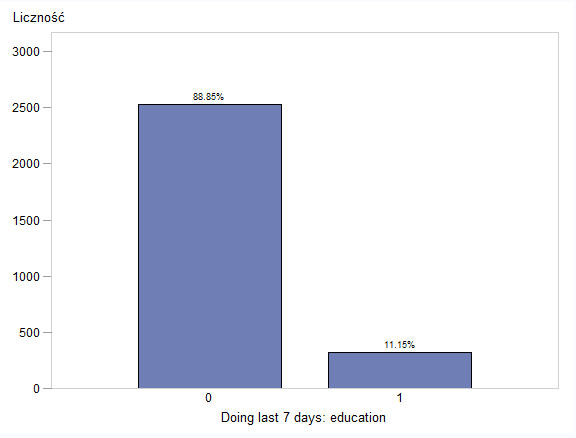


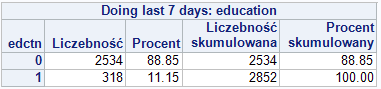


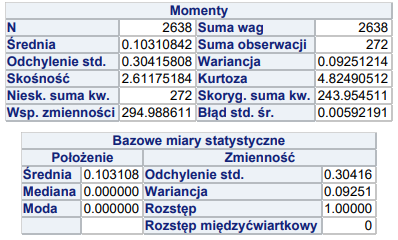
Wydawać by się mogło, że zawarcie związku małżeńskiego lub wejście w związek kohabitacyjny zwiększa dobrobyt jednostki, sugerując tym samym, że związek przyczynowy przebiega od małżeństwa lub konkubinatu do zadowolenia z życia. Stack i Eshleman (1998) [20] stwierdzili, że osoby rozwiedzione, owdowiałe i odseparowane miały niższy poziom dobrostanu w porównaniu z pojedynczymi osobami. Fakt, że analizowane są tylko kraje rozwinięte, jest jedną z wad ich analizy, ponieważ nie dostarcza wyraźnych dowodów na ewentualne różnice między bogatszymi a biedniejszymi krajami w odniesieniu do związku między stanem cywilnym a zadowoleniem z życia.

Respondenci wykazują wysoki poziom zadowolenia z życia, aż 44 % badanych jest bardzo zadowolonych ze swojego życia.

#### Zmienna objaśniająca - edctn - zmienna binarna określająca zajmowanie się edukacja w ostatnich 7 dniach od ankiety

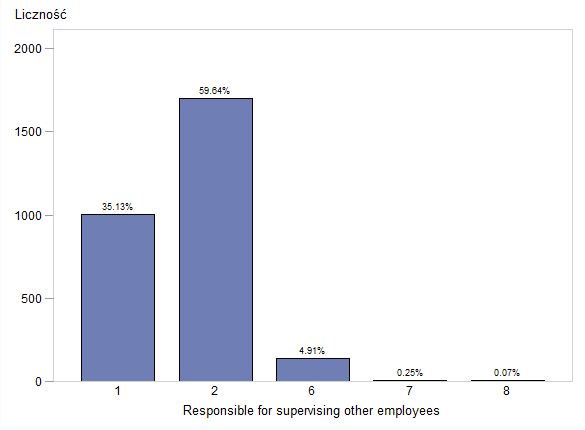


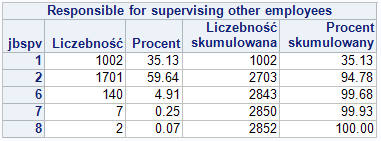


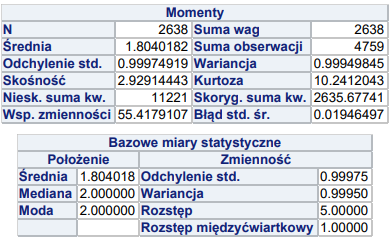


Zmienna odpowiada na pytanie o działalność naukowa respondenta w ostatnich 7 dniach od przeprowadzenia badania. Z rozkładu odpowiedzi wynika, że prawie 90% badanych nie dokształcali się w okresie bliskim do badania. Warto jednak zaznaczyć, że mowa tu o edukacji nie opłacanej przez pracodawcę, dlatego wszelkie szkolenia w pracy czy z nią związane nie zostały objęte w ramach odpowiedzi twierdzącej.

#### Zmienna objaśniająca - jbspv - pozycja w pracy wymagająca zarządzania innymi pracownikami

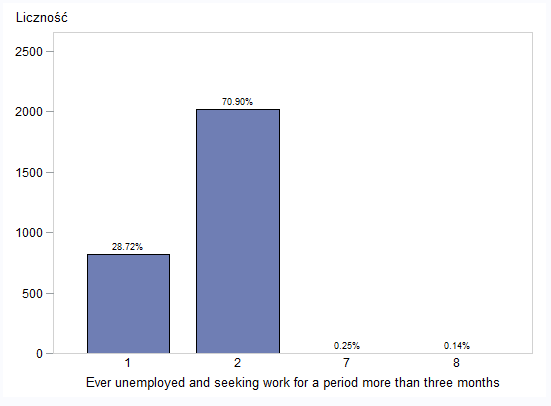


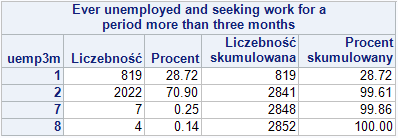




Osoby, pracujące na własny rachunek stanowią 6.9% rozwiedzionych par. [21] Zmienna opisuje, czy badany zajmuje w pracy stanowisko, które wymaga zarządzania innymi pracownikami.

#### Zmienna objaśniająca - uemp3m - okres przebywania na bezrobociu lub poszukiwania pracy równy lub dłuższy niż 3 miesięcy



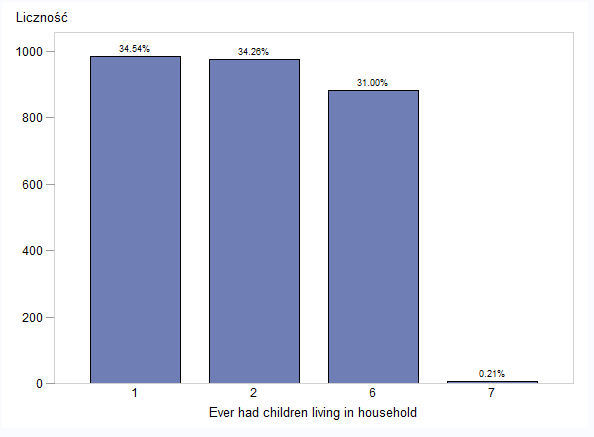


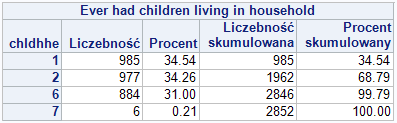


Osoby, które się rozwiodły i były bezrobotne lub bierne zawodowo to około 17 % rozwiedzionych par.   
83 % rozwiedzionych stanowiły osoby pracujące. [21]

W badaniach Paula R. Amato dopiero na 10 miejscu jako powód rozwodu znajdowały się problemy   
z zatrudnieniem (2.9 % odpowiedzi respondentów o przyczynę rozwodów) oraz na 13 miejscu problemy finansowe (2.0%). [22]

#### Zmienna objaśniająca - childhhe - Obecność dziecka w gospodarstwie domowym



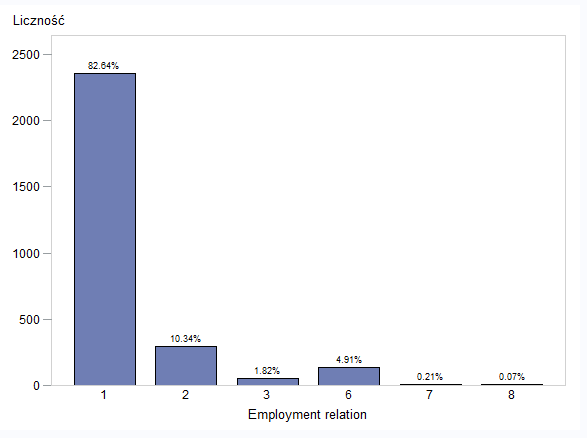


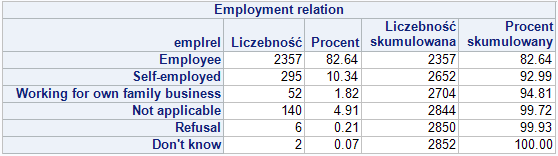


Na podstawie danych z pierwszej fali badania chińskiego panelu rodzinnego w 2010 r. [23] wynika, że istnieje związek pomiędzy ryzykiem rozwodu, a obecnością dzieci w chińskich rodzinach. Okazuje się, że ten czynnik ma znaczący wpływ na ryzyko rozwodu. W szczególności pary, które mają dzieci przedmałżeńskie, są bardziej narażone na rozwód, im większa liczba dzieci, tym bardziej stabilne małżeństwo, ale marginalny efekt maleje wraz ze wzrostem liczby dzieci, młodsze dzieci bardziej niż starsze zmniejszają ryzyko rozwodu oraz pary, które mają synów, rzadziej się rozwodzą. Wszystkie te wyniki są zgodne z wynikami badań przeprowadzonych w innych krajach i sugerują, że dzieci odgrywają ważną rolę w stabilizowaniu małżeństwa i rozwiązywaniu konfliktów w małżeństwie. Odkryto również, że pozytywny wpływ dzieci na stabilność małżeńską i wpływ płci dzieci są silniejsze na obszarach wiejskich, co wynika z różnic w modernizacji i wartości kulturowych między obszarami wiejskimi i miejskimi. Wyniki te sugerują, że zmiany wartości małżeństwa i spadek współczynnika dzietności to dwie ważne przyczyny niedawnego wzrostu liczby rozwodów w Chinach.

Na podstawie danych z Rocznika Demograficznego 2017 wydanego przez Główny Urząd Statystyczny można stwierdzić, że liczba rozwodów się zwiększyła. W roku 1980 było ich około 40 tysięcy, zaś już   
w 2016 około 63 i pół tysiąca. 41,3 % rozwiedzionych małżeństw to były małżeństwa bez dzieci. Rozwiedzione małżeństwa, które posiadały dzieci to 58,7%. Liczba dzieci w małżeństwie ma znaczenie, czy para się rozwiedzie. Obrazują to dane. Rozwiedzione małżeństwo z jednym dzieckiem to 36% rozwodów, z dwójką dzieci to 19,1%, z trójką to 3%, a z czwórką lub więcej to jedynie 0.7%. Na podstawie tych danych można stwierdzić, że jeśli nie chcemy się rozwieść to powinniśmy pomyśleć o przynajmniej czwórce dzieci. [21] . W dalszej analizie sprawdzimy, czy powyższe wyniki powyższych badań pokrywają się analizą dla Niemczech.

#### Zmienna objaśniająca - Emplrel - Rodzaj stosunku pracy



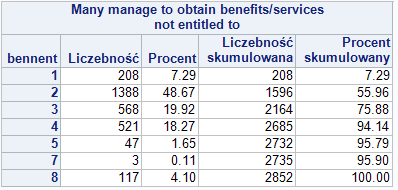




Następną zmienną objaśniającą wybraną do badania jest zmienna opisująca rodzaj stosunku pracy jaki wykonuje respondent. Co warto zauważyć, najistotniejsze z wyróżnionych kategorii to zatrudnienie, samozatrudnienie i praca w firmie rodzinnej. Socjologiczna analiza wpływu rodzaju stosunku pracy   
i wykonywania pracy jako takiej na możliwość wystąpienia rozwodu może prowadzić do nieoczywistych wniosków. Według badań [24], aktywność zawodowa kobiet jest ujemnie skorelowana z czerpaniem korzyści obojga partnerów z małżeństwa, co związane jest z brakiem efektu specjalizacji w związku rozumianego tutaj jako specjalizację mężczyzny w pracy zarobkowej przy specjalizacji kobiety w pracy wewnątrz gospodarstwa domowego. Dodatkowo, wynikająca z tego niesamodzielność finansowa kobiet utrudnia im decyzję o rozwiązaniu małżeństwa. Z drugiej zaś strony, zdarza się, że aktywność zawodowa kobiet wpływa pozytywnie na sytuację finansową związku, co stabilizuje relację [25]. Oczywiście, należy zaznaczyć, że bezrobocie zarówno u kobiet jak i u mężczyzn wpływa negatywnie na stabilność małżeństwa [9]. Dane przedstawione w badaniu wskazują na znaczną przewagę osób pracujących (82,5%) nad samozatrudnionymi i pracującymi w rodzinnej firmie (10,3% i 1,8%)

#### Zmienna objaśniająca - bennent - opinia na temat zasiłków w Państwie







W sytuacji braku wzrostu gospodarczego, wysokiej stopy bezrobocia oraz stagnującego wskaźnika zatrudnienia przy jednoczesnym obniżaniu się realnego poziomu wynagrodzenia niewystarczające wpływy składkowe prowadzą do deficytu systemu ubezpieczeń społecznych. Czynniki te przyczyniły się do poszukiwania w końcu lat 90. XX w. możliwości odciążenia tych systemów. Tradycyjnie mocno rozbudowana tzw. siatka zabezpieczeń społecznych miała zostać poluźniona. Cały pakiet reform zaplanowanych przez ówczesnego kanclerza, Gerharda Schrödera, mimo wszystko nie wszedł w życie. Zrealizowano natomiast reformę zasiłków dla bezrobotnych określaną potocznie reformą Hartz IV.

Z badań niemieckich [31], a także z porównawczych badań UE oraz OECD wynika, że Niemcy należą do krajów, w których transfery socjalne (zasiłki dla bezrobotnych, dodatek na dziecko, dodatek do czynszu   
i in. świadczenia) redukują

zagrożenie ubóstwem z 26% do 13% społeczeństwa.

Z zebranych danych wynika, że ponad 55% ankietowanych popiera zasiłki, 37% nie ma jasno określonego zdania i tylko niespełna 2% jest przeciwnych.

#### Zmienna objaśniająca - icwhct - posiadanie podstawowej lub stałej liczby godzin pracy







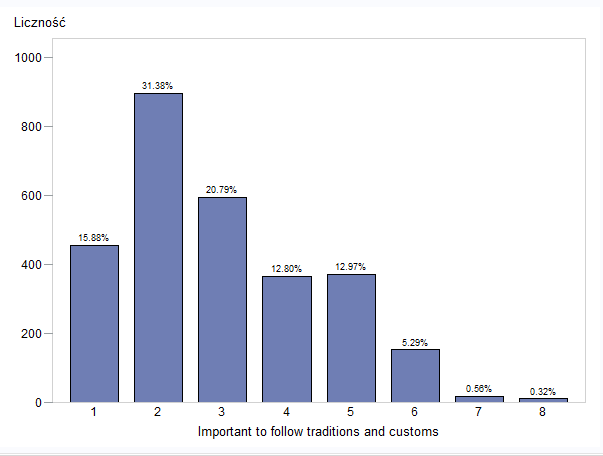
Opublikowane w American Sociological Review w 2016 [26] roku badania sugerują, że to nie tyle stan majątkowy pary może być określać ich szansę na rozstanie, ale wkład finansowy poczyniony przez małżonków.

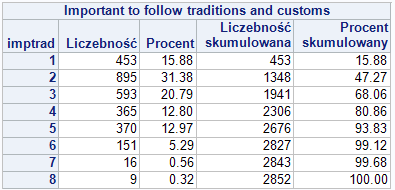
Badaczka Alexandra Killewald porównała pary małżeńskie, które pobrały się po 1975 roku. Okazało się, że związki, w których partner nie pracował w pełnym wymiarze godzin (np. na cały etat), miały o 3,3% większe szanse na rozwód już w kolejnym roku trwania małżeństwa, w porównaniu 2,5% szans wśród par, w których mąż pracował w pełnym wymiarze godzin.

Co zaskakujące, ilość godzin, którą partnerka poświęcała na pracę nie miały wpływu na potencjalne szanse rozwodu. Jak tłumaczą naukowcy, stereotyp męskiego żywiciela rodziny jest wciąż żywy i może mieć wpływ na stabilność związku.

Wśród respondentów aż 88,40% pracuje na etacie w pełnym wymiarze godzin, 8,7% na kontrakcie,   
a 4,91% nie udzieliło odpowiedzi.

#### Zmienna objaśniająca - Imptrad - Subiektywna ocena respondenta podobieństwa do osoby, dla której tradycje i wartości mają dużą wagę





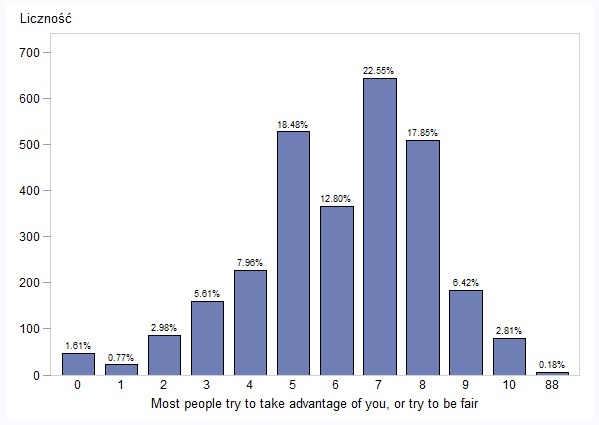


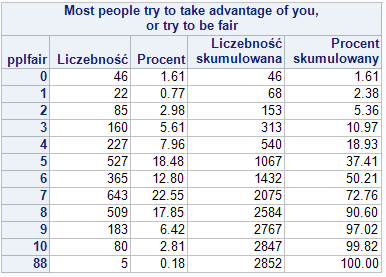
Najczęściej jako przyczynę rozwodu małżonkowie deklarują niezgodność charakterów (2/5 wszystkich rozwodów), kolejne przyczyny to niedochowanie wierności małżeńskiej (21.9%) , alkoholizm (16.7% rozwodów), zaś różnice światopoglądowe (0.1%). [17]

W badaniach Paula R. Amato najczęstszym powodem rozpadu małżeństwa była niewierność (18.1 % odpowiedzi respondentów o przyczynę rozwodów) oraz niedopasowanie (16.4%). [22]

Analizując modę, ze zbioru danych wynika, iż respondenci utożsamiają się z osobą, dla której tradycje   
i wartości mają dużą wagę.

#### Zmienna objaśniająca - pplfair - ocena zaufania do innych ludzi



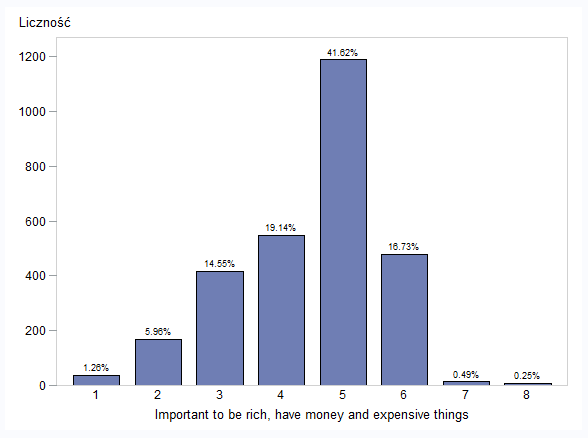


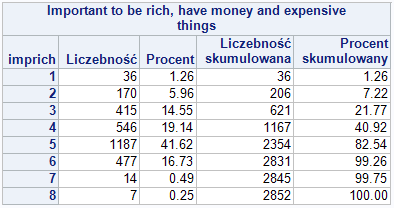


Dane respondentów wskazują, że w życiu społecznym Niemcy są raczej ufni. Tylko nieco ponad 10% ankietowanych wychodzi z założenia, że większości nie można ufać. Najczęstszą odpowiedzią była wartość 7 w skali 0-10, oznaczająca wysoki stopień zaufania do innych ludzi.

Wartość najczęściej występująca w próbie to 7, więc grupa badanych częściej wykazuje się ufnością do innych.

#### Zmienna objaśniająca - imprich - potrzeba bogactwa i posiadania pieniędzy





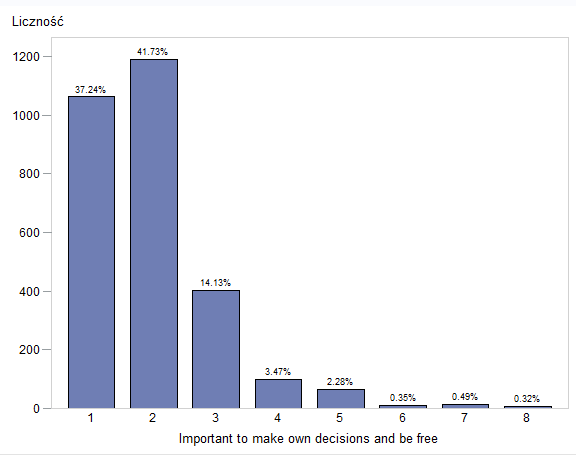


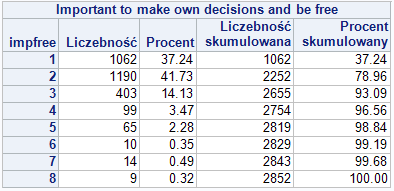
Chciwość to postawa, która sprawia, że rzeczy materialne stają się dla człowieka ważniejsze od osób. Chciwiec bardziej „kocha” pieniądze niż ludzi. W konsekwencji staje się niewrażliwy na los i na potrzeby bliźnich.

Sama chęć posiadania pieniędzy może nie być bezpośrednią przyczyną, ponieważ za nią idą działania takie jak pracoholizm, który z kolei zmniejsza ilość czasu poświęcanego rodzinie (tak jak w przypadku zmiennej *icwhct*).

41% respondentów silnie twierdzi, iż wcale nie traktuje bogactwa jako celu i nie ocenia chęci posiadania jako silnej, a kolejne 19% potwierdza brak takiego podejścia. Około 7% odpowiadających stawia sobie za cel zdobywanie pieniędzy i przypisuje temu dużą wagę. Miara dominanty wskazuje na wartość   
o największym prawdopodobieństwie wystąpienia wyniku z nie utożsamieniem się z pogonią za pieniądzem, ale z drugiej strony rozkład przesunięty lekko w prawo podpowiada nam większy udział egoistów wśród badanych.

#### Zmienna objaśniająca - impfree - potrzeba niezależności i wolności



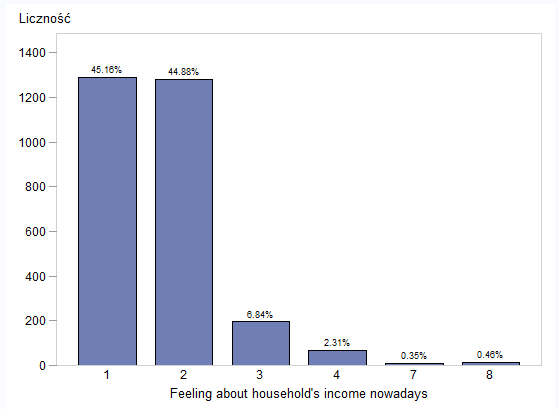


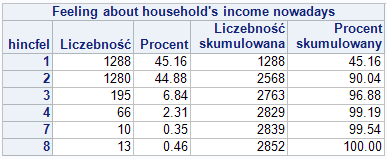


Niezależność definiowana wg PWN to niepodporządkowany komuś, czemuś, decydujący o sobie; też: świadczący o braku podporządkowania komuś lub czemuś, niebędący wyznaczonym, zdeterminowanym przez coś, wygłaszający bezstronne opinie, nie kierujący się interesem żadnej grupy społecznej. [27] Rośnie potrzeba niezależności i indywidualizmu, co skutkuje obniżeniem trwałości związków romantycznych. Ta niepewność, powoduje zmniejszenie inwestycji w kapitał specyficzny dla pary, taki jak wzajemna miłość, troska czy wsparcie. [28]

37% respondentów definiowało tę potrzebę jako bardzo ważną, a 41% jako ważną. Przy rosnącej potrzebie niezależności, ryzyko rozwodu może ulec podwyższeniu.

#### Zmienna objaśniająca - hincfel - poczucie komfortu / stabilności, w związku z obecnym stanem budżetu domowego







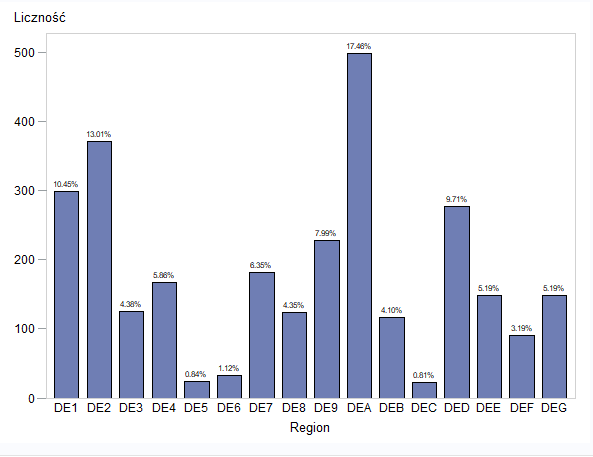
Czynniki, które różnicują poczucie własności pieniędzy, postawy wobec pieniędzy mogą wpłynąć na satysfakcję ze związku. Konflikt również może stanowić styl zarządzania pieniędzmi. Z badań   
C. Burgoyne i V. Morison (1997) z udziałem par, w których przynajmniej jeden z partnerów był wcześniej rozwiedziony, wynika ponadto, że kiedy małżeństwo jest postrzegane jako niestabilne czy niepewne, małżonkowie wybierają takie rozwiązanie, które sprawi, że finansowe rozdzielenie się będzie dla nich łatwiejsze (jeżeli małżeństwo nie przetrwa). [29]

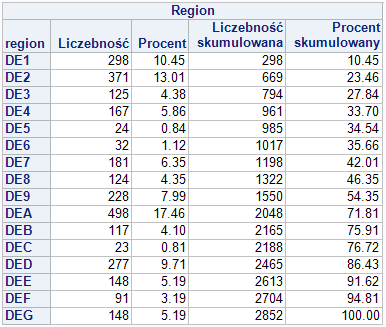
## 

Jak wynika z danych Ministerstwa Sprawiedliwości i GUS na ponad 66 tys. rozwodów, do jakich doszło   
w Polsce w 2013 r., ze względu na nieporozumienia na tle finansowym rozeszło się blisko 5,2 tys. par, czyli około 8 proc. wszystkich, które wówczas uzyskały rozwód. Po niezgodności charakterów, niewierności i nadużywaniu alkoholu finanse są czwartą co do popularności przyczyną rozwodów. [30]

45% respondentów prowadziło komfortowe życie przy obecnych dochodach oraz kolejne 45% radziło sobie nie narzekając na stan dochodowy. Pozostali respondenci skarżyli się bądź unikali odpowiedzi   
o satysfakcji z zarobków.

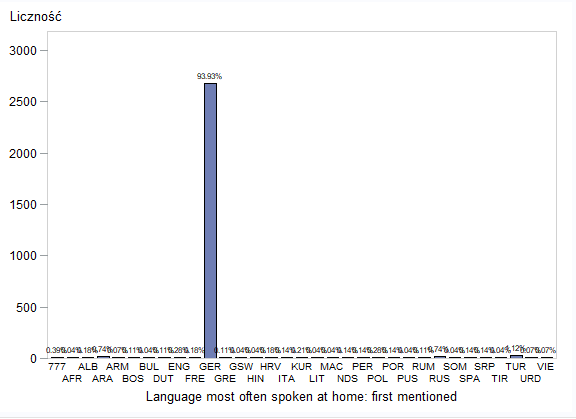
#### Zmienna objaśniająca - region - region zamieszkiwany przez respondenta





Niemcy jest to republika federacyjna, składająca się z 16 krajów związkowych, zwanych landami. Podobnie jak polskie województwa, czy też bliższe pod względem politycznym stany w Stanach Zjednoczonych landy różnią się między sobą zaludnieniem, zamożnością, poglądami politycznymi etc. Najwięcej respondentów przynależy odpowiednio do landów DEA, DE2 oraz DE1, którym odpowiadają landy Nordrhein-Westfalen, Bavaria oraz Baden-Württemberg, które są też odpowiednio landami   
o najwyższej populacji. W dalszej części analizy zmienna ta została skategoryzowana pod kątem podziału na dawne kraje NRD i RFN.

#### Zmienna objaśniająca - lnghorn1 - język najczęściej używany w domu



Jak możemy zaobserwować na wykresie, najwięcej osób mówi w domu w języku niemieckim - jest to aż 93.93%. Kolejnymi najliczniejszymi grupami są Turecki (TUR, 1.12%), Arabski (ARA) i Rosyjski (RUS) - po 0.74%.

# 8. Kategoryzacja zmiennych

Na potrzeby projektu, część zmiennych zdecydowaliśmy się przekształcić i skategoryzować.

#### Agea



Najmłodszy uczestnik badania miał 15 lat, zaś najstarszy 94 lata.

Średnia wieku ankietowanego w przybliżeniu to 48 lata i 7 miesięcy, zaś mediana to 50 lat.

Największą liczbę respondentów stanowiły osoby w wieku 55 lat.

Zauważmy że zmienną agea możemy podzielić na 3 kategorie - ludzie młodzi (*young*), w średnim wieku (*middle*) i starzy (*old*). Patrząc na rozkład zmiennych, za ludzi młodych przyjmujemy osoby w wieku do 45 lat, w średnim wieku 46 - 66 lat, a starych od 67 roku życia.

#### Anctry1



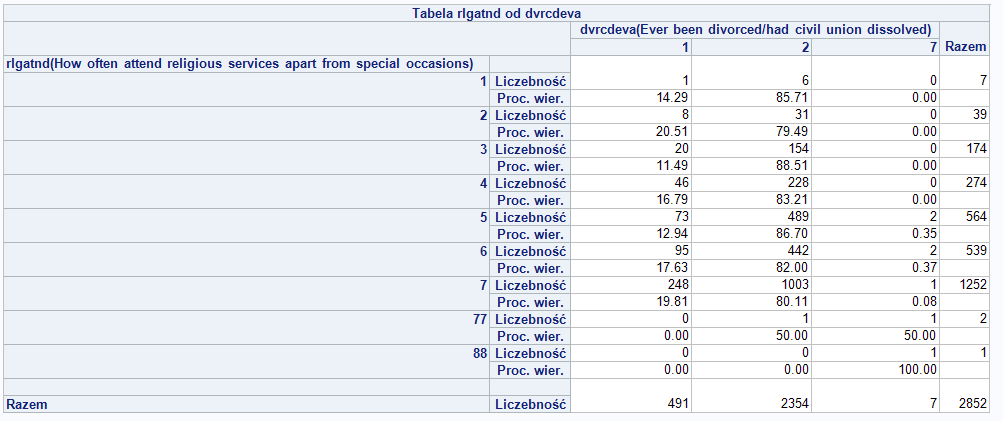
2551 spośród 2852 osób deklaruje pochodzenie jako “Niemieckie”, a więc wybrało opcję 11070, 11077, 11078, 11079. Ze względu na to, jak liczna jest to grupa, zmienne oznaczymy odpowiednio jako “GER”   
i “OTH”.

#### Hhmmb



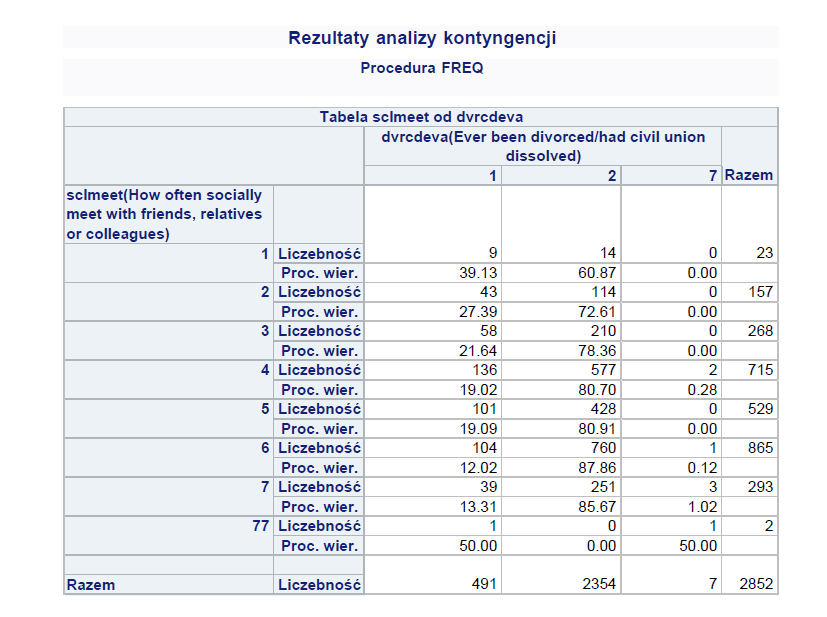
Bazując na tabeli możemy podzielić wartości na “low” i “high”. Jako “low”, a więc względnie małą liczbę domowników, przyjmiemy zakres od 1 do 3 osób. 4 do 6 mieszkańców klasyfikujemy jako “high”. Pozostałe wartości pomijamy, jako mało liczne.

#### Rlgatnd



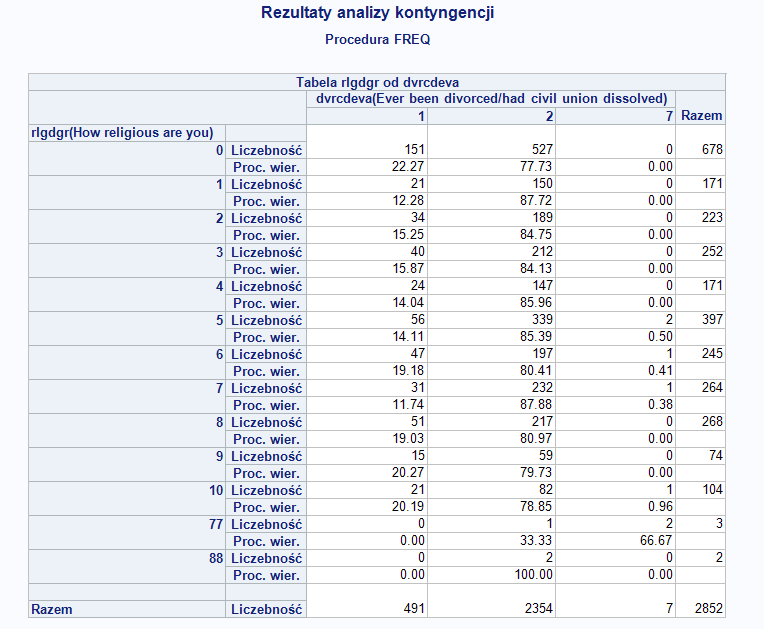
Zmienna ta określa zaangażowanie w praktyki religijne. Jak widzimy, ciężko będzie nam zagregować odpowiedzi w kategorie, dlatego też poprzestaniemy na usunięciu missingów (77 i 88).

#### Sclmeet



Patrząc na rezultaty analizy kontyngencji możemy wyodrębnić trzy kategorie. Pierwszą z nich stanowią osoby które rzadko spotykają się ze znajomymi, a więc zaznaczyły opcję 1 - “nigdy” oraz 2 - “rzadziej niż raz w miesiącu”. Średnio 31% z nich się rozwodzi. Kolejną grupą są osoby które ze znajomymi spotykają się “umiarkowanie” często, a więc 3 - “raz w miesiącu”, 4 - “kilka razy w miesiącu” i 5 - “raz w tygodniu”. Tutaj średnia rozwodów wynosi około 20%. Ostatnią kategorią są ludzie widzący się z bliskimi 6 - “kilka razy w tygodniu” i 7 - “każdego dnia”, a więc “często”. Dla nich średnia wynosi 12.5%.

#### Rlgdgr



Wartości z zakresu od 0 - niereligijny do 10 - bardzo religijny Analizując tablicę kontyngencji możemy wydzielić 3 kategorie - 0 jako osobę bardzo “niereligijna”, 1-5 określa człowieka “umiarkowanie” religijną, a od 6 do 10 “religijna”. Wartości 77 i 88 pomijamy.

#### Nwspol

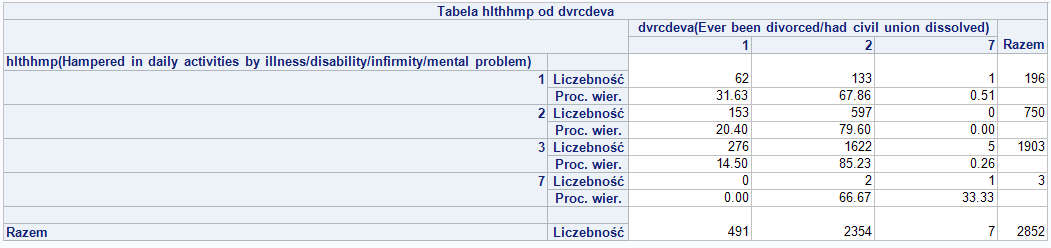
Możemy zauważyć że najczęściej zaznaczane były okrągłe wartości, jak 15, 30, 45 min, 1, 1.5, 2h.   
W przypadku tej zmiennej kategoryzujemy ją bazując na wartości procentu skumulowanego - 40.7% osób zadeklarowało wartości od 0 do 30 minut włącznie (a więc możemy założyć, że osoby te interesują się wiadomościami i polityką względnie rzadko - “rarely”). Następny wydzielony przedział to osoby które podały opcje 31 min do 1 godziny - zrobiło to bowiem 30.9% badanych. Nadamy mu etykietę średnio - “medium”. Pozostałe wartości zadeklarujemy jako “often” - od godziny wzwyż.

#### eduyrs



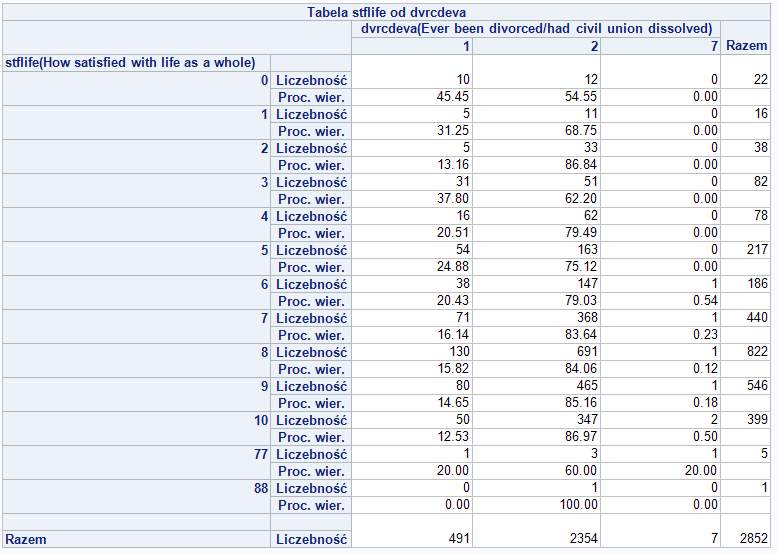
Analizując wykres rozkładu zdecydowaliśmy się podzielić zmienną eduyrs na 3 kategorie (low, medium, high) - osoby mające za sobą do 10 lat edukacji włącznie, od 10 do 18 i od 19 lat edukacji.

#### Hlthhmp



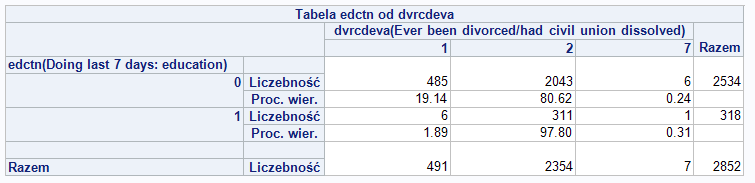
W zmiennej tej pomijamy missingi, a więc kategorię 7, z racji tego, że stanowią one jedynie 3   
z 2852 wartości. Pozostałe stanowią 3 kategorie.

#### Stflife



Zmienna Stflife opisuje satysfakcję z życia respondenta. Tak jak poprzednio odsiewamy missingi, które stanowią około 0.2% całości. Jeśli chodzi o pozostałe zmienne, musimy przeanalizować tablicę kontyngencji. Widzimy z niej, że wraz ze wzrostem satysfakcji z życia maleje stosunek osób rozwiedzionych do osób będących w stałym związku. Bazując na tabeli, zmienną Stflife możemy podzielić na 3 kategorie: “low”, dla wartości 0-3, “medium” - od 4 do 6, a pozostałe - (7 - 10) - jako “high”.

#### edctn



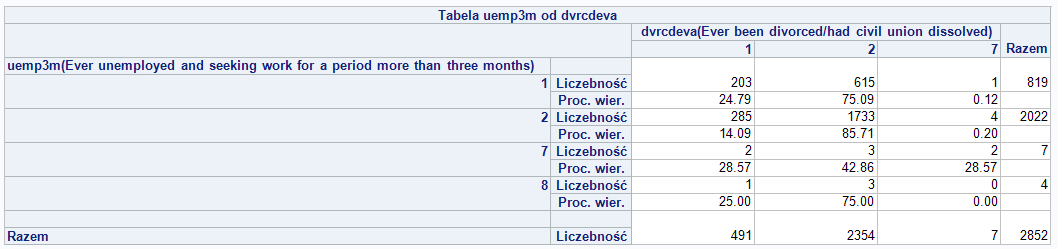
Ponieważ jest to zmienna binarna, nie potrzebna jest dodatkowa kategoryzacja.

#### Jbspv



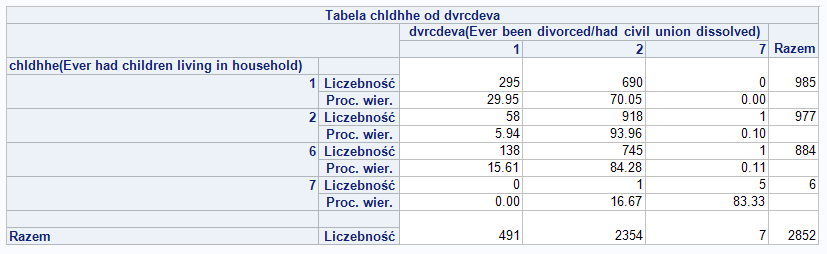
Analizując tabelę widzimy, że możemy pominąć missingi, a więc odpowiedzi 7 i 8. Pozostałe zmienne pozostają w pierwotnej postaci.

#### Uemp3m



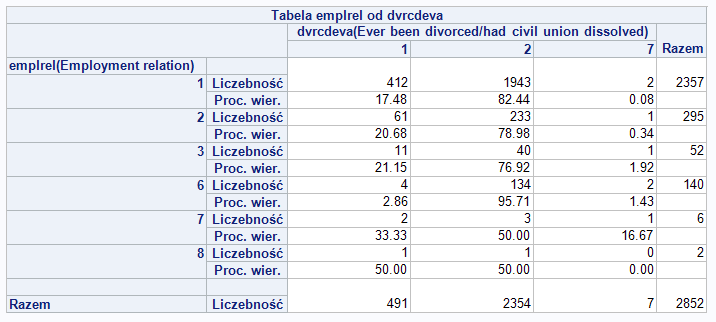
Bazując na tablicy kontyngencji oraz na fakcie, że wartości 7 i 8, czyli odpowiednio ‘Refusal’ i ‘Don't know’ stanowią łącznie 4 obserwacje, możemy je pominąć.

#### Chldhhe



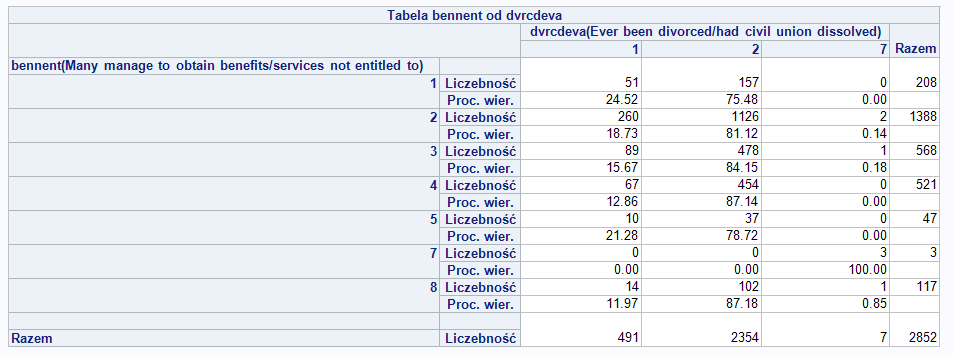
Analizując rozkład zmiennych widzimy, że każda z wartości 1,2 i 6, czyli ‘Yes’, ‘No’, ‘Not applicable’ ma   
w przybliżeniu podobną liczebność. Jednym wyjątkiem jest wartość 7, czyli ‘Refusal’, która stanowi 4 obserwacje. Ponadto każda z kategorii cechuje się dużą zmianą jeśli chodzi o odsetek rozwodników należących do niej. Ostatecznie pomijamy wartość “Refusal” i pozostawiamy 3 pozostałe kategorie.

#### Emplrel



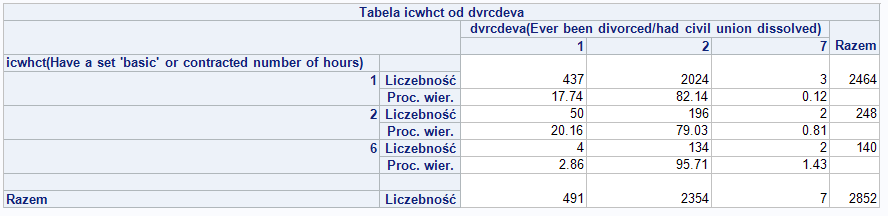
Pomijamy odpowiedź 7 i 8 jako mało liczną. Z tablicy kontyngencji wynika, że powinniśmy wydzielić opcję nr 6, czyli “Not applicable”, ponieważ na 140 osób 134 jest rozwodnikami. Pozostałe kategoryzujemy jako “business” - czyli działalność gospodarcza (swoja - 2 lub rodzinna - 3) oraz etat (1 - “position”).

#### Bennent



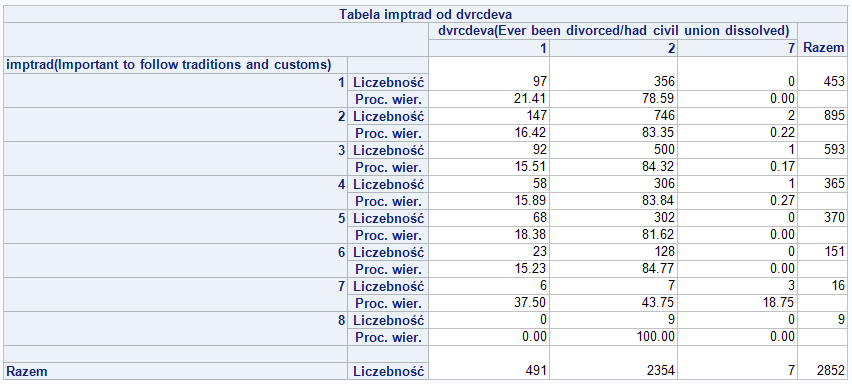
Bazując na analizie kontyngencji, możemy przydzielić osoby które zaznaczyły wartości skrajne 1, 2 i 5 do jednej grupy (“Extreme”), a do drugiej ci którzy wybrali wartości 3 - 4 (“Middle”). Wartości “Refusal” i “Don't know” również przypisujemy do kategorii “Middle”, ponieważ mają podobny odsetek rozwodników.

#### Icwhct



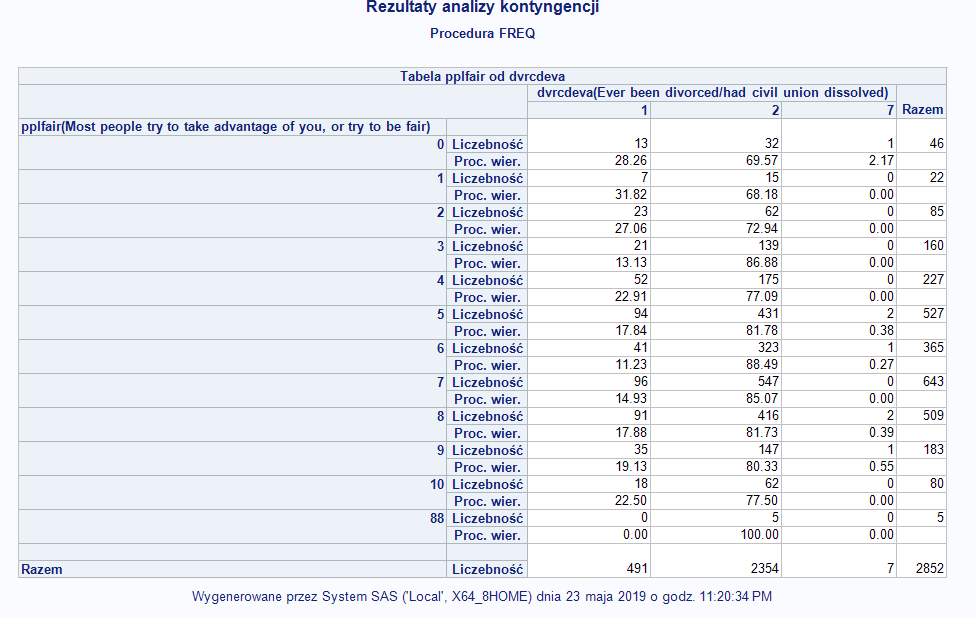
Bazując na rezultatach analizy kontyngencji możemy zakwalifikować osoby które odmówiły udzielenia odpowiedzi do grupy 1, czyli tych respondentów, którzy udzielili odpowiedzi twierdzącej na pytanie, ponieważ stosunek osób rozwiedzionych do trwających w związku jest w tych grupach podobny. Zatem możemy przyjąć 2 kategorie - “answer” dla 1 i 6 oraz “no answer” dla 2.

#### Imptrad



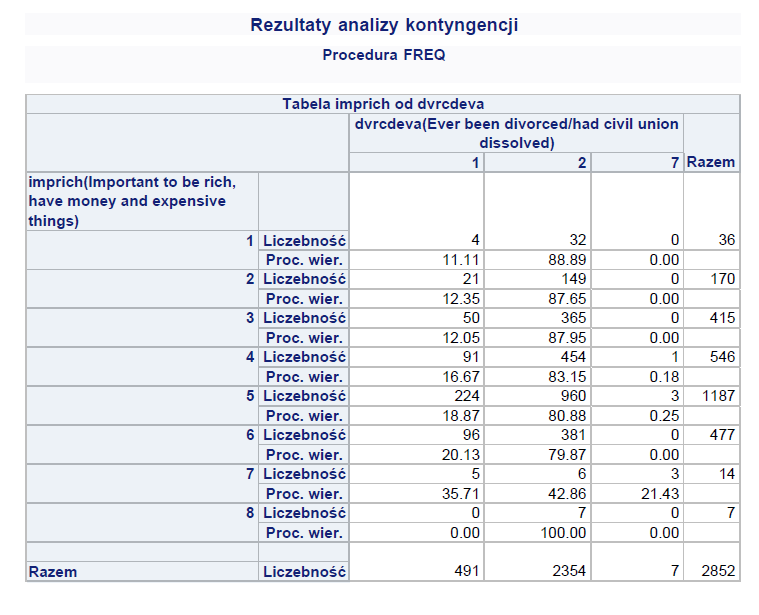
Na podstawie analizy tablic kontyngencji widzimy, że zmienna Imptrad nie ma wpływu na zmienną dvrcdeva - poza odpowiedzią “0” dla każdej z wartości zmiennej Imptrad stosunek osób rozwiedzionych i nierozwiedzionych jest podobny, i wynosi około 17% dla osób po rozwodzie do 83% w przeciwnym przypadku. Nie mamy zatem podstaw by zagregować daną zmienną i możemy jedynie pominąć opcję 7 i 8 jako missing’i.

#### Pplfair



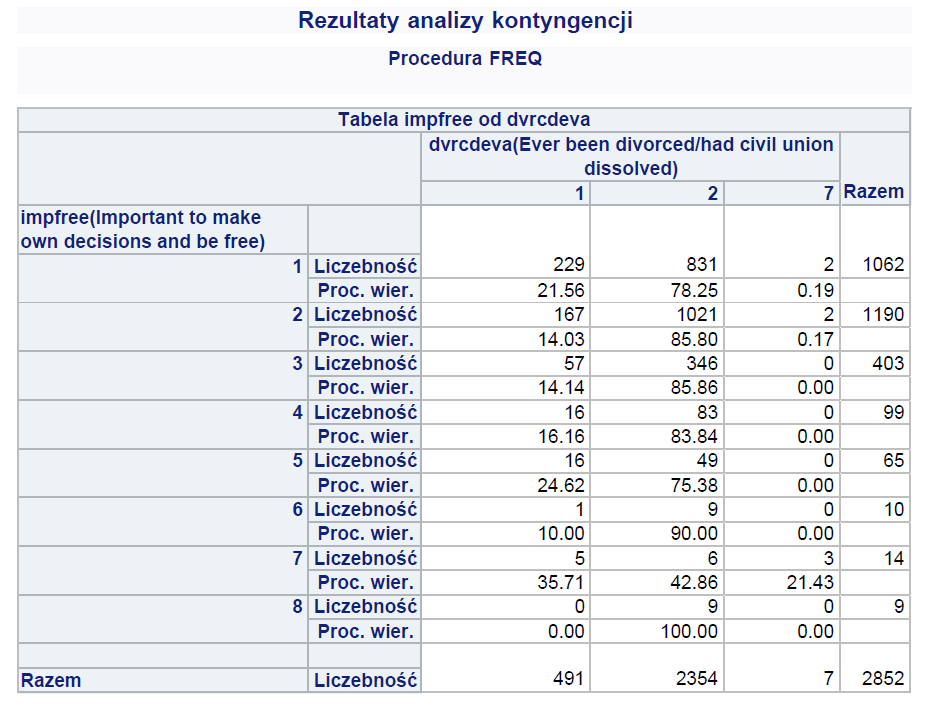
Na podstawie analizy tablic kontyngencji widzimy, że zmienna Pplfair może zostać zakwalifikowana jako nieufny (“cagey”), gdy przyjmuje wartości od 0 do 2 i ufny (“trustful”) dla wartości 3-10. Wartość 88 pomijamy jako brak odpowiedzi.

#### Imprich



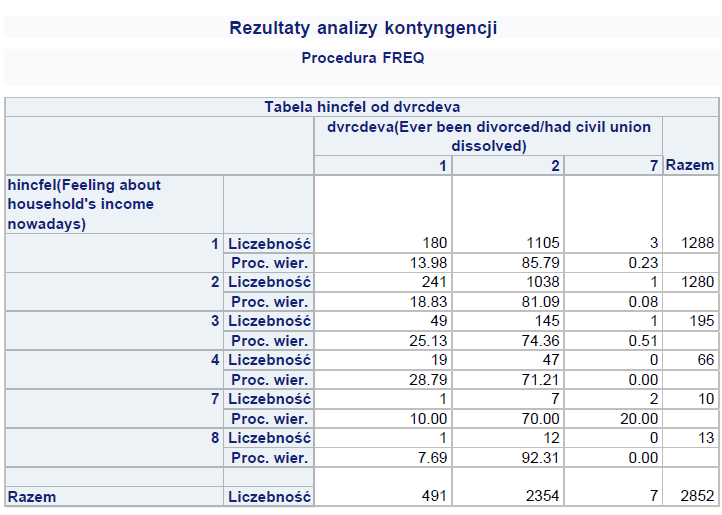
W przypadku zmiennej imprich na początku wyeliminujemy wartość 7 i 8 - oznaczają one odpowiednio “odmowę” i “brak odpowiedzi”, a do tego są mało liczne. Następnie pozostałe zmienne podzielimy na dwie kategorie - “yes” i “no”, gdzie za “yes” przyjmiemy osoby, dla których bogactwo jest istotne (1-3),   
a za “no” tych, których pieniądze nie interesują (odpowiedzi 4-6)

#### Impfree



Podobnie jak przy zmiennej imprich, możemy wykluczyć odpowiedzi 7 i 8, ponieważ oznaczają “odmowę” i “brak odpowiedzi”, a do tego są mało liczne. Tak samo możemy potraktować zmienną 6, ponieważ również zawiera małą próbkę. Patrząc na pozostałe zmienne, możemy połączyć odpowiedzi 2-4 w jedną kategorię, “medium”, 1 zakwalifikujemy jako ‘high”, a 5 jako “low”.

#### Hincfel



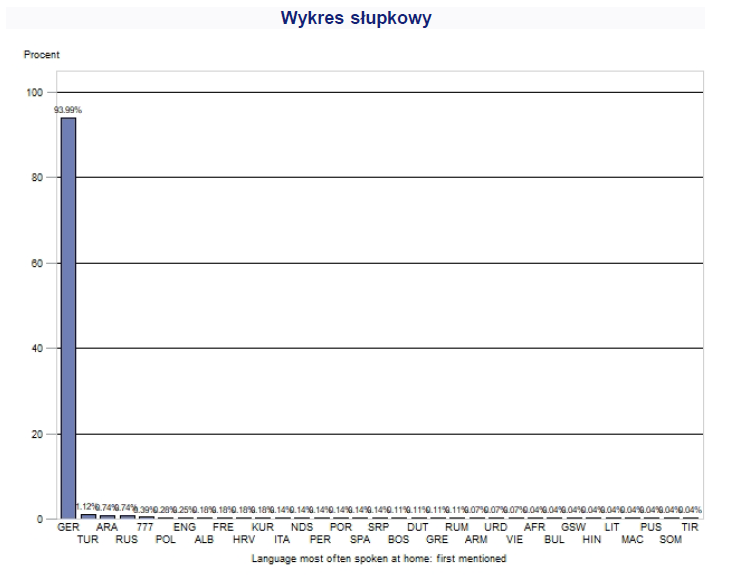
Możemy zauważyć że w przypadku poczucia stabilności, w związku z obecnym stanem budżetu domowego, osoby zadowolone ze swojej sytuacji materialnej (1-2) decydują się na rozwód o średnio 10% rzadziej, niż osoby niezadowolone ze swojego bytu. A zatem po odfiltrowaniu missing’ów (7 i 8) dostaniemy dwie kategorie - “yes”, czyli 1 i 2, oraz “no” - 3,4.

#### Region

Jedną z naszych tez badawczych jest sprawdzenie, czy miejsce zamieszkania *ceteris paribus* ma wpływ na wzrost lub spadek liczby rozwodów. W szczególności interesuje nas, czy osoby zamieszkujące tereny dawnego RFN rozwodzą się częściej, niż zamieszkujący obszar dawnego NRD. W związku z tym musimy skategoryzować wartości z kolumny Region.

|  |  |  |
| --- | --- | --- |
| Kod Regionu | Nazwa Regionu | Kategoria |
| DE1 | Baden-Württemberg | RFN |
| DE2 | Bayern | RFN |
| DE3 | Berlin | NRD |
| DE4 | Brandenburg | NRD |
| DE5 | Bremen | RFN |
| DE6 | Hamburg | RFN |
| DE7 | Hessen | RFN |
| DE8 | Mecklenburg-Vorpommern | NRD |
| DE9 | Niedersachsen | RFN |
| DEA | Nordrhein-Westfalen | RFN |
| DEB | Rheinland-Pfalz | RFN |
| DEC | Saarland | RFN |
| DED | Sachsen | NRD |
| DEE | Sachsen-Anhalt | NRD |
| DEF | Schleswig-Holstein | RFN |
| DEG | Thüringen | NRD |

#### Lnghom1



Zauważmy że zmienna lnghom1, oznaczająca pierwszy język, jakim posługuje się w gospodarstwie domowym respondenta, w 94% przypadków przyjmuje wartość “GER”. Oznacza to, że w tych domach pierwszym językiem jest Niemiecki. Pozostałe najpopularniejsze to Turecki (TUR, 1.12%), Arabski (ARA)   
i Rosyjski (RUS) - po 0.74%. W związku z tym postanowiliśmy podzielić wartości na dwie kategorie - GER, jeśli osoba mówi w domu w języku Niemieckim, oraz OTH, jako Other, czyli wszystkie inne wartości.

#### Dodatkowe uwagi

Dla wszystkich skategoryzowanych zmiennych do skrótu nazwy dodaliśmy przyrostek “cat”, na przykład “Regioncat”. W ten sposób rozróżnimy zmienne poddane kategoryzacji od tych, w których poprzestaliśmy na odfiltrowaniu missingów. Takimi zmiennymi są: chldhhe, Imptrad, Uemp3m, Hlthhmp, Rlgatnd.

# 9. Analiza

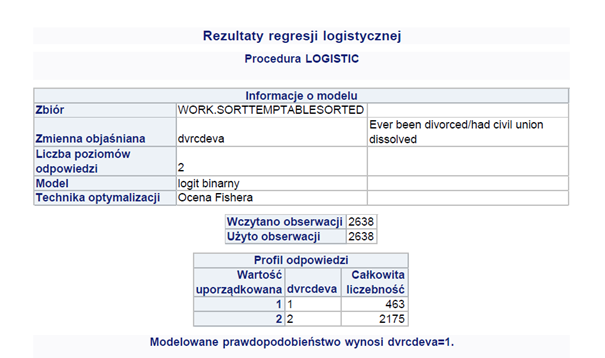
## 9.1 Światopogląd

Zaproponujemy teraz dwa zbiory modeli. Pierwszym jest zbiór zmiennych, na których kształt ma wpływ osoba badana - poglądy, podjęte przez niego życiowe decyzje (jak posiadanie dzieci) itp.

Przypomnijmy, że naszą zmienną objaśnianą jest **dvrcdeva**, czyli czy dana osoba rozwiodła się lub rozwiązała formalny związek partnerski.

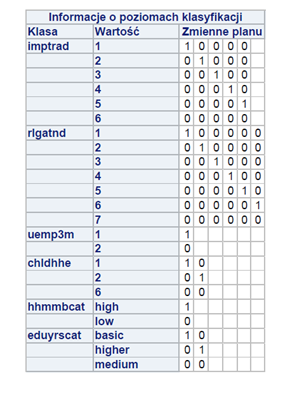
Zmienne objaśniające wybrane przez nas do modelu to:

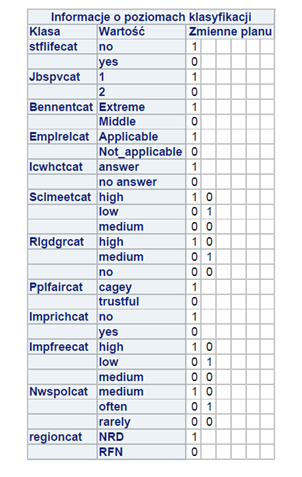
* **Regioncat** - skategoryzowana zmienna region. Oznacza ona miejsce zamieszkania, a to może zostać zmienione.
* **Hhmmbcat** - skategoryzowana liczba mieszkańców gospodarstwa domowego. Jeśli osoba badana założyła rodzinę bądź mieszka sama, ma ona wpływ na jej wartość.
* **Eduyrscat** - skategoryzowana liczba lat nauki - pomimo obowiązku szkolnego, który obejmuje dzieci i młodzież w wieku od 6 do 18 lat, po uzyskaniu pełnoletności to ankietowany decyduje czy będzie kontynuować naukę.
* **Stflifecat** - skategoryzowana zmienna “satysfakcja z życia”, opisująca jego opinię na temat tego, jak wiedzie się jemu lub jej w życiu.
* **Jbspvcat** - skategoryzowana zmienna która opisuje, czy badany zajmuje w pracy stanowisko, które wymaga zarządzania innymi pracownikami.
* **Bennentcat**  - dotycząca opinii, czy “wiele osób niezasłużenie otrzymuje pomoc od państwa”. Wartości pogrupowaliśmy na dwie kategorie - “poglądy skrajne” oraz “umiarkowane”.
* **Emplrelcat** - dotyczące typu/kategorii zatrudnienia
* **Icwhctcat** - Posiadanie podstawowej lub stałej liczby godzin. W zależności od typu umowy może pracować na etapie lub być freelancerem..
* **Sclmeetcat** - skategoryzowany stopień w jakim dana osoba lubi spotykać się ze znajomymi - wpływ na tą zmienną może mieć temperament lub charakter, a ten, jak wiadomo, można kształtować.
* **Rlgdgrcat** - skategoryzowany stopień, w jakim badany ocenia swoją religijność.
* **Pplfaircat** - jak bardzo dana osoba ufa ludziom
* **Imprichcat** - Opinia respondenta na temat bogactwa, posiadania pieniędzy i drogich rzeczy.
* **Impfreecat** - Opinia dotycząca poczucia własnej wolności i niezależności.
* **Nwspolcat** - Ile czasu dana osoba spędza na śledzeniu newsów.
* **Chldhhe** - Obecność dziecka w gospodarstwie domowym. Badany ma na to wpływ, jeśli sam posiada potomstwo.
* **Rlgatnd** - częstość uczestniczenia w praktykach religijnych z pominięciem specjalnych okazji - podobna do Rlgdgrcat.
* **Uemp3m** - czy badana osoba była kiedykolwiek bezrobotna lub poszukiwała pracy równy lub dłuższy niż 3 miesięcy. Na zmienną tą można było wpłynąć swoim poglądem na pracę (czy ktoś jest lekkoduchem czy też karierowiczem).
* **Imptrad** - opinia dotycząca przywiązania do tradycji



Źródło: wydruk z programu SAS

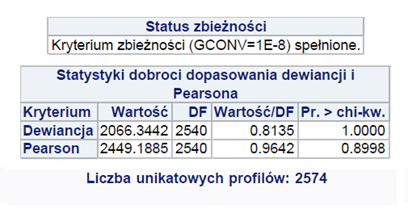
Powyżej został wygenerowany raport opisujący model regresji logistycznej. Do opisania modelu użyto 2638 obserwacji. 463 respondentów odpowiedziało pozytywnie na pytanie „Czy kiedykolwiek rozwiedli się lub czy zakończyli prawnie swój związek?„ i 2175 osób odpowiedziało przecząco.





Źródło: wydruk z programu SAS

W tabelach powyżej przedstawiono zmienne objaśniające użyte do modelu.



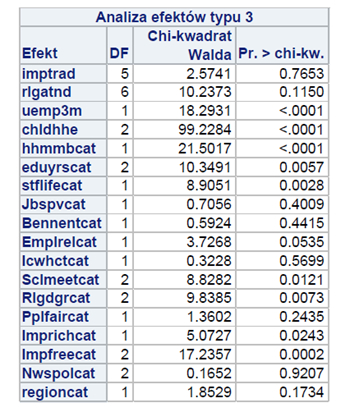
Źródło: wydruk z programu SAS

Zostało spełnione kryterium zbieżności. Liczba unikatowych profilów 2574 z 2638, świadczy o bardzo słabej agregacji.



Źródło: wydruk z programu SAS

Współczynnik pseudo R-kwadrat Cox’a i Snella wynosi 12.3 % oraz skorygowane R-kwadrat Negelkerke wynosi 20.4%. Świadczy to o niskim dopasowaniu danych do opracowanego modelu.



Źródło: wydruk z programu SAS

Przy przyjętym poziomie istotności 0,01 odrzucono hipotezę o nieistotności następujących zmiennych: *uemp3m, chldhhe, hhmmbcat, eduyrscat, stflifecat, rlgdrcat, impfreecat*. Pozostałe zmienne okazały się być nieistotne statystycznie. Prawdopodobnie trzeba będzie rozpatrzyć efekty drugiego rzędu.



Źródło: wydruk z programu SAS

Wartość p value = 0.22 jest wyższa od poziomu istotności 0.05. Nie świadczy ani za, ani przeciwko hipotezie zerowej. Wartość taka może oznaczać, że badanie miało zbyt niską moc statystyczną.

## 9.2 Środowisko

Drugim zbiorem są zmienne na które badana osoba mogła nie mieć wpływu - są to zmienne dotyczące środowiska w jakim osoba się wychowała.

zmienna objaśniana Dvrcdeva

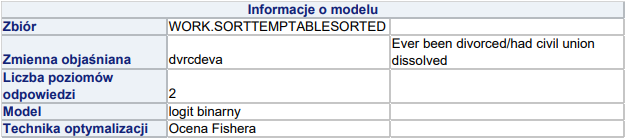
Zmienne objaśniające:

* **Ageacat** - grupa wiekowa do której należy respondent
* **Lnghom1cat** - język najczęściej używany w domu - w tym także rodzinnym, zwłaszcza   
  w przypadku mieszkania z rodzicami.
* **Eduyrscat** - Zmienna która cześciowo wynika z włąsnego poglądu na naukę, ale także ze względu na przepisy prawa (obowiązek szkolny do 18 roku życia)
* **Regioncat** - Chociaż miejsce zamieszkania można zmienić, część osób może mieszkać od urodzenia w tym samym mieście, ze względu np. Na studia i późniejszą pracę.
* **Anctry1cat** - pochodzenie etniczne danej osoby.
* hhmmbcat - skategoryzowana liczba mieszkańców gospodarstwa domowego. Wartość ta może być niezależna od ankietowanego w przypadku, gdy nadal mieszka on w domu rodzinnym
* **Rlgdgrcat** - pomimo że poglądy religijne mogą ulec zmianie, to jednak często jest to ważny element wychowania. Przykładem może być przedstawiciel mniejszości muzułmańskiej   
  w Niemczech, którego wiara wyniesiona z domu ma znaczący wpływ na światopogląd.
* **Rlgatnd** - analogicznie jak w przypadku Rlgdgrcat
* **Chldhhe** - Obecność dziecka w gospodarstwie domowym. Musimy uwzględnić sytuacje, gdy badany nie ma na to wpływu, bo np. mieszka w domu wielopokoleniowym, wraz z rodziną.
* **Imptrad** - Jak ważne jest przywiązanie do tradycji. Choć jest to opinia i można ją zmienić, to jednak osoby przywiązujące do niej wagę będą postępować zgodnie ze światopoglądem wpojonym przez środowisko, w którym się wychowały.
* **Hlthhmp** - utrudnienie w życiu codziennymi wynikające z choroby, niepełnosprawności, choroby umysłowej, w tym również tych wrodzonych.
* **Emplrelcat** - dotyczące typu/kategorii zatrudnienia. Umieściliśmy ją w modelu opisującym środowisko ze względu na to, że badany może pracować w firmie rodzinnej, a co za tym idzie, mógł nie mieć możliwości ścieżki kariery, lecz musiał kontynuować rodzinną tradycję czy też przejąć rodzinny biznes.

Do modelu dodaliśmy także zmienną **Stflifecat** - w przypadku gdy drugi model okaże się lepszy, potrzebujemy oszacować także tą zmienną, by uzyskać odpowiedź na drugą hipotezę badawczą.

Otrzymane wyniki - Środowisko

Model binarny ze wszystkimi zmiennymi objaśniającymi

Źródło: wydruk z programu SAS



Źródło: wydruk z programu SAS

W związku z tym, że zmienne wczytane do programu nie posiadają braków, model opisany w projekcie używa wszystkich wczytanych obserwacji, tj. 2638 obserwacji.

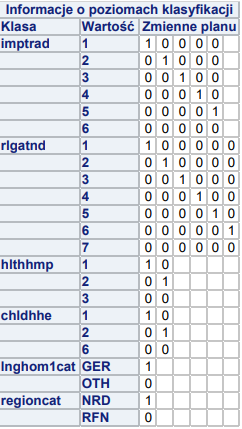
Profil odpowiedzi:

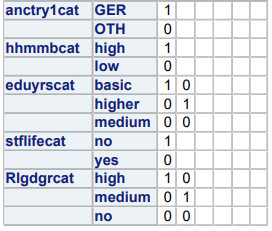


Źródło: wydruk z programu SAS

Powyższe dane uzyskane z modelu wskazują, że spośród badanych respondentów 463 osoby odpowiedziały twierdząco na pytanie, czy kiedykolwiek rozwiedli się lub czy zakończyli prawnie swój związek, natomiast 2175 osób nigdy nie miało takiego doświadczenia.

Kodowanie referencyjne zmiennych:





Źródło: wydruk z programu SAS

Powyższa tabela prezentuje sposób kodowania zmiennych referencyjnych wykorzystanych w modelu.

Status zbieżności:



Źródło: wydruk z programu SAS

Statystyki dopasowania dewiancji i Pearsona:



Źródło: wydruk z programu SAS

Powyższa tabela przedstawia wartości statystyki dewiancji i Pearsona, które obliczono dla badanego modelu w celu sprawdzenia, czy model jest dobrze dopasowany do danych. Wyliczona liczba unikatowych profilów wynosi 1517, co znacznie różni się od liczby obserwacji wykorzystanej do przygotowania modelu (2638). Na podstawie tych obserwacji, można wysnuć wniosek, że dane nie zostały dostatecznie dobrze zagregowane, w związku z czym powyższe statystyki nie dostarczają wystarczających informacji do ocenienia dobroci dopasowania modelu.

Statystyki dopasowania:



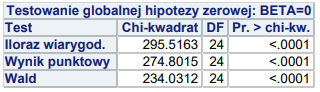
Źródło: wydruk z programu SAS

Statystyka R kwadrat:



Źródło: wydruk z programu SAS

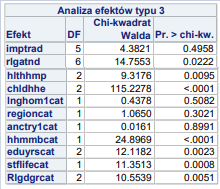
Testowanie globalnej hipotezy zerowej BETA=0



Źródło: wydruk z programu SAS

Kolejnym testem potrzebnym do sprawdzenia modelu jest przetestowanie hipotezy zerowej , która zakłada, że wszystkie współczynniki w modelu są równe 0 i co za tym idzie nie są istotne statystycznie. Jak można zauważyć powyżej, w przypadku każdego z testów wartość p-value jest mniejsza niż 0,05, co pozwala na odrzucenie hipotezy zerowej na rzecz hipotezy alternatywnej, która mówi, że przynajmniej jeden ze współczynników w modelu jest istotny statystycznie.

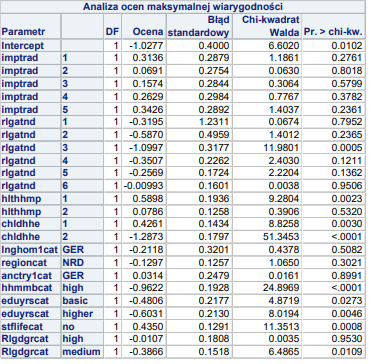
Analiza efektów typu 3:



Źródło: wydruk z programu SAS

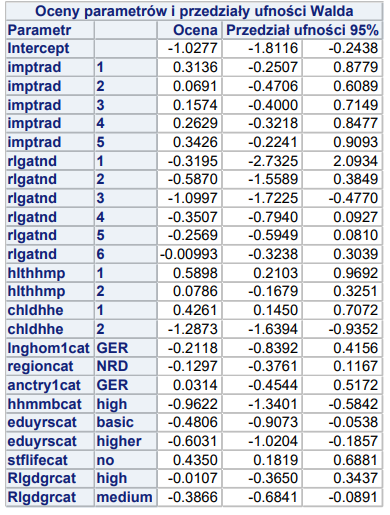
Powyższa analiza ma za zadanie wykazanie, które z wykorzystanych zmiennych są istotne statystycznie. Warunkiem niezbędnym do tego, aby zmienna była istotna jest uzyskanie dla niej wartości mniejszej niż 0,05 dla analizy chi-kwadrat. Spośród zmiennych wybranych do analizy, cztery nie spełniają tego założenia (imptrad, lnghom1cat, regioncat, anctry1cat). Powyższe wyniki mogą stanowić podstawę wykluczenia tych zmiennych z modelu.

Analiza ocen maksymalnej wiarygodności:



Źródło: wydruk z programu SAS

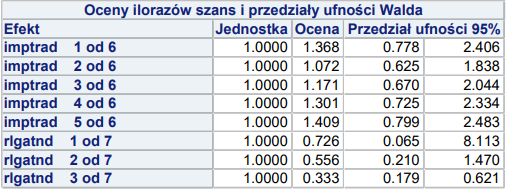
Oceny parametrów i przedziały ufności Walda:

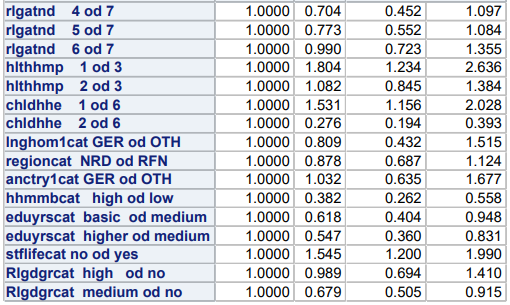


Źródło: wydruk z programu SAS

Powyższa tabel przedstawia analizę przedziałów ufności Walda dla ilorazów szans. Zasadą jest, że dana zmienna jest istotna na wybranym poziomie α, jeżeli przedział ufności dla exp(β) z współczynnikiem 1-α zawiera liczbę 1. Jak można zauważyć, w modelu występuje 95% przedział ufności dla zmiennych (α=0,05). W przypadku gdy przedział zawiera 0, dana ocena jest nieistotna statystycznie. Zmienne, które są nieistotne statystycznie w podanym modelu to: wszystkie zmienne imptrad, zmienne rlgatnd (bez kategorii 3), hlthhmp (kategoria 2), lnghom1cat, regioncat, anctry1cat i obie zmienne rlgdgrcat.

Oceny ilorazów szans i przedziału ufności Walda

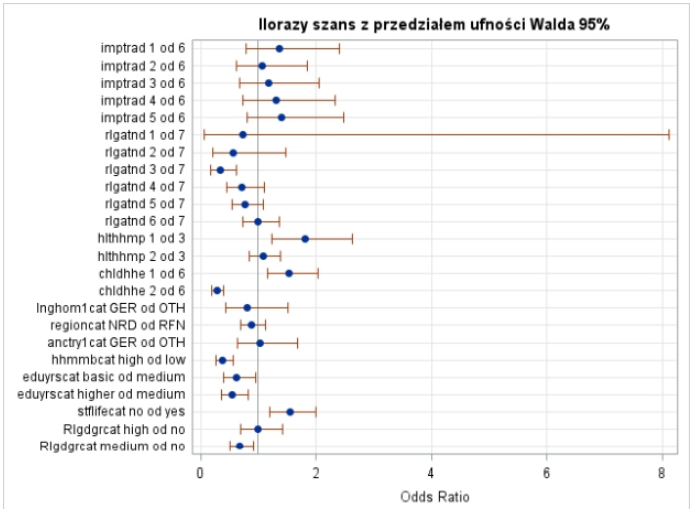




Źródło: wydruk z programu SAS

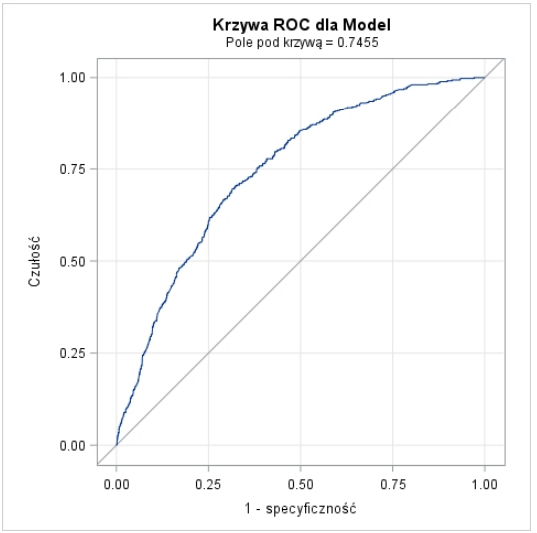
W powyżej tabeli zostały przedstawione oceny ilorazów szans i przedziałów ufności Walda. W tym przypadku można mówić o nieistotności, jeśli przedział zawiera liczbę 1. Wyniki zamieszczone w tabeli powyżej prezentują występowanie istotności dla zmiennych: rlfatnd (3 od 7), hlthhmp (1 od 3), chldhhe (obie kategorie), hhmmbcat (basic od medium), eduyrscat (obie kategorie), stflifcat i Rlgdgrcat (medium od no). Te same informacje mogą być odczytane na poniższej graficznej reprezentacji tabeli. Pionowa linia dla Odds Ratio = 1 umożliwia interpretację, czy dla danej zmiennej ocena jest istotna statystycznie.

Ilorazy szans z przedziałem ufności wiarygodności profilu 95%



Źródło: wydruk z programu SAS

Krzywa ROC dla modelu



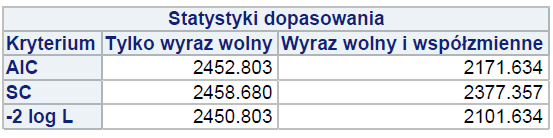
Źródło: wydruk z programu SAS

Powyższy wykres prezentuje krzywą ROC obliczoną dla modelu. Krzywa ROC jest to graficzna reprezentacja efektywności modelu predykcyjnego, która powstaje poprzez wykreślenie charakterystyki jakościowej klasyfikatorów binarnych powstałych z modelu przy zastosowaniu wielu różnych punktów odcięcia. Krzywa ta reprezentuje zależność między prawdopodobieństwem prawidłowego sklasyfikowania jednostki, która doświadczyła zdarzenia, od prawdopodobieństwa niepoprawnego zaklasyfikowania obserwacji, dla której zdarzenie nie zaszło. Pole pod wykresem krzywej ROC zwane pod nazwą AUC (Area Under Curve) jest najczęściej interpretowaną częścią tego wykresu. Należy zauważyć, że wartości tej statystyki należą do przedziału <0,1>. Oczekiwaną wartością parametru AUC jest wartość jak najbardziej zbliżona do 1. Dla analizowanego modelu pole pod krzywą wynosi 0,7455, co oznacza, że model jest dopasowany w 74,55% do danych. Jednakże, z uwagi na to, że ostatecznym celem projektu jest jak najlepsza ocena efektów, ta interpretacja nie została uznana za najbardziej istotną.

## 9.3 Porównanie obu modeli

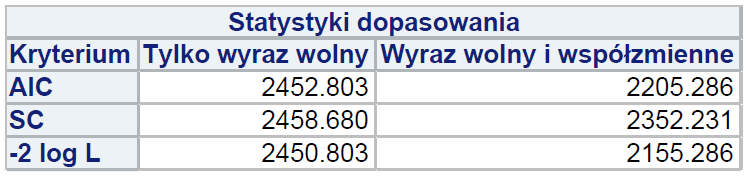
Po tym jak rozpatrzyliśmy dwa różne modele musimy podjąć decyzje, który z nich wybrać do objaśnienia zmiennej dvrcdeva (rozwody). W tym celu wykorzystamy wartości kryterium informacyjnego Akaike’a oraz kryterium Schwartz’a by uzyskać model, który charakteryzuje się największą wiarygodnością.

Model zmiennych światopoglądu:



Źródło: wydruk z programu SAS

Model zmiennych środowiska:



Źródło: wydruk z programu SAS

Wartości kryterium Schwatrz’a i Akaike’a maleją wraz ze wzrostem wiarygodności i rosną wraz z liczbą szacowanych parametrów. Zatem im niższa wartość tym lepszy udało nam się stworzyć model. Różnica między jednym i drugim kryterium jest taka, że kryterium AIC w przeciwieństwie do SC jest mniej restrykcyjne i nie karze nas za dodatkowe zmienne w modelu.

Ponieważ na tym etapie nie możemy jeszcze rozstrzygnąć, który z modeli jest lepszy, postanowiliśmy przeprowadzić analizę obu modeli raz jeszcze, tym razem jednak pozwolimy SASowi na “odsianie” nieistotnych zmiennych.

Do wyboru mamy trzy metody:

* **Wybór następnych** -Metoda Wybór następnych początkowo nie zawiera zmiennych w modelu, dopiero dodaje zmienne poprzez porównywanie wartości p dla statystyk F z poziomem istotności, określonym w polu tekstowym Wstawienie do modelu.
* **Eliminacja poprzednich** - Metoda Eliminacja poprzednich na początku uwzględnia wszystkie zmienne ilościowe, a następnie usuwa zmienne do momentu, aż wszystkie pozostałe zmienne utworzą statystyki F znaczące na poziomie istotności, określonym w polu tekstowym Pozostanie w modelu.
* **Metoda krokowa** - Metoda krokowa jest modyfikacją metody wyboru następnych, różniącą się tym, że zmienne znajdujące się już w modelu nie muszą w nim pozostać. Proces krokowy kończy się, gdy ani jedna ze zmiennych pozostających poza modelem nie ma statystyki F istotnej na poziomie istotności wystarczającym do wstawienia do modelu oraz gdy wszystkie zmienne pozostające w modelu są znaczące na poziomie istotności wystarczającym do pozostawienia   
  w modelu lub jeśli zmienna dodana do modelu okaże się zmienną już wcześniej usuniętą z modelu.

W tej analizie wybraliśmy metodę numer 3, a więc zastosowanie metody krokowej.

Porównamy teraz raz jeszcze wartości statystyk dopasowania, już po odfiltrowaniu zmiennych nieistotnych:

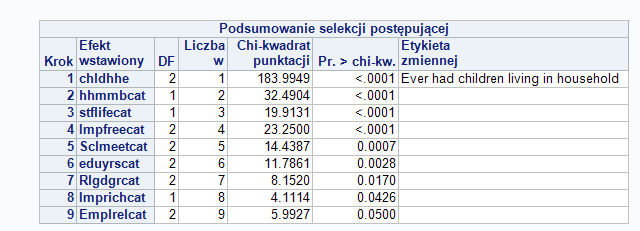
|  |  |  |
| --- | --- | --- |
| Model | AIC | SC |
| Światopogląd | 2171.747 | 2265.792 |
| Środowisko | 2226.049 | 2325.971 |

A zatem ostatecznie lepszym modelem jest model badający zmienne dotyczące światopoglądu, który nie zawiera zmiennych nieistotnych.

# 

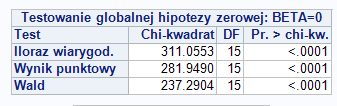
# 10. Analiza ostatecznego modelu regresji

Po wykonaniu metody krokowej w naszym modelu pozostały następujące zmienne:



Źródło: wydruk z programu SAS

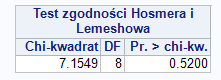
Jak widzimy, każda z nich jest istotna co najmniej na poziomie 95%.



Źródło: wydruk z programu SAS

Odrzucamy hipotezę zerową na korzyść hipotezy alternatywnej (“przynajmniej jeden z parametrów jest istotnie różny od 0”) także w tym modelu, bowiem wszystkie wartości p-value są mniejsze od domyślnego poziomu istotności, który wynosi α=0.05.

## 10.1 Wynik testu zgodności Hosmera i Lemeshowa



Źródło: wydruk z programu SAS

Wartość p value = 0.52 jest wyższa od poziomu istotności 0.05. Nie świadczy ani za, ani przeciwko hipotezie zerowej. Wartość taka może oznaczać, że badanie miało zbyt niską moc statystyczną.

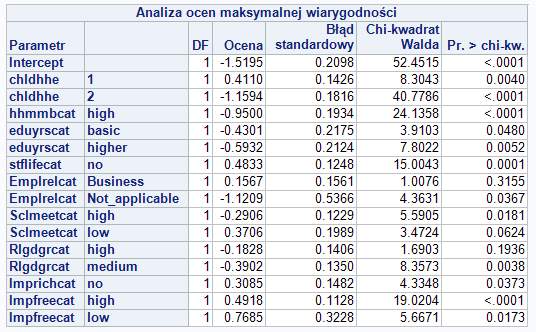
## 10.2 Statystyka dobroci dopasowania dewiancji i Pearsona

## 

Źródło: wydruk z programu SAS

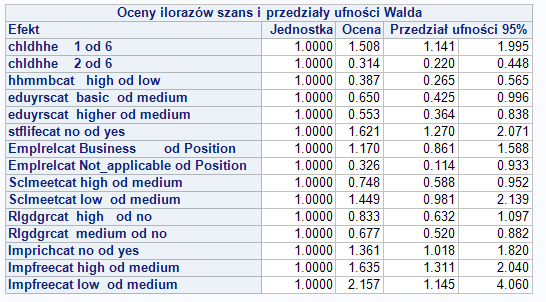
Zarówno dla statystyk dewiancji jak i Pearsona wartość krytyczna prawdopodobieństwa jest wyższa niż poziom istotności 𝛼 = 0.05. W rezultacie nie mamy podstaw, aby odrzucić hipotezę zerową mówiącą   
o dobrym dopasowaniu modelu do danych.

## 10.3 Interpretacja ilorazów szans i weryfikacja postawionych hipotez

.

Źródło: wydruk z programu SAS

Powyższa tabela zawiera oszacowania parametrów modelu wraz z błędem oszacowań, statystyk Chi-kwadrat Walda wraz z krytycznymi poziomami istotności oraz wartości ilorazu szans.



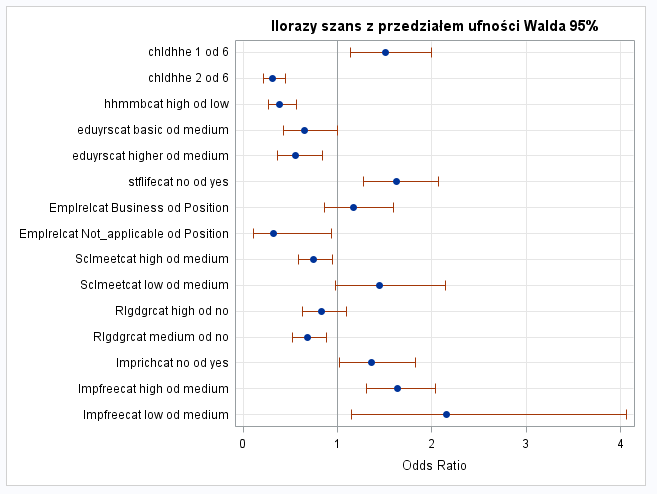
Źródło: wydruk z programu SAS

Powyższa tabela posłuży nam do interpretacji wyników.

* Osoby, w których gospodarstwie domowym znajduje się dziecko zwiększa ryzyko rozwodu   
  w stosunku do osób, których pytanie to nie dotyczy o 50.8% (chldhhe 1 od 6). Natomiast osoby które nie mają dziecka w gospodarstwie domowym mają o 68.6% niższą szansę rozwodu (chldhhe 2 od 6).
* Osoby które posiadają dużą liczbę współlokatorów mają o 61.3% mniejsze szanse na rozwód   
  w stosunku do osób, u których w gospodarstwie domowym mieszka od 1 do 3 osób.
* Osoby, które deklarowały liczbę lat edukacji mieszczącą się w kategorii “niska” mają o 35% niższe szanse na rozwód w stosunku do osób, które deklarowały “przeciętną” liczbę lat edukacji. Natomiast osoby, które deklarowały liczbę lat edukacji mieszczącą się w kategorii “wysoka” mają o 44.7% niższe szanse na rozwód w stosunku do osób, które deklarowały “przeciętną” liczbę lat edukacji.
* Osoby, które deklarowały że nie są zadowolone z życia, mają o 62.1% większe szanse   
  na rozwód od osób zadowolonych z życia.
* Osoby, które są przedsiębiorcami mają o 17% wyższe szanse na rozwód w stosunku do osób zatrudnionych na etacie. Osoby, których pytanie to nie dotyczy mają o 67.4% niższe szanse   
  na rozwód w stosunku do osób zatrudnionych na etacie.
* Osoby, które bardzo lubią spotkania ze znajomymi mają o 25.2% mniejsze szanse na rozwód   
  w stosunku do osób, które średnio lubią spotkania ze znajomymi. Osoby, które nie lubią spotkań ze znajomymi mają o 44.9% wyższe szanse na rozwód w stosunku do osób, które średnio lubią spotkania ze znajomymi.
* Osoby które są bardzo religijne mają o 16.7% niższe szanse na rozwód od osób deklarujących się jako niereligijne. Osoby które są umiarkowanie religijne mają o 32.3% niższe szanse   
  na rozwód od osób deklarujących się jako niereligijne.
* Osoby które deklarują, że bogactwo jest dla nich nieistotne mają o 36.1% wyższe szanse   
  na rozwód w stosunku do osób, dla których pieniądze są istotne.
* Osoby które deklarują, że ich wolność osobista jest dla nich bardzo istotna mają o 63.5% wyższe szanse na rozwód w stosunku do osób, dla których wolność osobista jest przeciętnie istotna. Osoby które deklarują, że ich wolność osobista jest dla nich mało istotna mają o 115.7% wyższe szanse na rozwód w stosunku do osób, dla których wolność osobista jest przeciętnie istotna.

Należy wspomnieć, że wszystkie zmiany są *ceteris paribus*, co na polski tłumaczy się zwykle jako „przy pozostałych warunkach równych” lub „przy tych samych okolicznościach”.

Graficzną prezentację ilorazów szans wraz z przedziałem ufności Walda przedstawia poniższy wykres. Co więcej, daje on możliwość szybkiego wykrycia kategorii, które nie są istotne.



Źródło: wydruk z programu SAS

## 10.4 Wyniki a postawione hipotezy badawcze

Przypomnijmy, że na początku naszych rozważań postawiliśmy sobie dwie hipotezy badawcze. Teraz, gdy stworzyliśmy model, możemy spróbować udzielić na nie odpowiedzi.

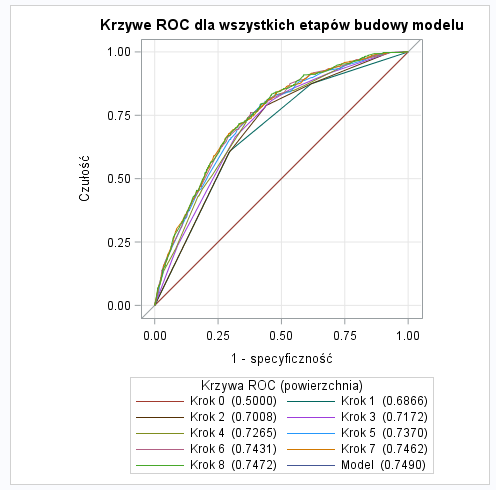
* **Czy osoby pochodzące z obszaru dawnej Republiki Federalnej Niemiec (RFN) rozwodzą się częściej, niż ludzie zamieszkujący dawną Niemiecką Republikę Demokratyczną (NRD)?**

Pomimo uwzględnienia zmiennej “Region” w modelu, w trakcie badań okazało się, że pochodzenie geograficzne nie ma znaczenia - wartość ta jest bowiem statystycznie nieistotna i została odrzucona na etapie zastosowania metody krokowej.

* **Czy osoby zadowolone z życia rozwodzą się rzadziej, niż niezadowolone?**

Na podstawie badań możemy stwierdzić, że osoby, które deklarowały że nie są zadowolone z życia, mają o 62.1% większe szanse na rozwód od osób zadowolonych z życia, a zatem osobny zadowolone rozwodzą się znacznie rzadziej. Przy poziomie istotności 𝛼 = 0.05 wartość ta waha się pomiędzy 27% a 107.1%.

## 10.5 Krzywa ROC

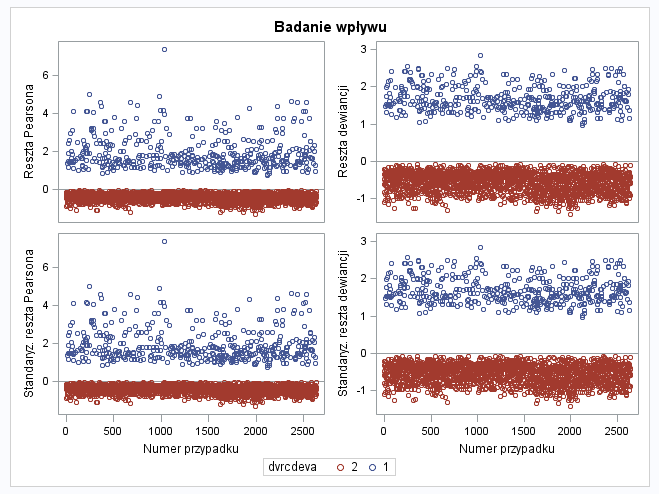


Źródło: wydruk z programu SAS

Ostatnim punktem naszych rozważań jest przedstawienie krzywej ROC. Przypomnijmy, że krzywa ROC to jeden ze sposobów wizualizacji jakości klasyfikacji [32], pokazujący zależności wskaźników TPR (True Positive Rate) oraz FPR (False Positive Rate). W naszym przypadku, końcowy model ma 74.9% skuteczności. Idealny model miałby 100% skuteczności, a punkt przegięcia znajdowałby się w punkcie (X,Y) = (0,1), czyli swoistość i czułość równe byłby 1. W naszym przypadku nie zależy nam jednak na budowie modelu predykcyjnego, więc nie musimy przywiązywać do tej wartości zbytniej wagi.

## 10.6 Badanie wpływu

Na koniec przedstawimy jeszcze wykres badania wpływu. Zasada działania jest następująca - usuwamy daną zmienną z modelu. Następnie sprawdzamy, jaki wpływ miało to na nasz model.

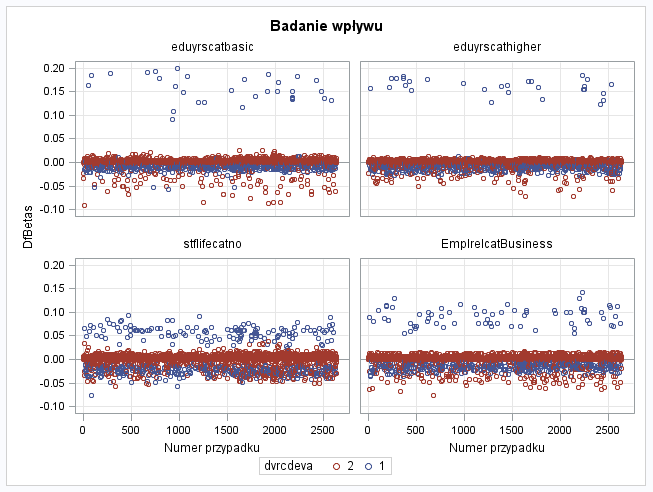


Źródło: wydruk z programu SAS

W powyższym badaniu sprawdzamy, jaki wpływ ma dana obserwacja kolejno na:

* W lewym górnym rogu - reszty Pearsona - mówią nam o wkładzie obserwacji do statystyki reszt Pearsona.
* W prawym górnym rogu - reszty dewiacji - mówią nam o zależności poszczególnych obserwacji wyników i reszt dewiacji
* W lewym dolnym rogu - standaryzowane reszty Pearsona - mówią nam o wkładzie obserwacji do standaryzowanej statystyki reszt Pearsona.
* W prawym górnym rogu - standaryzowane reszty dewiacji - mówią nam o zależności poszczególnych obserwacji i wyników standaryzowanych reszt dewiacji

Analogicznie prezentuje się analiza wpływu obserwacji wobec innych miar i statystyk, jak chociażby różnica usunięcia dewiacji, chi-kwadrat, albo usunięcie niektórych kategorii - jak choćby poszczególne wartości z kategorii eduyrscat.



Źródło: wydruk z programu SAS

# 

# 11. Podsumowanie

W niniejszym raporcie podjęto próbę wyjaśnienia jakie czynniki mają wpływ podejmowanie decyzji o rozwodzie. W tym celu zbudowano dwa modele regresji logistycznej binarnej, których zmienna objasnianą była odpowiedź na pytanie, czy respondent zdecyduje się na rozstanie.

Początkowo przeanalizowaliśmy specjalistyczną literaturę, by dowiedzieć się, jakie czynniki mają wpływ na szczęśliwe pożycie małżeńskie. Następnie, bazując na wiedzy eksperckiej, wyselekcjonowaliśmy 24 zmienne spośród 533 możliwych, by skonstruować jak najlepszy model.

Kolejnym etapem było przygotowanie uzyskanych danych do dalszej pracy. Każdą zmienną rozpatrzyliśmy pod kątem liczby obserwacji, średniej, skośności czy też kurtozy. Wartości podzieliliśmy na grupy, a tam, gdzie nie było to konieczne, zadbaliśmy o usunięcie bezwartościowych rekordów.

Gdy mieliśmy już przygotowany zbiór danych, mogliśmy zacząć przygotowywać odpowiedni model. W tym miejscu zadaliśmy sobie pytanie - jaka grupa zmiennych ma większy wpływ - opinie, przekonania respondenta, czynniki które pojawiły się wraz kolejnymi latami jego życia (jak choćby praca czy związek), czy też może te, które wynikały z jego pochodzenia - czynniki z którymi się urodził, lub które wyniósł z domu. W tym miejscu wyklarowały się nam także dwie hipotezy badawcze - czy fakt, że dana osoba mieszka na terenie dawnego RFN i NRD ma znaczenie, oraz czy częściej rozwiedzie się osoba szczęśliwa, czy też niezadowolona z życia.

Budowa modelu wraz ze wszystkimi zmiennymi pokazała, że niektóre zmienne ze statystycznego punktu widzenia są nieistotne. Dzięki temu wyselekcjonowano tylko te zmienne, które miały znaczący wpływ na główne pytanie. Do zmiennych nieistotnych statystycznie należały zmienne objaśniające takie jak język używany w domu lub region zamieszkania respondenta. Natomiast do zmiennych istotnie statystycznie zakwalifikowaliśmy czynniki takie jak częstość uczestnictwa obrzędach religijnych, utrudnienia w życiu codziennym takie jak choroba, niepełnosprawność, obecność dziecka lub czas przeznaczany na edukację.

Po tym, jak stworzyliśmy dwa modele, rozpoczęliśmy ich analizę. Wstępne wyniki jakie otrzymaliśmy nie były dla nas w pełni satysfakcjonujące, dlatego też postanowiliśmy przeprowadzić analizę raz jeszcze, tym razem oczyszczając modele ze zmiennych nieistotnych, a następnie wybierając najlepszy. Po zastosowaniu algorytmu metody krokowej wybraliśmy model ze zmiennymi opisującymi przekonania respondenta. Po analizie statystyk takich jak dobroci dopasowania dewiancji, Pearsona czy analizie interpretacji ilorazów szans, mogliśmy wreszcie uzyskać odpowiedzi na nurtujące nas pytania. Okazało się, że pochodzenie geograficzne nie ma znaczenia, ale ma fakt czy osoba jest szczęśliwa z życia. Co zaskakujące, najwięszy wpływ zaobserwowaliśmy dla zmiennej impfree, a więc fakt że osoby które deklarują, że ich wolność osobista jest dla nich bardzo istotna mają o 63.5% wyższe szanse na rozwód w stosunku do osób, dla których wolność osobista jest przeciętnie istotna.

# 

# 12. Bibliografia

[1] G. C. Kitson, K. B. Babri, M. J. Roach, Who Divorces and Why: A Review, Journal of Family Issues 1985, nr 6, s. 3.

[2] E.Rosset, Rozwody, Warszawa 1986.

[3] Frejka, T. (2008). Determinants of family formation and childbearing during the societal transition in Central and Eastern Europe. Demographic Research, 19 (7), 139–170.

[4] Kotowska, I., Józwiak, J., Matysiak, A., Baranowska, A. (2008). Poland: Fertility decline as a response to profound societal and labour market changes? Demographic Research, 19 (22), 795–854

[5] Matysiak, A. (2009). Employment Þ rst, then childbearing: Wo - men’s strategy in post-socialist Poland. Population Studies, 63 (3), 253–276

[6] [http://kolegia.sgh.waw.pl/pl/KAE/struktura/ISiD/Documents/KSIĄZKA%20ISD%20wersja%2011-03-2019%20.pdf](http://kolegia.sgh.waw.pl/pl/KAE/struktura/ISiD/Documents/KSIA%CC%A8ZKA%20ISD%20wersja%2011-03-2019%20.pdf)

[7] <https://www.poradnikzdrowie.pl/psychologia/relacje/od-fascynacji-do-prawdziwej-milosci-aa-s9XS-W71U-Mx81.html>

[8] <https://www.rp.pl/Rodzina/307219993-Rozwodow-w-Niemczech-jest-coraz-mniej.html>

[9] Rydzewski P., 2010, Socjologiczne analizy rozwodów. Aspekty teoretyczne, empiryczne i metodologiczne, Wydawnictwo WSPA, Lublin.

[10] <https://www.europeansocialsurvey.org/docs/round7/survey/ESS7_appendix_a11_e01_0.pdf> str 2

[11] <http://dspace.uni.lodz.pl:8080/xmlui/bitstream/handle/11089/5686/WNS.pdf?sequence=1> Zawieranie i rozpad związków małżeńskich a zmiany w strukturze gospodarstw domowych i rodzin, Wanda Nowak-Sapota, Uniwersytet Łódzki

[12] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2857783/>

[13] <https://journals.sagepub.com/doi/10.1177/0265407505054516>

[14] <http://sd.pan.pl/old/images/stories/pliki/Archiwum/2010_1-2_4_ms.pdf> str 38

[15] <http://www.teologia.pl/m_k/kkk1s20.htm> KKK2384

[16] Lyngstad T.H., Jalovaara M., 2010, A review of the antecedents of union dissolution, „Demographic Research”, 23(10): 257-292.

[17] <https://stat.gov.pl/obszary-tematyczne/roczniki-statystyczne/roczniki-statystyczne/rocznik-demograficzny-2017,3,11.html> Rocznik Demograficzny 2017, str. 246, 248, 254, 256

[18] <http://kolegia.sgh.waw.pl/pl/KAE/struktura/ISiD/projekty/famwell/konferencja/Documents/7_StyrcMatysiak_Niestabilnosc_malzenstw.pdf>

[19] Shelby B. Scott, Galena K. Rhoades, Scott M. Stanley, Elizabeth S. Allen, and Howard J. Markman, 2013, Reasons for Divorce and Recollections of Premarital Intervention: Implications for Improving Relationship Education; <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4012696/>

[20] *Journal of Marriage and Family*, Steven Stack and J. Ross Eshleman, [National Council on Family Relations](https://www.jstor.org/publisher/ncfr)

[21] <https://stat.gov.pl/obszary-tematyczne/roczniki-statystyczne/roczniki-statystyczne/rocznik-demograficzny-2017,3,11.html> Rocznik Demograficzny 2017, str. 246, 248, 254, 256

[22] <https://pdfs.semanticscholar.org/8cac/04a71b433c465b781738bb77423d4af79528.pdf>

[23] [ttps://link.springer.com/article/10.1186/s40711-015-0003-0](https://link.springer.com/article/10.1186/s40711-015-0003-0) Qi Xu, Jianning YuZeqi Qiu, 2015, The impact of children on divorce risk, The Journal of Chinese Sociology.

[24] <http://sd.pan.pl/old/images/stories/pliki/Archiwum/2010_1-2_4_ms.pdf> (31)

[25] Wieczorek M., 1999, Zmiany procesu rozpadu małżeństw, [w:] Kotowska I.E. (red.), Przemiany demografi czne w Polsce w latach 90. w świetle koncepcji drugiego przejścia demografi cznego, Monografi e i Opracowania 461, Szkoła Główna Handlowa, Warszawa.

[26] <http://www.asanet.org/sites/default/files/attach/journals/aug16asrfeature.pdf>, Money, Work, and Marital Stability: Assessing Change in the Gendered Determinants of Divorce, Alexandra Killewald, American Sociological Review 2016

[27] <https://sjp.pwn.pl/sjp/niezalezny;2490308.html>

[28] I. Janicka, Kohabitacja a małżeństwo w perspektywie psychologicznej. Studium porównawcze, Wyd. UŁ, Łódź 2006, s. 19–22.

[29] Burgoyne, C., Morison, V. (1997). Money in Remarriage: Keeping Things Simple – and Separate. The Sociological Review, 45(3), 363–95

[30] <https://stat.gov.pl/cps/rde/xbcr/gus/rs_rocznik_demograficzny_2013.pdf>

[31] <http://ur.edu.pl/pliki/Zeszyt17/31.pdf> Reformy systemu zasiłków a wzrost nierówności

w Niemczech, Dr Lilianna Jodkowska, Hochschule für Technik und Wirtschaft Berlin

[32] [https://algolytics.pl/tutorial-jak-ocenic-jakosc-i-poprawnosc-modeli-klasyfikacyjnych-czesc-4-krzywa-roc/](https://algolytics.pl/tutorial-jak-ocenic-jakosc-i-poprawnosc-modeli-klasyfikacyjnych-czesc-4-krzywa-roc/?fbclid=IwAR3hTrAzomDfEOJMsqbM6ttPk1E6zaeRnA8NpaWx5FrMKt1GfAfzf_Qrgq4)

# 13. Kody SAS

### Filtrowanie danych

LIBNAME TMP00001 "D:\Michal\Dokumenty\STUDIA - SGH\II\Regresja logistyczna\Projekt\ESS8DE.sas";

%\_eg\_conditional\_dropds(WORK.FILTER\_FOR\_ESS8DE\_SAS7BDAT);

PROC SQL;

CREATE TABLE WORK.FILTER\_FOR\_ESS8DE\_SAS7BDAT AS

SELECT t1.dvrcdeva,

t1.yrbrn,

t1.anctry1,

t1.hhmmb,

t1.maritalb,

t1.agea,

t1.rlgatnd,

t1.sclmeet,

t1.rlgdgr,

t1.nwspol,

t1.eduyrs,

t1.hlthhmp,

t1.stflife,

t1.edctn,

t1.jbspv,

t1.uemp3m,

t1.chldhhe,

t1.bennent,

t1.emplrel,

t1.icwhct,

t1.imptrad,

t1.iplylfr,

t1.lnghom1,

t1.region

FROM TMP00001.ess8de t1;

QUIT;

### Kategoryzacja zmiennych

PROC SQL;

CREATE TABLE WORK.SORTED\_AND\_CATEGORIZED AS

SELECT t1.dvrcdeva,

t1.agea,

t1.anctry1,

t1.hhmmb,

t1.Imptrad,

t1.Nwspol,

t1.rlgatnd,

t1.sclmeet,

t1.rlgdgr,

t1.eduyrs,

t1.hlthhmp,

t1.stflife,

t1.edctn,

t1.jbspv,

t1.uemp3m,

t1.chldhhe,

t1.emplrel,

t1.bennent,

t1.icwhct,

t1.pplfair,

t1.imprich,

t1.impfree,

t1.hincfel,

t1.region,

t1.lnghom1

FROM WORK.FILTER\_FOR\_ESS8DE\_SAS7BDAT\_0000 as T1

;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized1 AS

select \*, case when 0<= agea <= 45 then "young"

when 46<= agea <= 66 then "middle"

else "old" end as ageacat

from WORK.SORTED\_AND\_CATEGORIZED;

;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized2 AS

select \*, case when lnghom1 = 'GER' then "GER"

else "OTH" end as lnghom1cat

from WORK.SORTTempTableCategorized1;

;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized3 AS

select \*, case when region = 'DE4' or

region = 'DE3' or

region = 'DE8' or

region = 'DE3' or

region = 'DED' or

region = 'DEE' or

region = 'DEG' then "NRD"

else "RFN" end as regioncat

from WORK.SORTTempTableCategorized2;

;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized4 AS

select \*, case when anctry1 = 11070 or

anctry1 = 11077 or

anctry1 = 11078 or

anctry1 = 11079 then "GER"

else "OTH" end as anctry1cat

from WORK.SORTTempTableCategorized3;

;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized5 AS

select \*, CASE when 1 <= hhmmb <= 3 then 'low'

else 'high' end as hhmmbcat

from WORK.SORTTempTableCategorized4

where hhmmb le 6 ;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized6 AS

select \*

from WORK.SORTTempTableCategorized5

where Rlgatnd lt 77 ;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized7 AS

select \*, case when 0 < eduyrs <= 10 then "basic"

when 10 < eduyrs <= 18 then "medium"

else "higher" end as eduyrscat

from WORK.SORTTempTableCategorized6

where eduyrs ne 99 and eduyrs ne 88;

;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized8 AS

select \*

from WORK.SORTTempTableCategorized7

where Hlthhmp NE 7;

;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized9 AS

select \*, case when 0 < stflife <= 6 then "no"

else "yes" end as stflifecat

from WORK.SORTTempTableCategorized8

where stflife NE 77 and stflife NE 88;

;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized10 AS

select \*, case when Jbspv = 1 then 1

else 2 end as Jbspvcat

from WORK.SORTTempTableCategorized9

where Jbspv NE 7 and Jbspv NE 8

;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized11 AS

select \* from WORK.SORTTempTableCategorized10

where Uemp3m NE 7 and Uemp3m NE 8

;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized12 AS

select \* from WORK.SORTTempTableCategorized11

where chldhhe NE 8

;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized13 AS

select \*, case when 0 < Bennent <= 2 then "Extreme"

when Bennent = 5 then "Extreme"

else "Middle" end as Bennentcat

from WORK.SORTTempTableCategorized12

where Bennent NE 8 and Bennent NE 7

;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized14 AS

select \*, case when Emplrel = 6 then "Not\_applicable"

when Emplrel = 1 then "Position"

else "Business" end as Emplrelcat

from WORK.SORTTempTableCategorized13

where Emplrel NE 8 and Emplrel NE 7

;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized15 AS

select \*,case when Icwhct = 2 then 'no answer'

else 'answer' end as Icwhctcat

from WORK.SORTTempTableCategorized14

;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized16 AS

select \*,case when 1 <= Sclmeet <= 2 then 'low'

when 3 <= Sclmeet <= 5 then "medium"

else 'high' end as Sclmeetcat

from WORK.SORTTempTableCategorized15

where Sclmeet ne 77;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized17 AS

select \*,case when Rlgdgr = 0 then 'no'

when 1 <= Rlgdgr <= 5 then "medium"

else 'high' end as Rlgdgrcat

from WORK.SORTTempTableCategorized16

where Rlgdgr ne 77 and Rlgdgr ne 88;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized18 AS

select \*,case when 0 <= Pplfair <= 2 then 'cagey'

else 'trustful' end as Pplfaircat

from WORK.SORTTempTableCategorized17

where Pplfair ne 88;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized19 AS

select \*,case when 0 <= Imprich <= 3 then 'yes'

else 'no' end as Imprichcat

from WORK.SORTTempTableCategorized18

where Imprich ne 7 and Imprich ne 8;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized20 AS

select \*,case when Impfree = 1 then 'high'

when 2 <= Impfree <= 4 then 'medium'

else 'low' end as Impfreecat

from WORK.SORTTempTableCategorized19

where Impfree ne 7 and Impfree ne 8 and Impfree ne 6;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized21 AS

select \*,case when 0 <= Nwspol <= 30 then 'rarely'

when 31 <= Nwspol <= 60 then 'medium'

else 'often' end as Nwspolcat

from WORK.SORTTempTableCategorized20;

proc sql;

CREATE VIEW WORK.SORTTempTableCategorized22 AS

select \* from WORK.SORTTempTableCategorized21

where Imptrad LT 7;

proc sql;

CREATE TABLE WORK.FINALSortedAndCategorizedTable AS

select \*, case when 1 <= Hincfel <= 2 then 'yes'

else 'no' end as Hincfelcat

from WORK.SORTTempTableCategorized22

where Hincfel ne 8 and Hincfel ne 7;

### Regresja dla modelu zmiennych dotyczących wpływu środowiska

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

Kod wygenerowany przez zadanie SAS-a

Wygenerowany dnia: niedziela, 26 maja 2019 o godz. 19:10:59

Przez zadanie: enviroment\_filtered

Dane wejściowe: Local:WORK.FINALSORTEDANDCATEGORIZEDTABLE

Serwer: Local

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

ODS GRAPHICS ON;

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

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

Sortowanie zbioru Local:WORK.FINALSORTEDANDCATEGORIZEDTABLE

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

PROC SQL;

CREATE VIEW WORK.SORTTempTableSorted AS

SELECT T.dvrcdeva, T.imptrad, T.rlgatnd, T.hlthhmp, T.chldhhe, T.lnghom1cat, T.regioncat, T.anctry1cat, T.eduyrscat, T.Emplrelcat, T.Rlgdgrcat

FROM WORK.FINALSORTEDANDCATEGORIZEDTABLE 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 imptrad (PARAM=EFFECT) rlgatnd (PARAM=EFFECT) hlthhmp (PARAM=EFFECT) chldhhe (PARAM=EFFECT) lnghom1cat (PARAM=EFFECT) regioncat (PARAM=EFFECT) anctry1cat (PARAM=EFFECT) eduyrscat (PARAM=EFFECT) Emplrelcat (PARAM=EFFECT)

Rlgdgrcat (PARAM=REF);

MODEL dvrcdeva (Event = '1')=imptrad rlgatnd hlthhmp chldhhe lnghom1cat regioncat anctry1cat eduyrscat Emplrelcat Rlgdgrcat /

SELECTION=STEPWISE

SLE=0.05

SLS=0.05

INCLUDE=0

INFLUENCE

LACKFIT

AGGREGATE SCALE=NONE

RSQUARE

CTABLE

LINK=LOGIT

CLPARM=WALD

CLODDS=WALD

ALPHA=0.05

;

RUN;

QUIT;

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

Koniec kodu zadania

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

RUN; QUIT;

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

TITLE; FOOTNOTE;

ODS GRAPHICS OFF;

### Regresja dla modelu zmiennych dotyczących wpływu światopoglądu

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

Kod wygenerowany przez zadanie SAS-a

Wygenerowany dnia: niedziela, 26 maja 2019 o godz. 19:07:52

Przez zadanie: opinions\_filtered

Dane wejściowe: Local:WORK.FINALSORTEDANDCATEGORIZEDTABLE

Serwer: Local

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

ODS GRAPHICS ON;

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

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

Sortowanie zbioru Local:WORK.FINALSORTEDANDCATEGORIZEDTABLE

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

PROC SQL;

CREATE VIEW WORK.SORTTempTableSorted AS

SELECT T.dvrcdeva, T.rlgatnd, T.chldhhe, T.hhmmbcat, T.eduyrscat, T.stflifecat, T.Jbspvcat, T.Bennentcat, T.Emplrelcat, T.Icwhctcat, T.Sclmeetcat, T.Rlgdgrcat, T.Pplfaircat, T.Imprichcat, T.Impfreecat, T.Nwspolcat

FROM WORK.FINALSORTEDANDCATEGORIZEDTABLE as T

;

QUIT;

TITLE;

TITLE1 "opinions\_filtered";

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 rlgatnd (PARAM=REF) chldhhe (PARAM=REF) hhmmbcat (PARAM=REF) eduyrscat (PARAM=REF) stflifecat (PARAM=REF) Jbspvcat (PARAM=REF) Bennentcat (PARAM=REF) Emplrelcat (PARAM=REF) Icwhctcat (PARAM=REF) Sclmeetcat (PARAM=REF)

Rlgdgrcat (PARAM=REF) Pplfaircat (PARAM=REF) Imprichcat (PARAM=REF) Impfreecat (PARAM=REF) Nwspolcat (PARAM=REF);

MODEL dvrcdeva (Event = '1')=rlgatnd chldhhe hhmmbcat eduyrscat stflifecat Jbspvcat Bennentcat Emplrelcat Icwhctcat Sclmeetcat Rlgdgrcat Pplfaircat Imprichcat Impfreecat Nwspolcat /

SELECTION=FORWARD

SLE=0.05

INCLUDE=0

INFLUENCE

LACKFIT

AGGREGATE SCALE=NONE

RSQUARE

CTABLE

LINK=LOGIT

CLPARM=WALD

CLODDS=WALD

ALPHA=0.05

;

RUN;

QUIT;

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

Koniec kodu zadania

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

RUN; QUIT;

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

TITLE; FOOTNOTE;

ODS GRAPHICS OFF;