



Szkoła Główna Handlowa w Warszawie

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	3
2. Metodologia	4
3. Dane	5
4. Opis narzędzi oraz statystyk	6
5. Hipotezy badawcze	9
6. Korelaty demograficzne i psychologiczne - objaśnienie	10
7. Opis wybranych zmiennych	15
8. Kategoryzacja zmiennych	52
9. Analiza	68
9.1 Światopogląd	68
9.2 Środowisko	72
9.3 Porównanie obu modeli	82
10. Analiza ostatecznego modelu regresji	84
10.1 Wynik testu zgodności Hosmera i Lemeshowa	84
10.2 Statystyka dobroci dopasowania dewiancji i Pearsona	85
10.3 Interpretacja ilorazów szans i weryfikacja postawionych hipotez	85
10.4 Wyniki a postawione hipotezy badawcze	88
10.5 Krzywa ROC	88
10.6 Badanie wpływu	90
11. Podsumowanie	91
12. Bibliografia	92
13. Kody SAS	94

1. Wstęp

“Małżeństwo zwykle 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 podwyższonemu 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 któregoś 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 rozvodu 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 ciąża 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 rozvodu (prognozy, czynniki zwiększające prawdopodobieństwo rozvodu, np. małżeństwa wymuszone, zbyt młody wiek zawarcia związku), czynniki rozvodu (industrializacja, urbanizacja, aktywizacja zawodowa kobiet, tolerancja opinii publicznej itp.) przyczyny rozvodu (alkoholizm, brak wierności itp.).

Podczas gdy byli małżonkowie podają za przyczyny rozpadu związku bezpośrednie winy partnera lub partnerki, w niniejszej pracy zostaną rozważone strukturalno-demograficzne korelaty rozvodu, 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 socio-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 związki stały się mniej trwałe [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

W części opisowej dla zmiennych korelatów rozwodów społeczno-demograficzne rozpatrywanych w literaturze zostały wyselekcjonowane 24 zmienne, 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). Próbką 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 liczego zbioru przypuszczalnych zmiennych objaśniających, eliminację 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 staje 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 pierwsza 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 zależną. 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 zależną są w pierwszej kolejności rozpatrzone pod względem ich usunięcia. Jeśli zmienne te spełniają określone kryteria zostają usunięte. Po usunięciu pierwszej zmiennej, kolejną braną pod uwagę do usunięcia jest ta zmienna pozostająca w równaniu, która ma najmniejszą korelację cząstkową ze zmienną zależną. Procedura ulega zakończeniu gdy w równaniu nie występują inne zmienne spełniające 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 główne 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. Pożądanym 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 i Lemeshowa 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	hhmb	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	edyurs	Zmienna ciągła	Czas edukacji respondenta w latach	Wartość z zakresu
9	hlthmp	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 miesiące	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

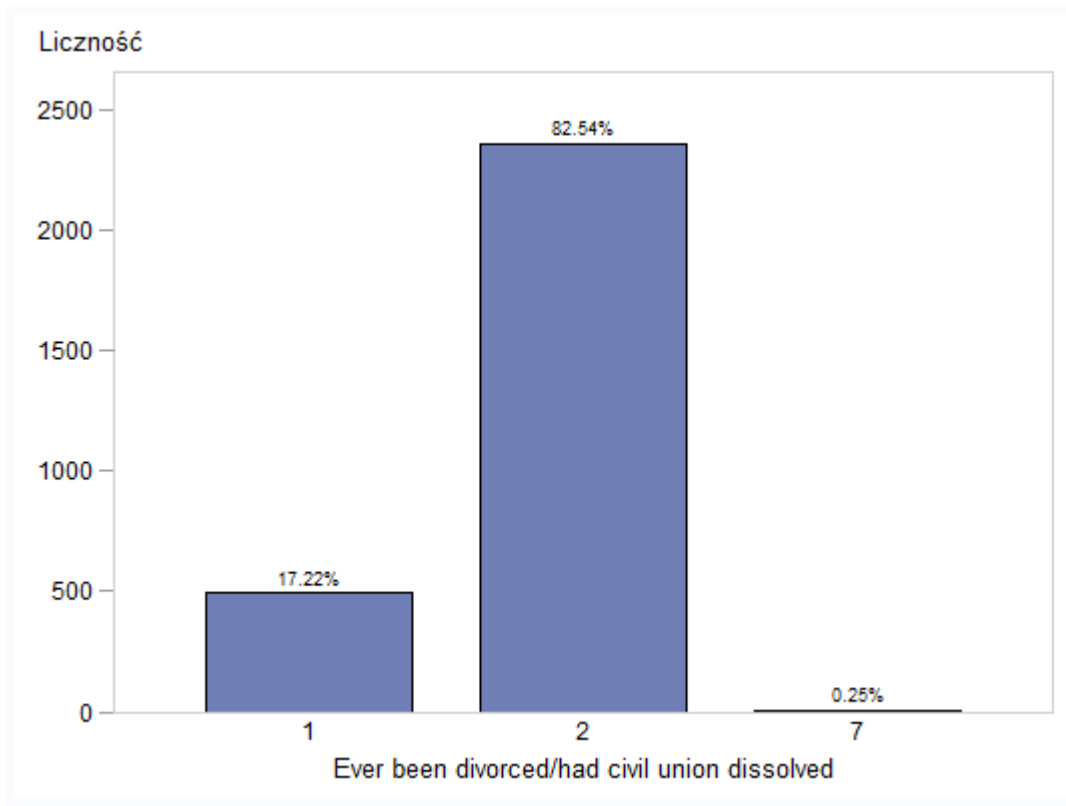
18	imptrad	Zmienna porządkowa	Subiektywna ocena respondenta podobieństwa do osoby, dla której tradycje i wartości mają dużą wagę	1 - bardzo podobna 2 - podobna 3 - w pewnym stopniu podobna 4 - trochę podobna 5 - niepodobna 6 - bardzo niepodobna 7 - odmowa 8 - nie wie 9 - brak odpowiedzi
19	pplfair	Zmienna 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 drogie 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-Württemberg 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 - dvrcreva



Ever been divorced/had civil union dissolved				
dvrcreva	Liczność	Procent	Liczność skumulowana	Procent skumulowany
1	491	17.22	491	17.22
2	2354	82.54	2845	99.75
7	7	0.25	2852	100.00

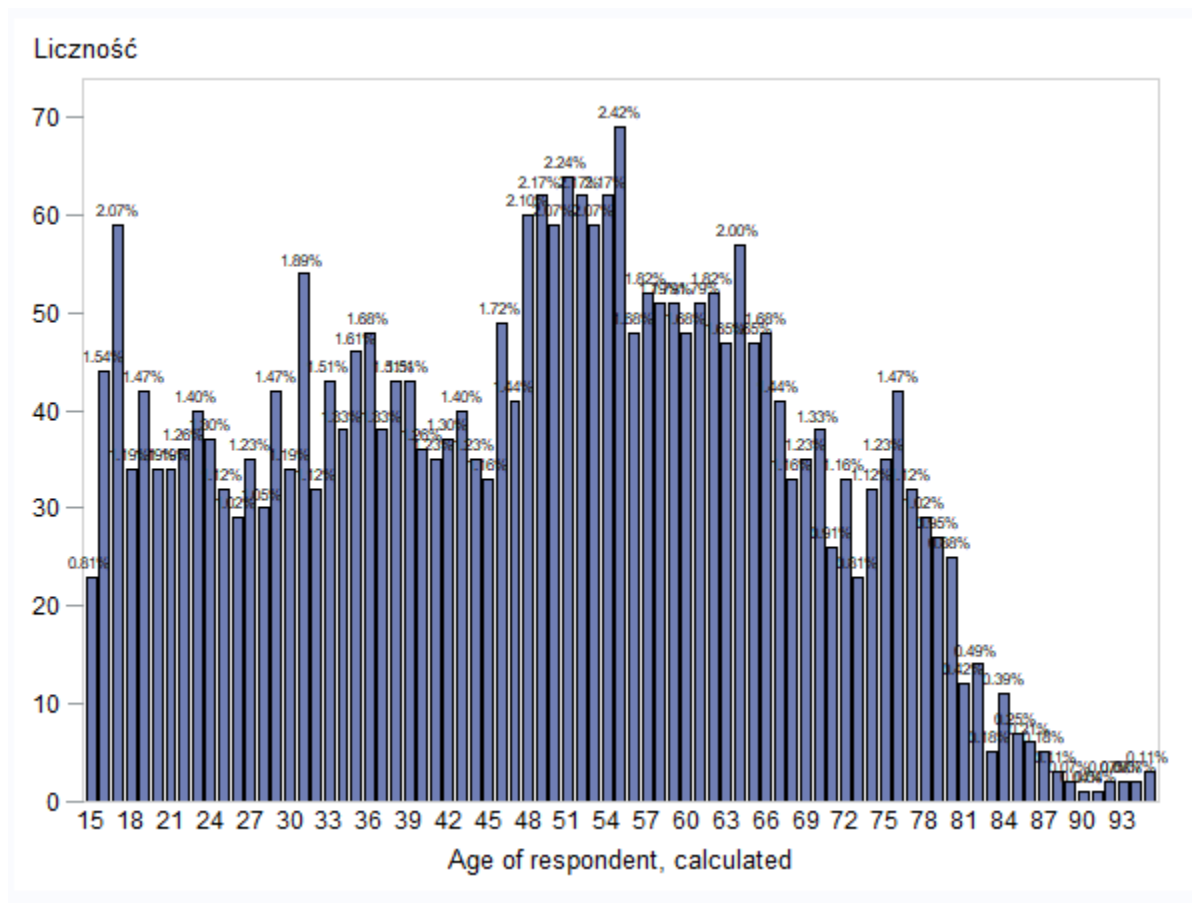
Procedura UNIVARIATE
Zmienna: dvrcdeva (Ever been divorced/had civil union dissolved)

Momenty			
N	2638	Suma wag	2638
Średnia	1.82448825	Suma obserwacji	4813
Odchylenie std.	0.38047635	Wariancja	0.14476225
Skośność	-1.7069889	Kurtoza	0.91450375
Niesk. suma kw.	9163	Skoryg. suma kw.	381.738059
Wsp. zmienności	20.8538668	Błąd std. śr.	0.00740782

Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	1.824488	Odchylenie std.	0.38048
Mediana	2.000000	Wariancja	0.14476
Moda	2.000000	Rozstęp	1.00000
		Rozstęp międzykwartkowy	0

Za zmienną 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



Procedura UNIVARIATE
Zmienna: agea (Age of respondent, calculated)

Momenty			
N	2638	Suma wag	2638
Średnia	48.9162244	Suma obserwacji	129041
Odchylenie std.	18.2059228	Wariancja	331.455626
Skośność	-0.0522893	Kurtoza	-0.9112237
Niesk. suma kw.	7186247	Skoryg. suma kw.	874048.486
Wsp. zmienności	37.2185774	Błąd std. śr.	0.35446659

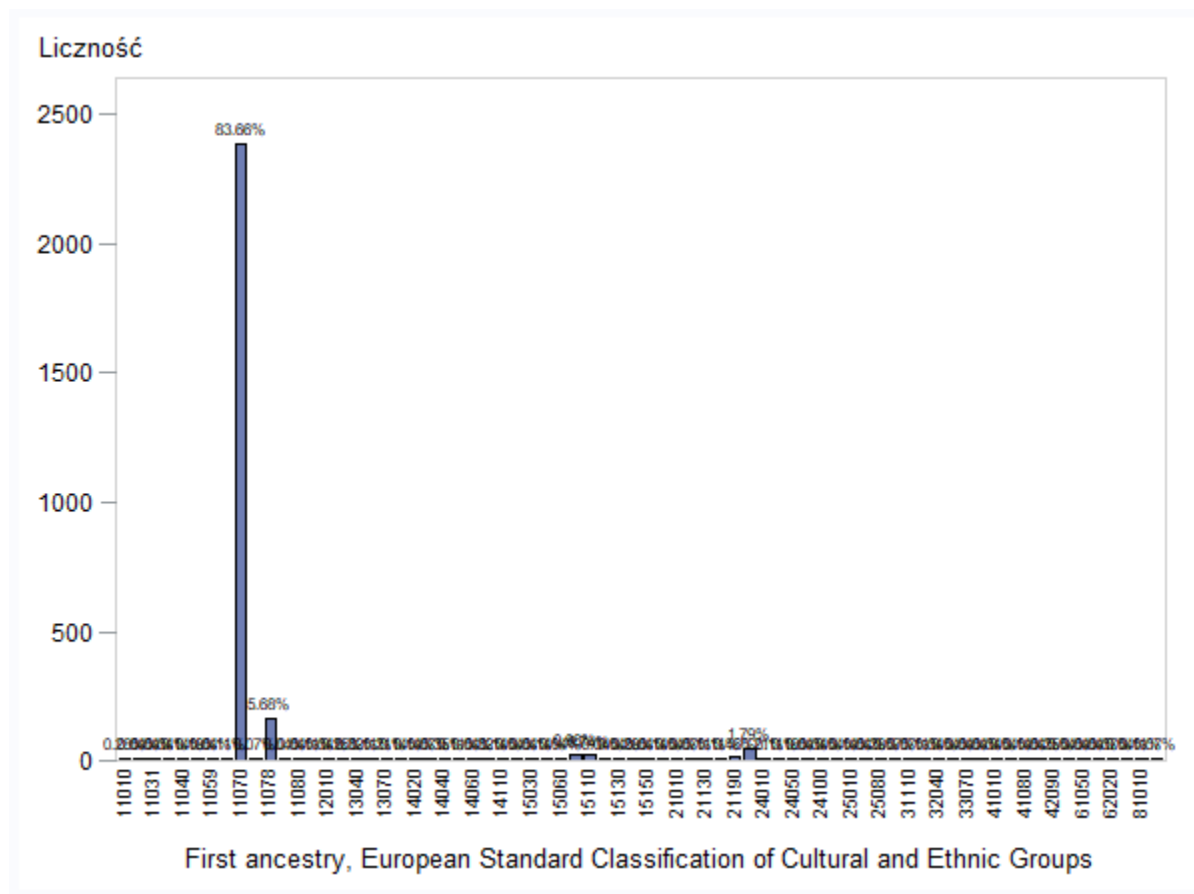
Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	48.91622	Odchylenie std.	18.20592
Mediana	50.00000	Wariancja	331.45563
Moda	55.00000	Rozstęp	79.00000
		Rozstęp międzykwartkowy	29.00000

Najwięcej osób, to te które w momencie orzeczenia rozvodu 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



Procedura UNIVARIATE

Zmienna: anctry1 (First ancestry, European Standard Classification of Cultural and Ethnic Groups)

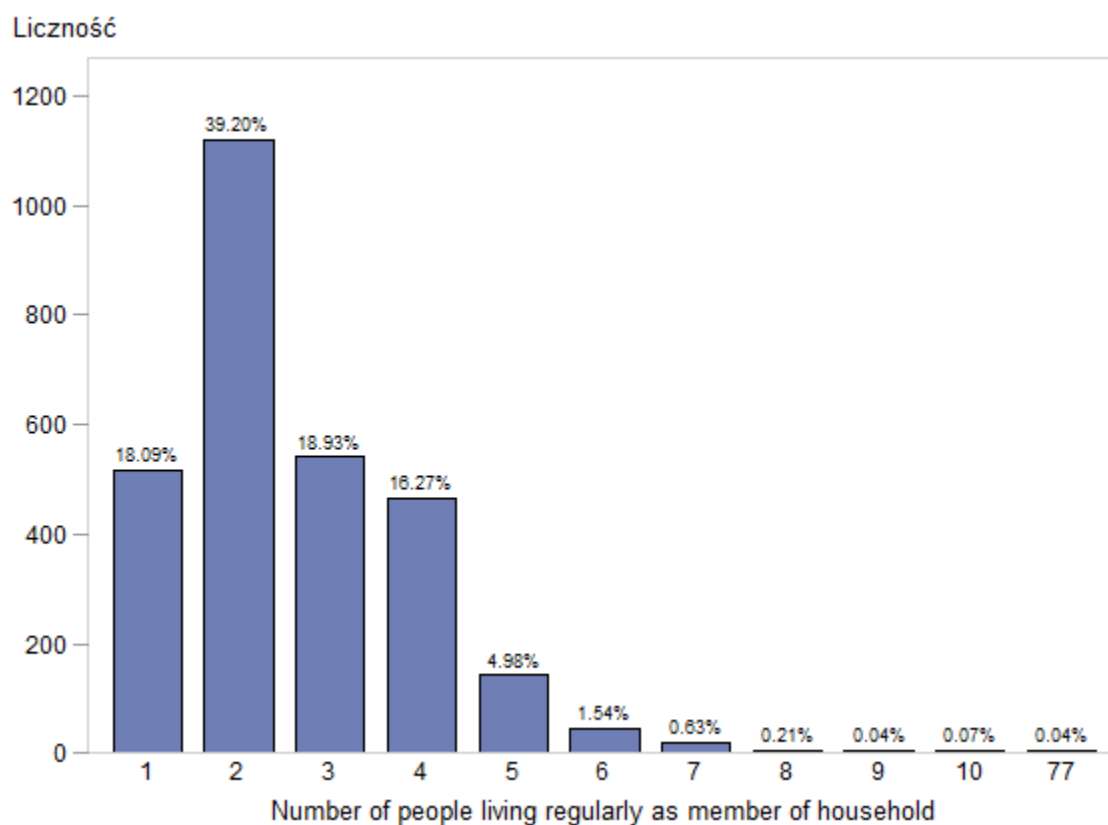
Momenty			
N	2638	Suma wag	2638
Średnia	11946.2244	Suma obserwacji	31514140
Odchylenie std.	4437.80612	Wariancja	19694123.1
Skośność	8.89796252	Kurtoza	103.776746
Niesk. suma kw.	4.28408E 11	Skoryg. suma kw.	5.19334E 10
Wsp. zmienności	37.1481898	Błąd std. śr.	86.4034192

Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	11946.22	Odchylenie std.	4438
Mediana	11070.00	Wariancja	19694123
Moda	11070.00	Rozstęp	70000
		Rozstęp międzykwartkowy	0

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 objaśniająca - Hhmb - Liczba mieszkańców gospodarstwa domowego



Number of people living regularly as member of household				
hhmb	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
1	516	18.09	516	18.09
2	1118	39.20	1634	57.29
3	540	18.93	2174	76.23
4	464	16.27	2638	92.50
5	142	4.98	2780	97.48
6	44	1.54	2824	99.02
7	18	0.63	2842	99.65
8	6	0.21	2848	99.86
9	1	0.04	2849	99.89
10	2	0.07	2851	99.96
77	1	0.04	2852	100.00

Zmienna: hhmb (Number of people living regularly as member of household)

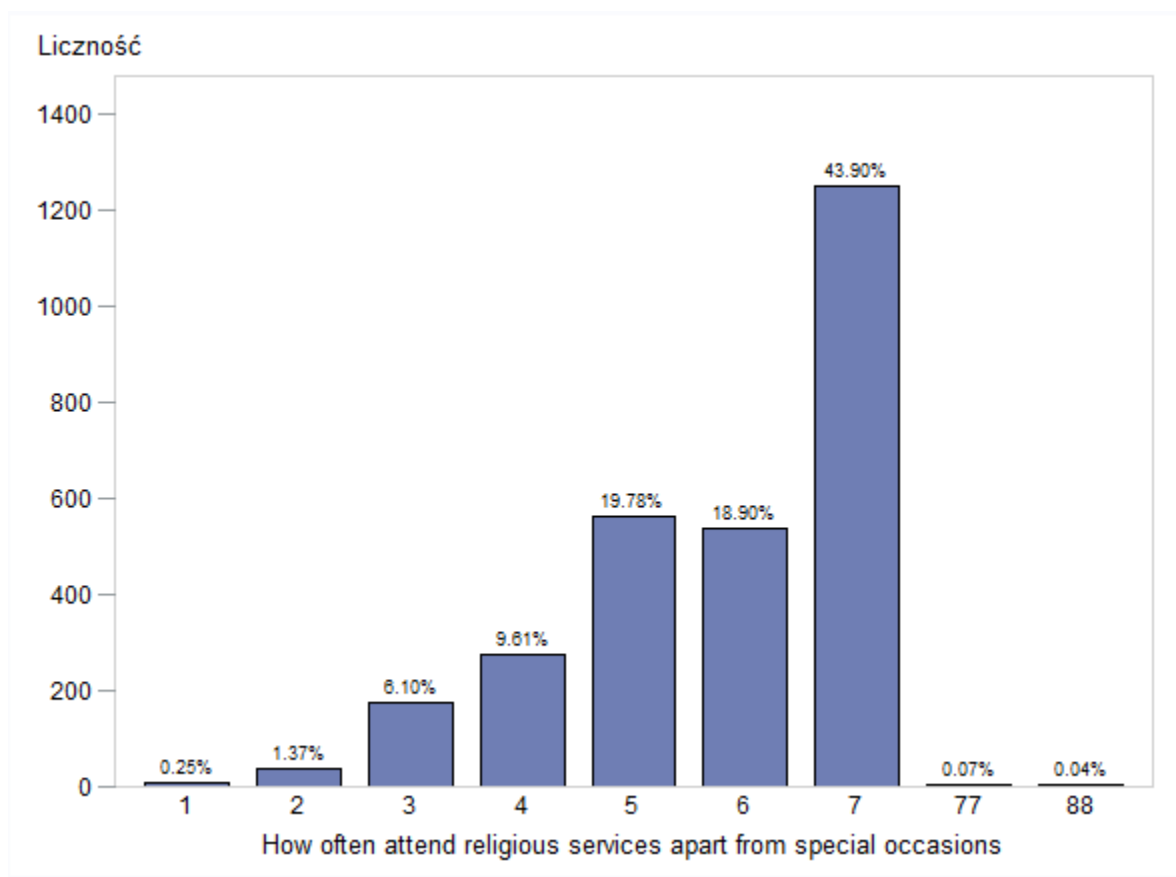
Momenty			
N	2638	Suma wag	2638
Średnia	2.54169826	Suma obserwacji	6705
Odchylenie std.	1.19042336	Wariancja	1.41710777
Skośność	0.67103467	Kurtoza	-0.1437292
Niesk. suma kw.	20779	Skoryg. suma kw.	3736.91319
Wsp. zmienności	46.8357467	Błąd std. śr.	0.02317736

Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	2.541698	Odchylenie std.	1.19042
Mediana	2.000000	Wariancja	1.41711
Moda	2.000000	Rozstęp	5.00000
		Rozstęp międzykwartkowy	1.00000

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



How often attend religious services apart from special occasions				
rlgatnd	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
Every day	7	0.25	7	0.25
More than once a week	39	1.37	46	1.61
Once a week	174	6.10	220	7.71
At least once a month	274	9.61	494	17.32
Only on special holy days	564	19.78	1058	37.10
Less often	539	18.90	1597	56.00
Never	1252	43.90	2849	99.89
Refusal	2	0.07	2851	99.96
Don't know	1	0.04	2852	100.00

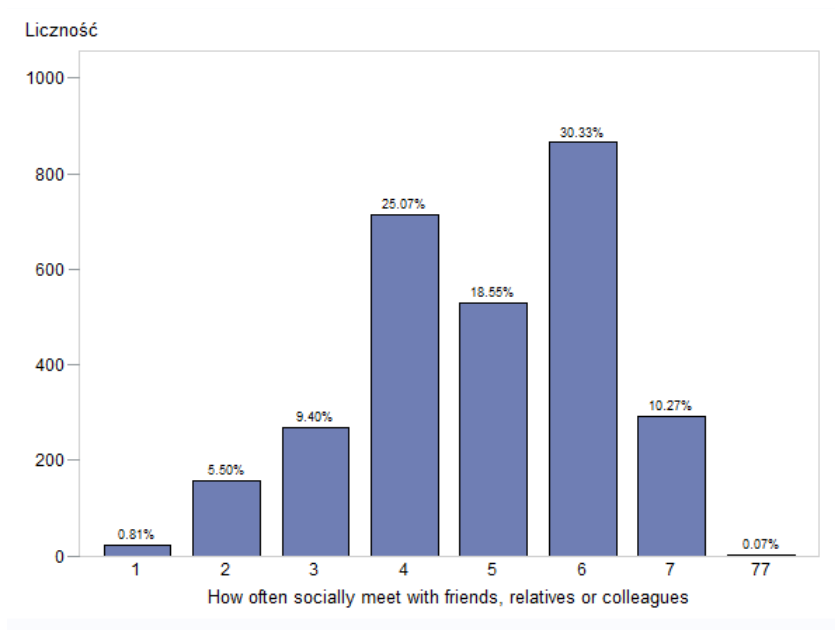
Momenty			
N	2638	Suma wag	2638
Średnia	5.8191812	Suma obserwacji	15351
Odchylenie std.	1.32716855	Wariancja	1.76137635
Skośność	-0.9224212	Kurtoza	-0.0107905
Niesk. suma kw.	93975	Skoryg. suma kw.	4644.74943
Wsp. zmienności	22.8067919	Błąd std. śr.	0.02583977

Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	5.819181	Odchylenie std.	1.32717
Mediana	6.000000	Wariancja	1.76138
Moda	7.000000	Rozstęp	6.00000
		Rozstęp międzykwartkowy	2.00000

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.



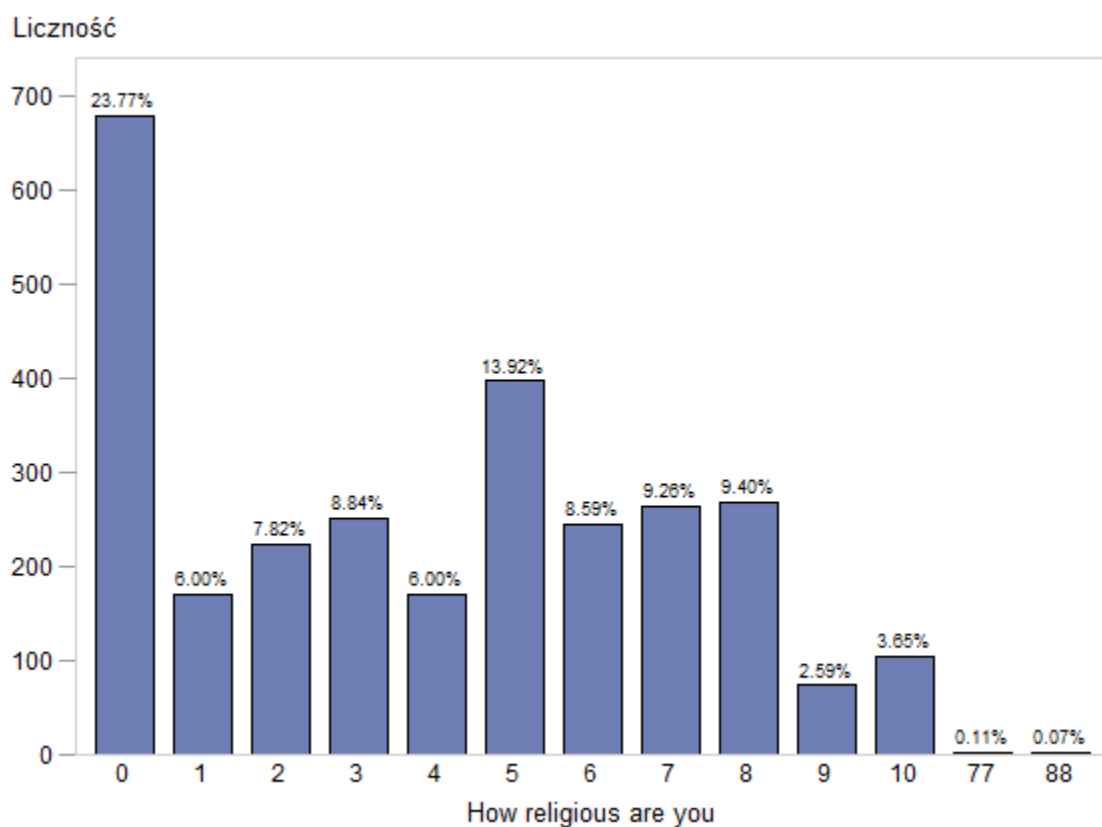
How often socially meet with friends, relatives or colleagues				
sclmeet	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
Never	23	0.81	23	0.81
Less than once a month	157	5.50	180	6.31
Once a month	268	9.40	448	15.71
Several times a month	715	25.07	1163	40.78
Once a week	529	18.55	1692	59.33
Several times a week	865	30.33	2557	89.66
Every day	293	10.27	2850	99.93
Refusal	2	0.07	2852	100.00

Momenty			
N	2638	Suma wag	2638
Średnia	4.85670963	Suma obserwacji	12812
Odchylenie std.	1.38108961	Wariancja	1.90740851
Skośność	-0.390132	Kurtoza	-0.5455461
Niesk. suma kw.	67254	Skoryg. suma kw.	5029.83624
Wsp. zmienności	28.4367342	Błąd std. śr.	0.02688961

Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	4.856710	Odchylenie std.	1.38109
Mediana	5.000000	Wariancja	1.90741
Moda	6.000000	Rozstęp	6.00000
		Rozstęp międzykwartkowy	2.00000

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



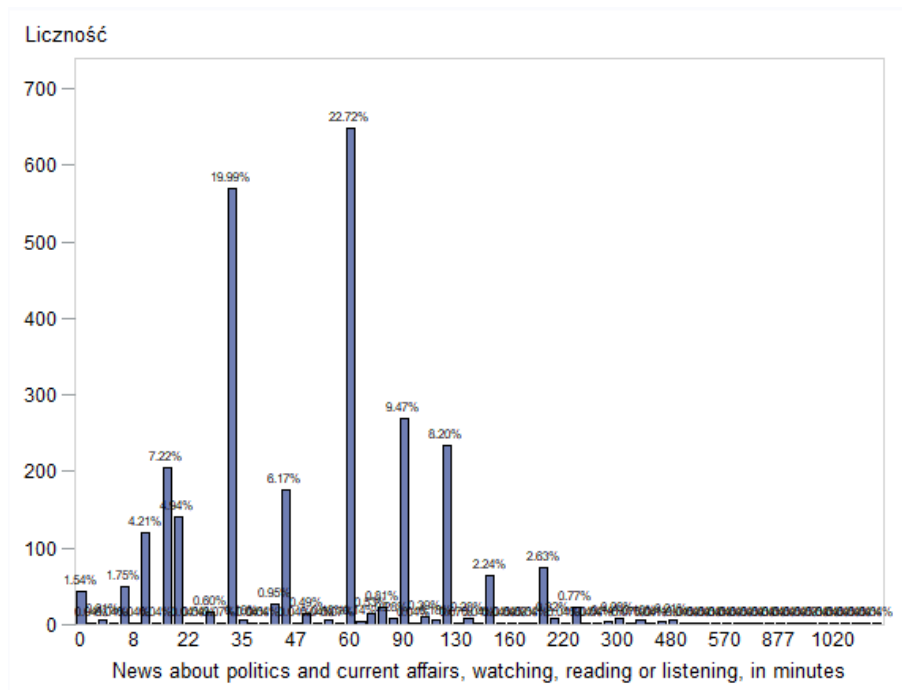
How religious are you				
rlgdgr	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
Not at all religious	678	23.77	678	23.77
1	171	6.00	849	29.77
2	223	7.82	1072	37.59
3	252	8.84	1324	46.42
4	171	6.00	1495	52.42
5	397	13.92	1892	66.34
6	245	8.59	2137	74.93
7	264	9.26	2401	84.19
8	268	9.40	2669	93.58
9	74	2.59	2743	96.18
Very religious	104	3.65	2847	99.82
Refusal	3	0.11	2850	99.93
Don't know	2	0.07	2852	100.00

Momenty			
N	2638	Suma wag	2638
Średnia	3.89158453	Suma obserwacji	10266
Odchylenie std.	3.07293426	Wariancja	9.44292498
Skośność	0.16382605	Kurtoza	-1.1981779
Niesk. suma kw.	64852	Skoryg. suma kw.	24900.9932
Wsp. zmienności	78.9635748	Błąd std. śr.	0.05982957

Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	3.891585	Odchylenie std.	3.07293
Mediana	4.000000	Wariancja	9.44292
Moda	0.000000	Rozstęp	10.00000
		Rozstęp międzykwartkowy	5.00000

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

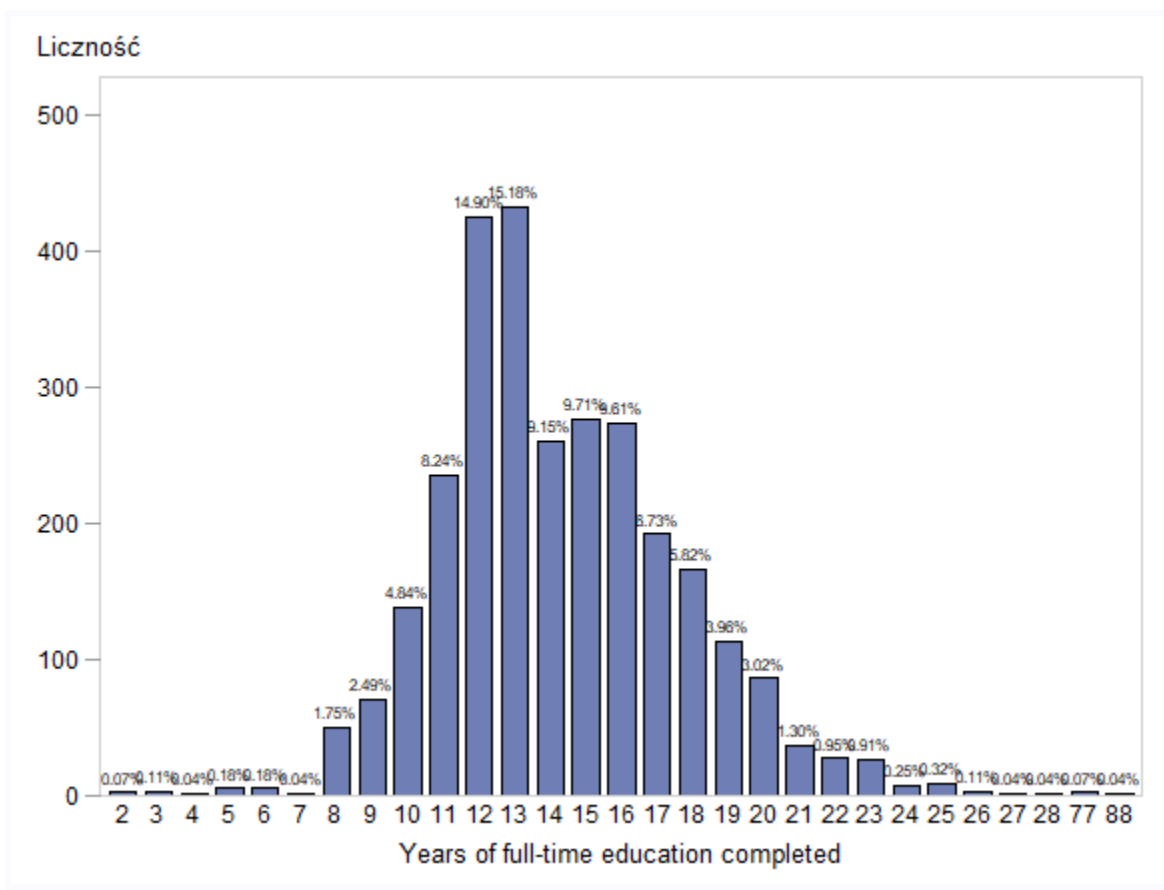


Momenty			
N	2638	Suma wag	2638
Średnia	67.373768	Suma obserwacji	177732
Odchylenie std.	81.4855517	Wariancja	6639.89513
Skośność	6.45331827	Kurtoza	63.810405
Niesk. suma kw.	29483878	Skoryg. suma kw.	17509403.5
Wsp. zmienności	120.945516	Błąd std. śr.	1.58651146

Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	67.37377	Odchylenie std.	81.48555
Mediana	60.00000	Wariancja	6640
Moda	60.00000	Rozstęp	1200
		Rozstęp międzykwartkowy	60.00000

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 - eduysr - czas edukacji respondenta w latach



Momenty			
N	2638	Suma wag	2638
Średnia	14.3400303	Suma obserwacji	37829
Odchylenie std.	3.29246809	Wariancja	10.8403462
Skośność	0.45023152	Kurtoza	0.51928662
Niesk. suma kw.	571055	Skoryg. suma kw.	28585.9928
Wsp. zmienności	22.95998	Błąd std. śr.	0.06410386

Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	14.34003	Odchylenie std.	3.29247
Mediana	14.00000	Wariancja	10.84035
Moda	13.00000	Rozstęp	26.00000
		Rozstęp międzykwartkowy	4.00000

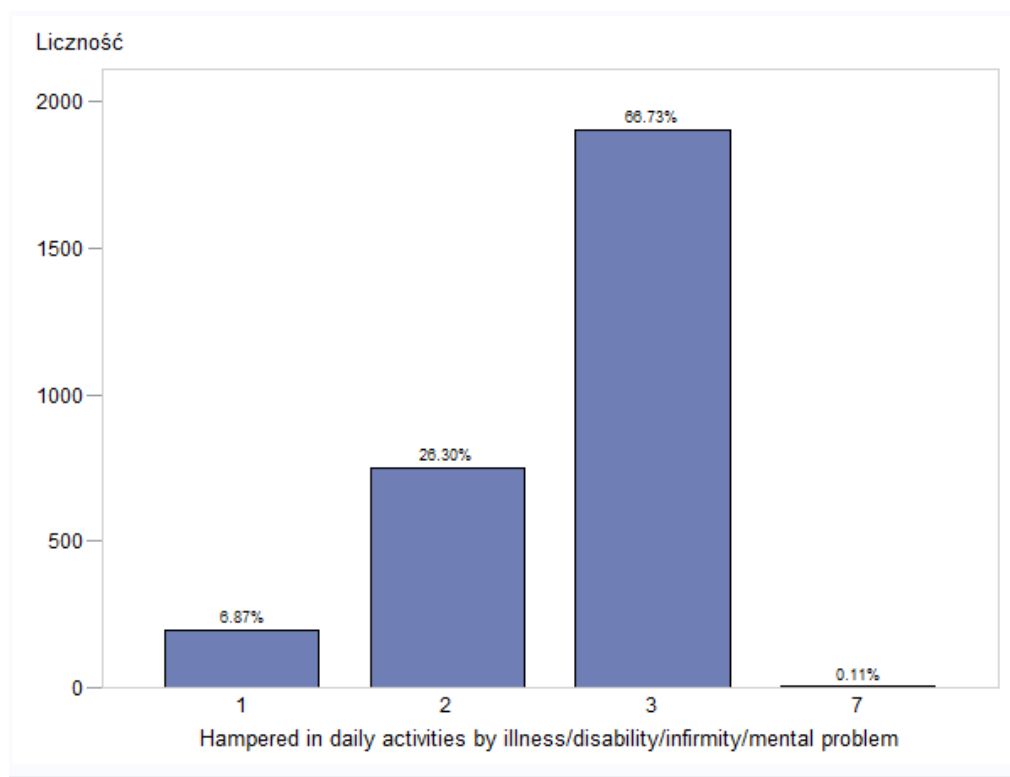
Zmienna analizowana: eduysr Years of full-time education completed						
Kurtoza	Skośność	Średnia	Odch. std.	Minimum	Maksimum	Wsp. zmienności
0.5190996	0.4607689	14.2570126	3.3122653	2.0000000	28.0000000	23.2325338

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 - hlthmp - utrudnienie w życiu codziennymi wynikające z choroby, niepełnosprawności, choroby umysłowej etc



Hampered in daily activities by illness/disability/infirmity/mental problem				
hlthmp	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
1	196	6.87	196	6.87
2	750	26.30	946	33.17
3	1903	66.73	2849	99.89
7	3	0.11	2852	100.00

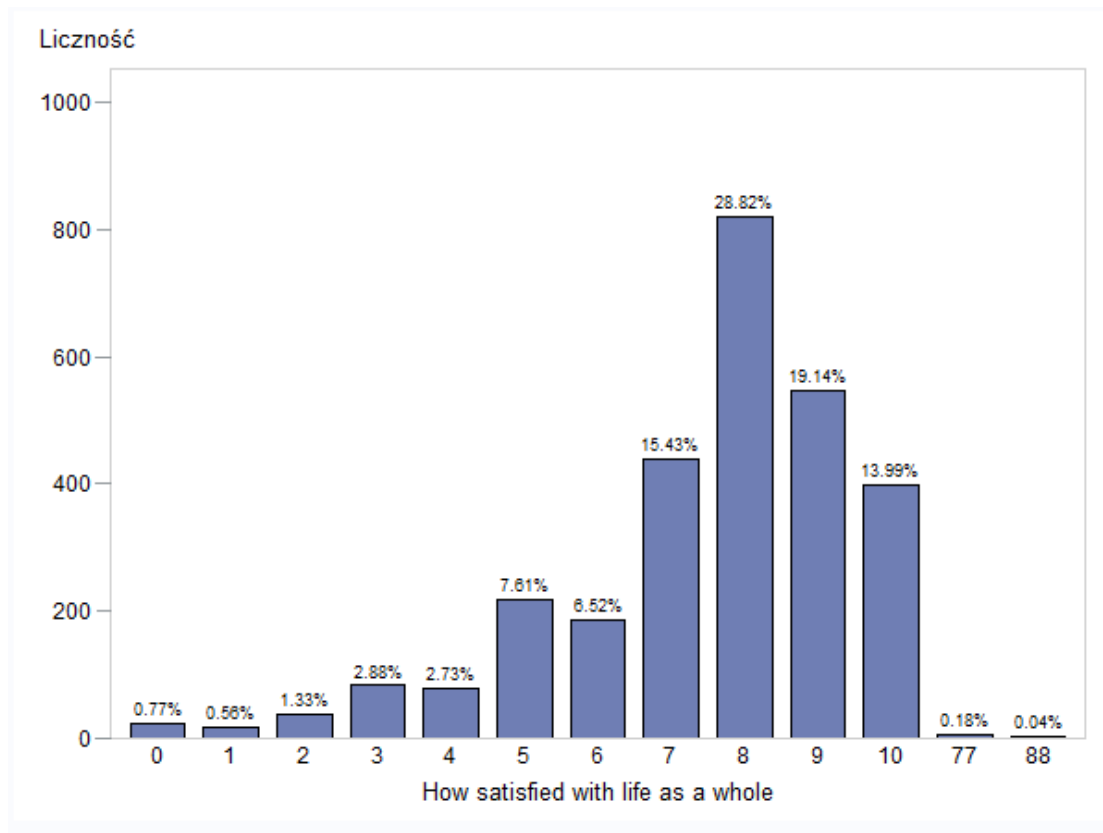
Momenty			
N	2638	Suma wag	2638
Średnia	2.60121304	Suma obserwacji	6862
Odchylenie std.	0.61162922	Wariancja	0.3740903
Skośność	-1.2713844	Kurtoza	0.52542171
Niesk. suma kw.	18836	Skoryg. suma kw.	986.476118
Wsp. zmienności	23.5132305	Błąd std. śr.	0.01190833

Bazowe miary statystyczne			
Polozenie		Zmienność	
Średnia	2.601213	Odchylenie std.	0.61163
Mediana	3.000000	Wariancja	0.37409
Moda	3.000000	Rozstęp	2.00000
		Rozstęp międzykwartkowy	1.00000

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

Najczęściej zgłaszanymi przyczynami rozvodu 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 rozvodu. [19]

Zmienna objaśniająca - stflife - satysfakcja z życia



How satisfied with life as a whole				
stflife	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
0	22	0.77	22	0.77
1	16	0.56	38	1.33
2	38	1.33	76	2.66
3	82	2.88	158	5.54
4	78	2.73	236	8.27
5	217	7.61	453	15.88
6	186	6.52	639	22.41
7	440	15.43	1079	37.83
8	822	28.82	1901	66.65
9	546	19.14	2447	85.80
10	399	13.99	2846	99.79
77	5	0.18	2851	99.96
88	1	0.04	2852	100.00

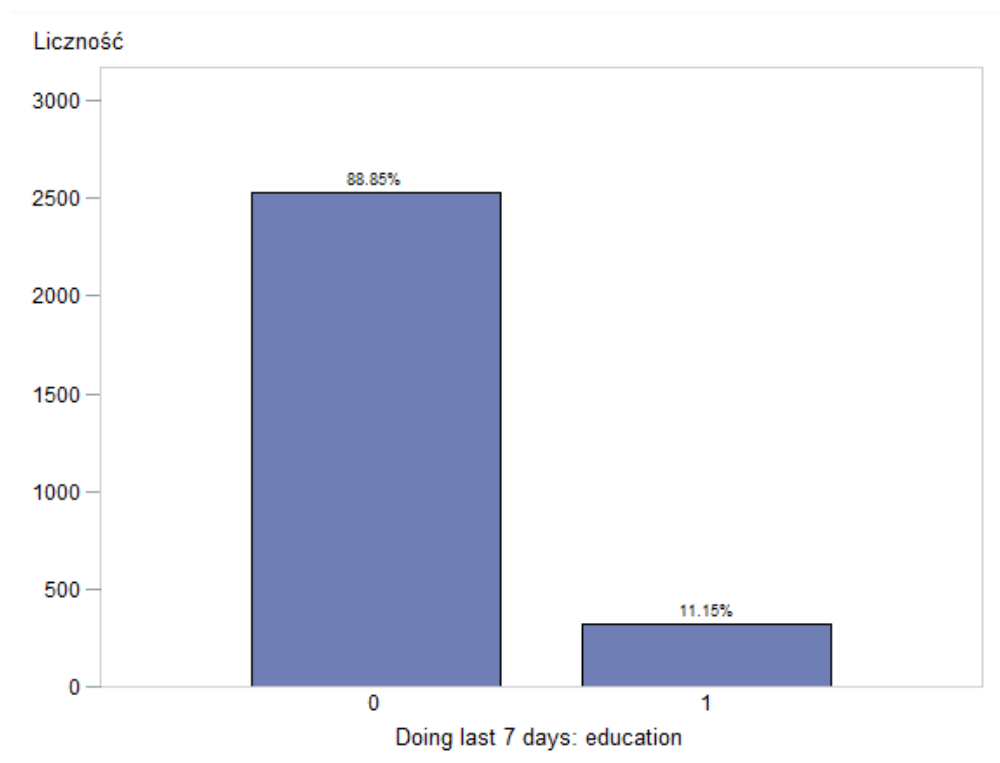
Momenty			
N	2638	Suma wag	2638
Średnia	7.5242608	Suma obserwacji	19849
Odchylenie std.	1.98782293	Wariancja	3.95144001
Skośność	-1.1955868	Kurtoza	1.57120537
Niesk. suma kw.	159769	Skoryg. suma kw.	10419.9473
Wsp. zmienności	26.4188468	Błąd std. śr.	0.03870261

Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	7.524261	Odchylenie std.	1.98782
Mediana	8.000000	Wariancja	3.95144
Moda	8.000000	Rozstęp	10.00000
		Rozstęp międzyćwiartkowy	2.00000

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 edukacją w ostatnich 7 dniach od ankiety



Doing last 7 days: education				
edctn	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
0	2534	88.85	2534	88.85
1	318	11.15	2852	100.00

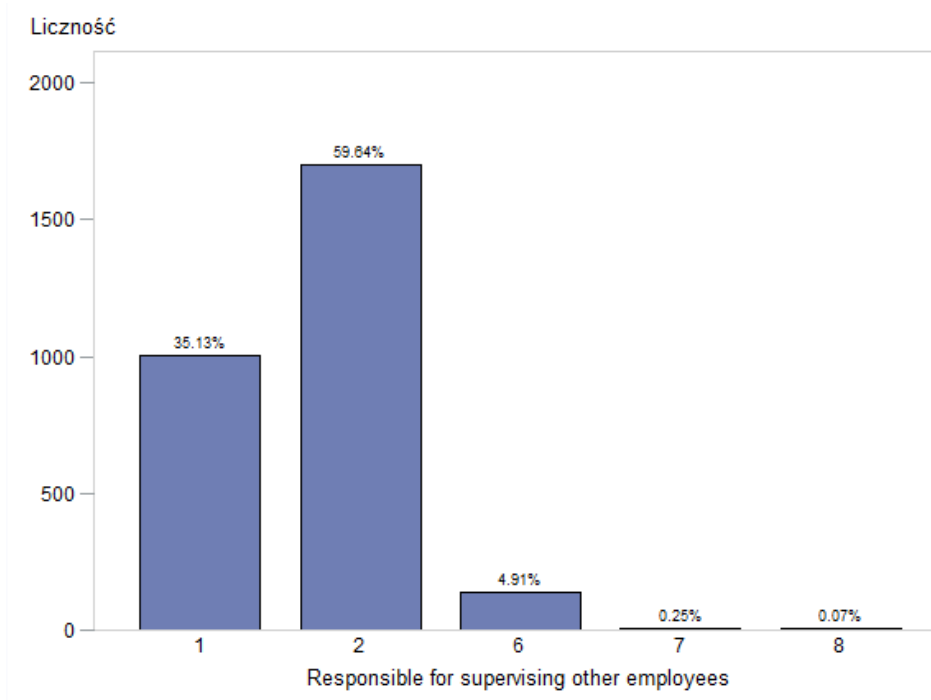
Momenty			
N	2638	Suma wag	2638
Średnia	0.10310842	Suma obserwacji	272
Odchylenie std.	0.30415808	Wariancja	0.09251214
Skośność	2.61175184	Kurtoza	4.82490512
Niesk. suma kw.	272	Skoryg. suma kw.	243.954511
Wsp. zmienności	294.988611	Błąd std. śr.	0.00592191

Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	0.103108	Odchylenie std.	0.30416
Mediana	0.000000	Wariancja	0.09251
Moda	0.000000	Rozstęp	1.00000
		Rozstęp międzykwartkowy	0

Zmienna odpowiada na pytanie o działalność naukową respondenta w ostatnich 7 dniach od przeprowadzenia badania. Z rozkładu odpowiedzi wynika, że prawie 90% badanych nie dokończyło 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



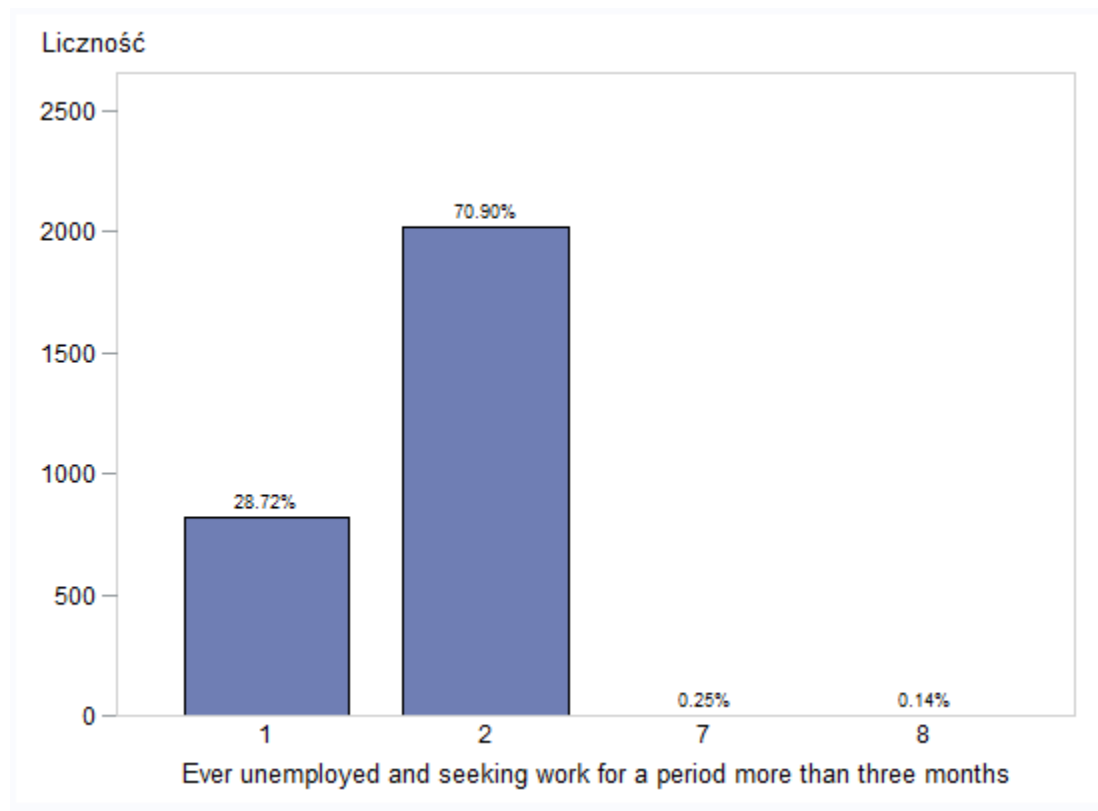
Responsible for supervising other employees				
jbspv	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
1	1002	35.13	1002	35.13
2	1701	59.64	2703	94.78
6	140	4.91	2843	99.68
7	7	0.25	2850	99.93
8	2	0.07	2852	100.00

Momenty			
N	2638	Suma wag	2638
Średnia	1.8040182	Suma obserwacji	4759
Odchylenie std.	0.99974919	Wariancja	0.99949845
Skośność	2.92914443	Kurtoza	10.2412043
Niesk. suma kw.	11221	Skoryg. suma kw.	2635.67741
Wsp. zmienności	55.4179107	Błąd std. śr.	0.01946497

Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	1.804018	Odchylenie std.	0.99975
Mediana	2.000000	Wariancja	0.99950
Moda	2.000000	Rozstęp	5.00000
		Rozstęp międzykwartkowy	1.00000

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



Ever unemployed and seeking work for a period more than three months				
uemp3m	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
1	819	28.72	819	28.72
2	2022	70.90	2841	99.61
7	7	0.25	2848	99.86
8	4	0.14	2852	100.00

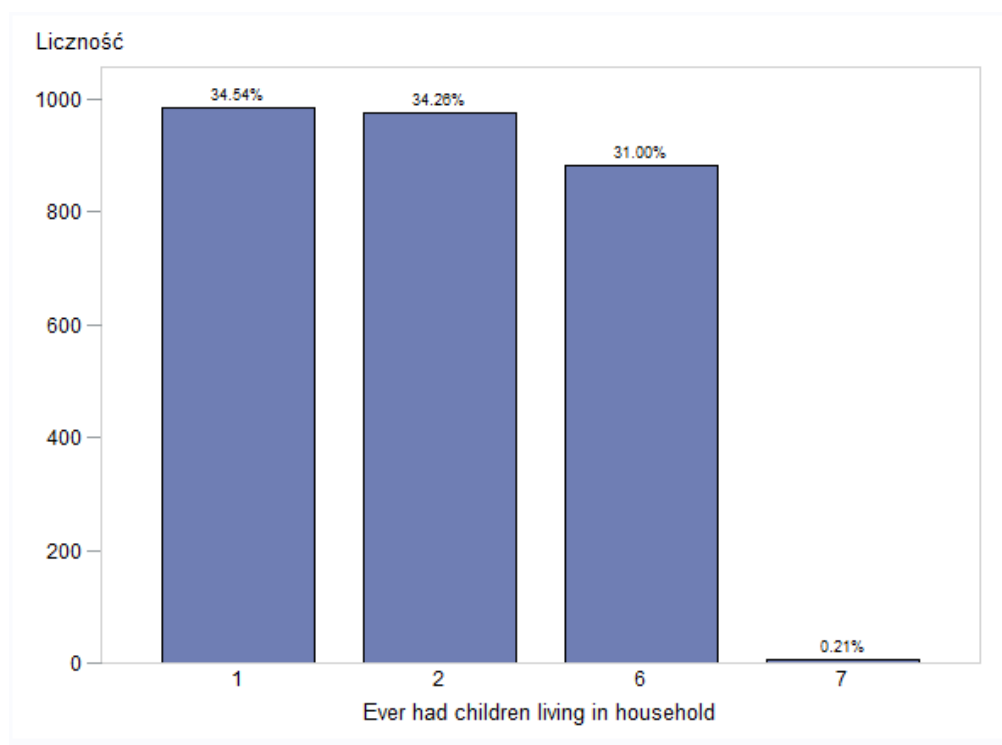
Momenty			
N	2638	Suma wag	2638
Średnia	1.70962851	Suma obserwacji	4510
Odchylenie std.	0.45401986	Wariancja	0.20613403
Skośność	-0.9241338	Kurtoza	-1.1468467
Niesk. suma kw.	8254	Skoryg. suma kw.	543.575436
Wsp. zmienności	26.5566381	Błąd std. śr.	0.0088397

Bazowe miary statystyczne			
Polozenie		Zmienność	
Średnia	1.709629	Odchylenie std.	0.45402
Mediana	2.000000	Wariancja	0.20613
Moda	2.000000	Rozstęp	1.00000
		Rozstęp międzykwartkowy	1.00000

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 rozvodu 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



Ever had children living in household				
chldhhe	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
1	985	34.54	985	34.54
2	977	34.26	1962	68.79
6	884	31.00	2846	99.79
7	6	0.21	2852	100.00

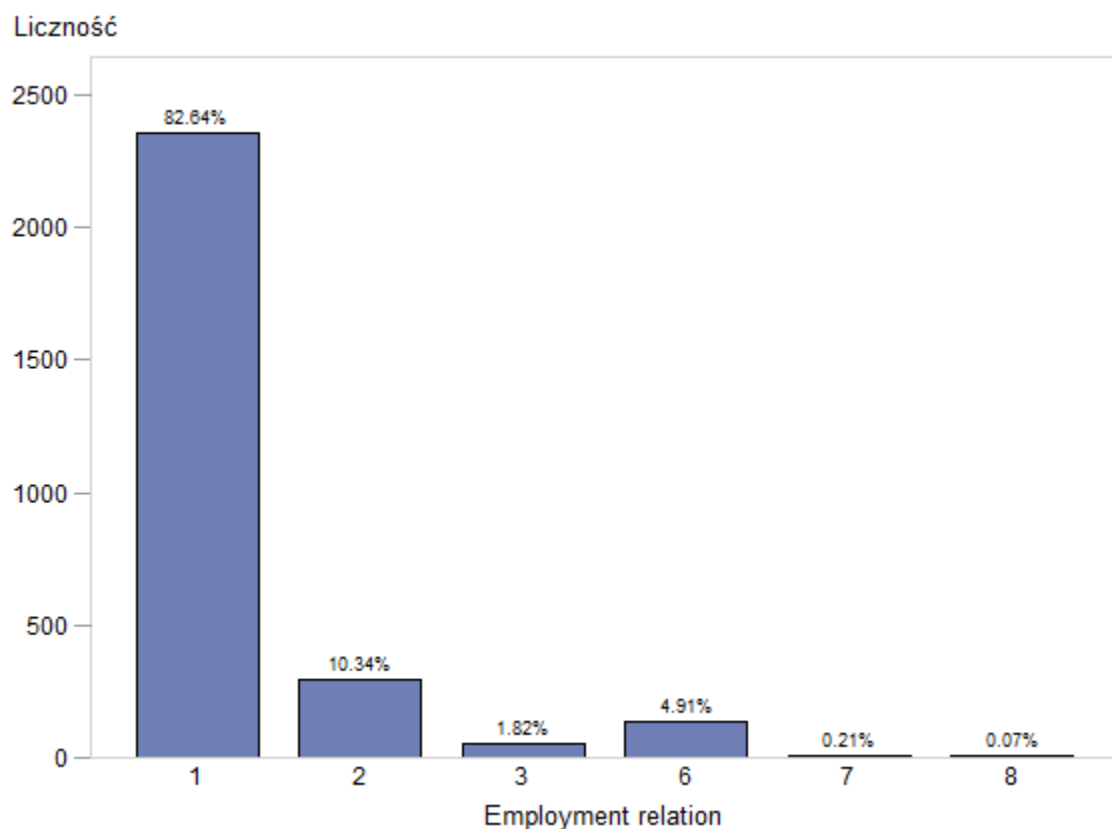
Momenty			
N	2638	Suma wag	2638
Średnia	2.89461713	Suma obserwacji	7636
Odchylenie std.	2.13043599	Wariancja	4.53875751
Skośność	0.69346599	Kurtoza	-1.3619238
Niesk. suma kw.	34072	Skoryg. suma kw.	11968.7036
Wsp. zmienności	73.5999233	Błąd std. śr.	0.04147927

Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	2.894617	Odchylenie std.	2.13044
Mediana	2.000000	Wariancja	4.53876
Moda	1.000000	Rozstęp	5.00000
		Rozstęp międzykwartkowy	5.00000

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



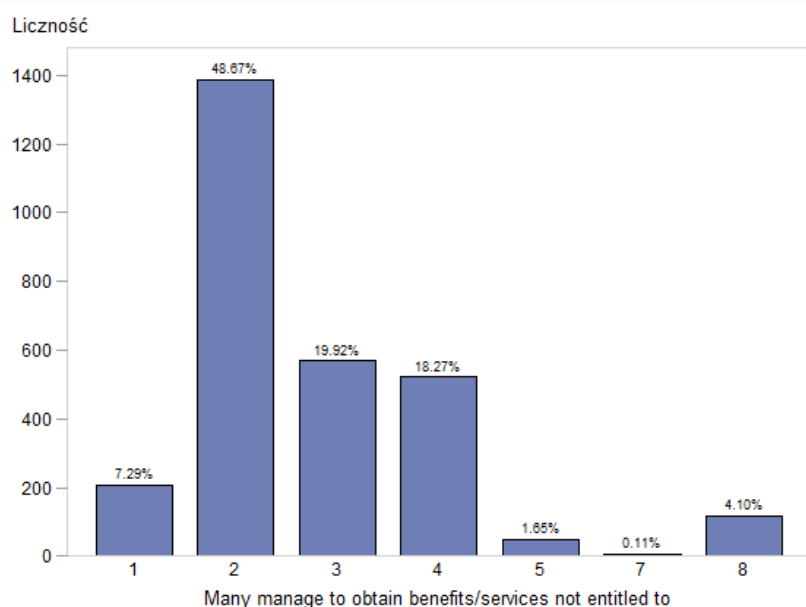
Employment relation				
emplrel	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
Employee	2357	82.64	2357	82.64
Self-employed	295	10.34	2652	92.99
Working for own family business	52	1.82	2704	94.81
Not applicable	140	4.91	2844	99.72
Refusal	6	0.21	2850	99.93
Don't know	2	0.07	2852	100.00

Momenty			
N	2638	Suma wag	2638
Średnia	1.35216073	Suma obserwacji	3567
Odchylenie std.	1.05190343	Wariancja	1.10650083
Skośność	3.69718965	Kurtoza	13.2127514
Niesk. suma kw.	7741	Skoryg. suma kw.	2917.84268
Wsp. zmienności	77.7942599	Błąd std. śr.	0.0204804

Bazowe miary statystyczne			
Polożenie		Zmienność	
Średnia	1.352161	Odchylenie std.	1.05190
Mediana	1.000000	Wariancja	1.10650
Moda	1.000000	Rozstęp	5.00000
		Rozstęp międzykwartkowy	0

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



Many manage to obtain benefits/services not entitled to				
bennent	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
1	208	7.29	208	7.29
2	1388	48.67	1596	55.96
3	568	19.92	2164	75.88
4	521	18.27	2685	94.14
5	47	1.65	2732	95.79
7	3	0.11	2735	95.90
8	117	4.10	2852	100.00

Momenty			
N	2638	Suma wag	2638
Średnia	2.55913571	Suma obserwacji	6751
Odchylenie std.	0.93397403	Wariancja	0.87230748
Skośność	0.54471151	Kurtoza	-0.5193974
Niesk. suma kw.	19577	Skoryg. suma kw.	2300.27483
Wsp. zmienności	36.4956818	Błąd std. śr.	0.01818433

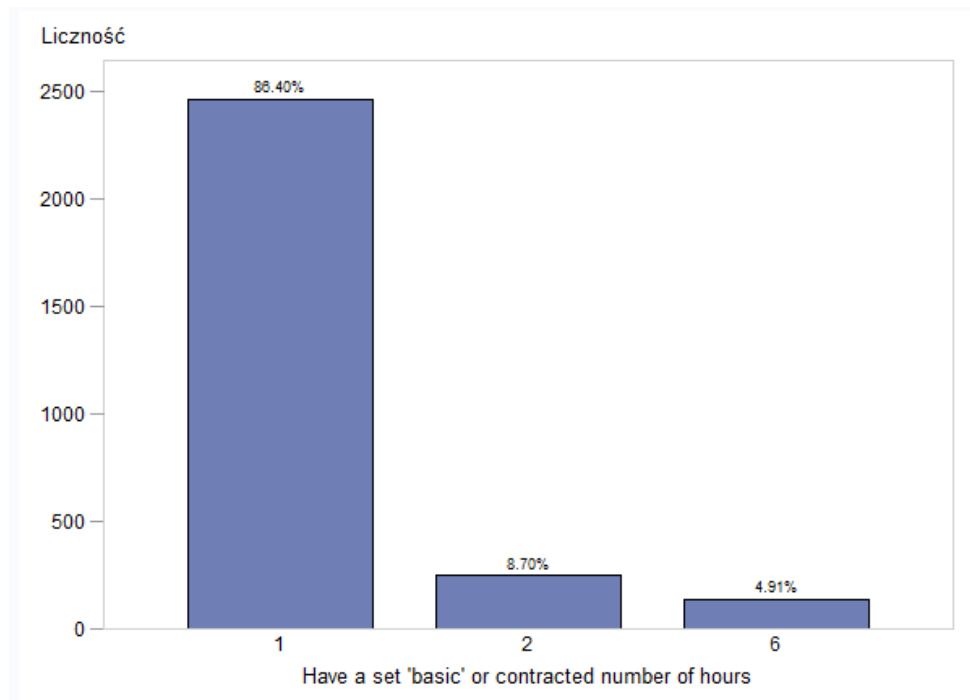
Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	2.559136	Odchylenie std.	0.93397
Mediana	2.000000	Wariancja	0.87231
Moda	2.000000	Rozstęp	4.00000
		Rozstęp międzykwartkowy	1.00000

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 poluzniona. 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



Have a set 'basic' or contracted number of hours				
icwhct	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
1	2464	86.40	2464	86.40
2	248	8.70	2712	95.09
6	140	4.91	2852	100.00

Momenty			
N	2638	Suma wag	2638
Średnia	1.29416224	Suma obserwacji	3414
Odchylenie std.	1.02448783	Wariancja	1.04957531
Skośność	4.08927303	Kurtoza	15.802592
Niesk. suma kw.	7186	Skoryg. suma kw.	2767.7301
Wsp. zmienności	79.1622406	Błąd std. śr.	0.01994662

Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	1.294162	Odchylenie std.	1.02449
Mediana	1.000000	Wariancja	1.04958
Moda	1.000000	Rozstęp	5.00000
		Rozstęp międzykwartkowy	0

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.

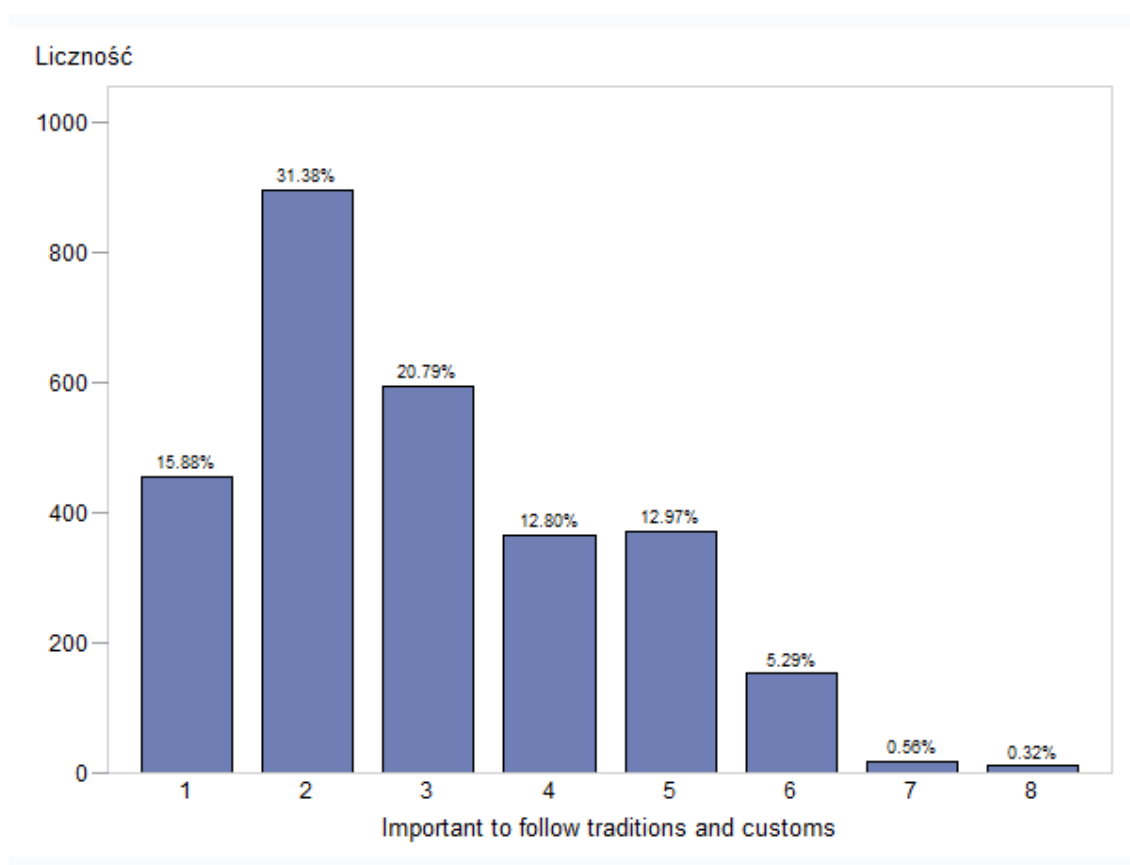
Badaczka Alexandra Killewald porównała pary małżeńskie, które pobrały się po 1975 roku. Okazało się, że związki, w których partner nie pracował w pełnym wymiarze godzin (np. na cały etat), miały o 3,3% większe

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



Important to follow traditions and customs				
imptrad	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
1	453	15.88	453	15.88
2	895	31.38	1348	47.27
3	593	20.79	1941	68.06
4	365	12.80	2306	80.86
5	370	12.97	2676	93.83
6	151	5.29	2827	99.12
7	16	0.56	2843	99.68
8	9	0.32	2852	100.00

Momenty			
N	2638	Suma wag	2638
Średnia	2.92304776	Suma obserwacji	7711
Odchylenie std.	1.4493573	Wariancja	2.10063659
Skośność	0.51268674	Kurtoza	-0.7501327
Niesk. suma kw.	28079	Skoryg. suma kw.	5539.3787
Wsp. zmienności	49.5837708	Błąd std. śr.	0.02821877

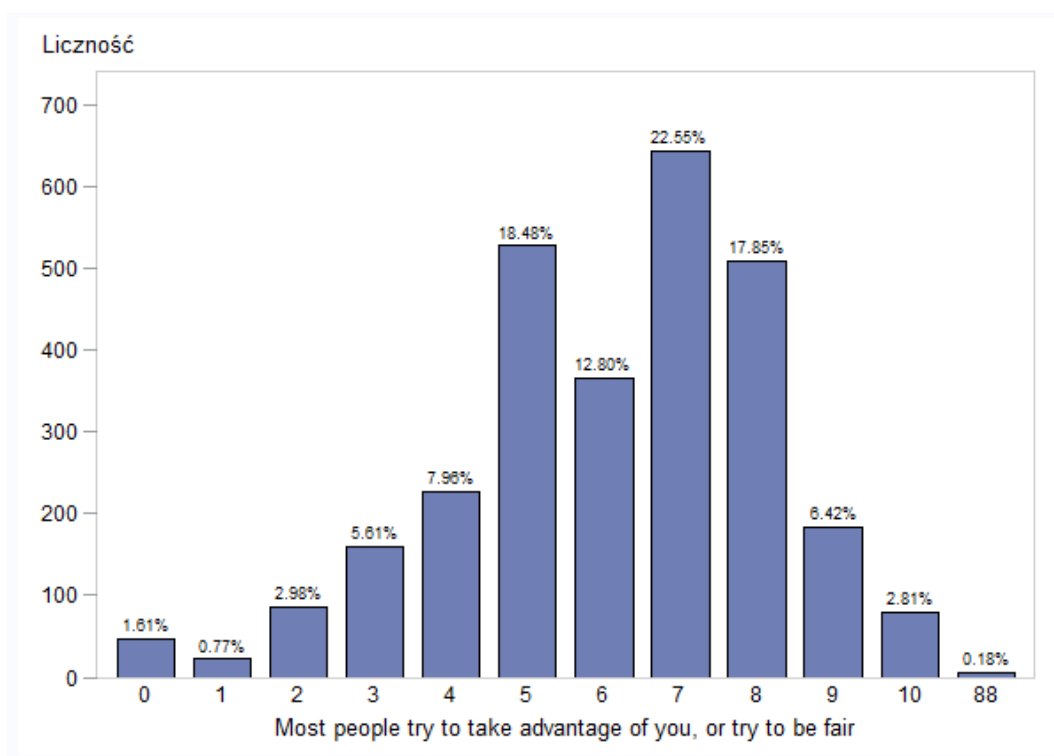
Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	2.923048	Odchylenie std.	1.44936
Mediana	3.000000	Wariancja	2.10064
Moda	2.000000	Rozstęp	5.00000
		Rozstęp międzykwartkowy	2.00000

Najczęściej jako przyczynę rozvodu 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



Most people try to take advantage of you, or try to be fair				
pplfair	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
0	46	1.61	46	1.61
1	22	0.77	68	2.38
2	85	2.98	153	5.36
3	160	5.61	313	10.97
4	227	7.96	540	18.93
5	527	18.48	1067	37.41
6	365	12.80	1432	50.21
7	643	22.55	2075	72.76
8	509	17.85	2584	90.60
9	183	6.42	2767	97.02
10	80	2.81	2847	99.82
88	5	0.18	2852	100.00

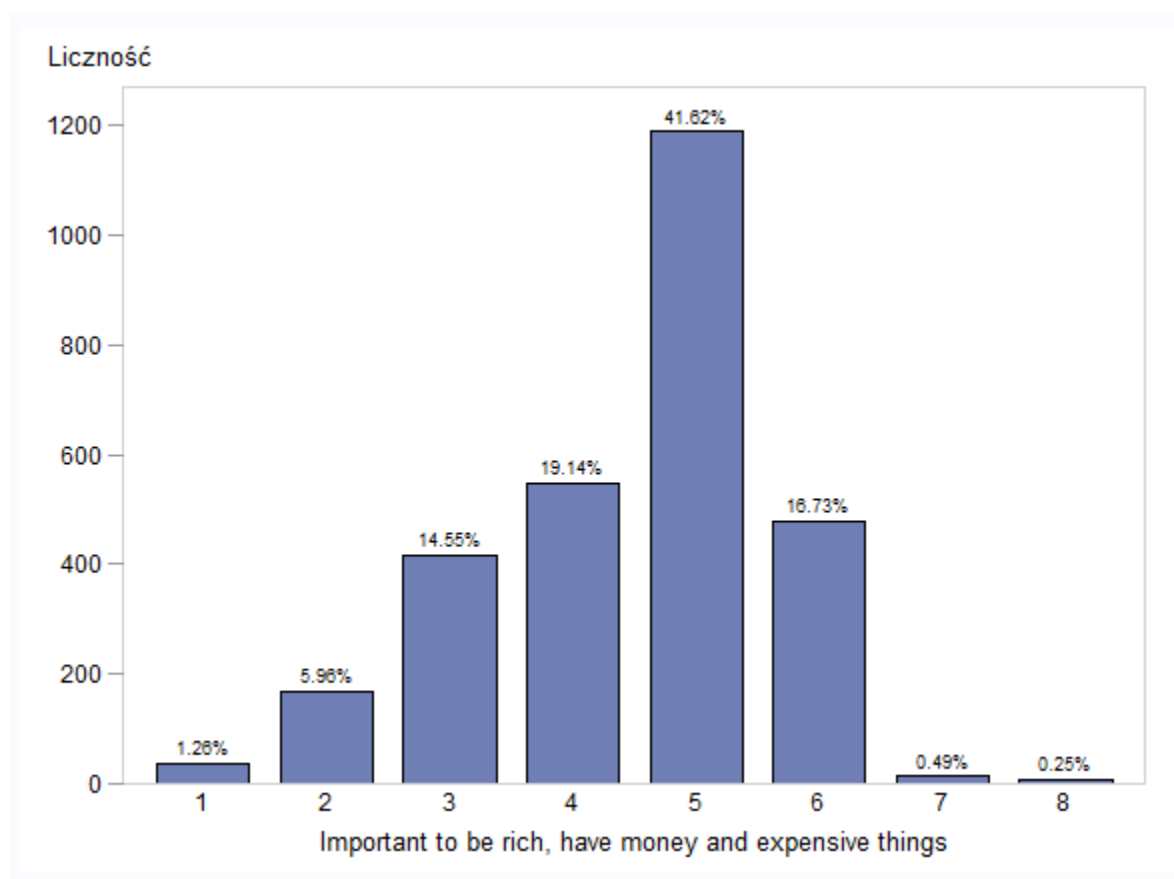
Momenty			
N	2638	Suma wag	2638
Średnia	6.13495072	Suma obserwacji	16184
Odchylenie std.	2.03756507	Wariancja	4.15167142
Skośność	-0.5974262	Kurtoza	0.24779593
Niesk. suma kw.	110236	Skoryg. suma kw.	10947.9575
Wsp. zmienności	33.2124114	Błąd std. śr.	0.03967109

Bazowe miary statystyczne			
Położenie		Zmienność	
Średnia	6.134951	Odchylenie std.	2.03757
Mediana	6.500000	Wariancja	4.15167
Moda	7.000000	Rozstęp	10.00000
		Rozstęp międzykwartkowy	3.00000

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



Important to be rich, have money and expensive things				
imprich	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
1	36	1.26	36	1.26
2	170	5.96	206	7.22
3	415	14.55	621	21.77
4	546	19.14	1167	40.92
5	1187	41.62	2354	82.54
6	477	16.73	2831	99.26
7	14	0.49	2845	99.75
8	7	0.25	2852	100.00

Momenty			
N	2638	Suma wag	2638
Średnia	4.45792267	Suma obserwacji	11760
Odchylenie std.	1.17114036	Wariancja	1.37156974
Skośność	-0.7243754	Kurtoza	-0.0537181
Niesk. suma kw.	56042	Skoryg. suma kw.	3616.82942
Wsp. zmienności	26.2709887	Błąd std. śr.	0.02280193

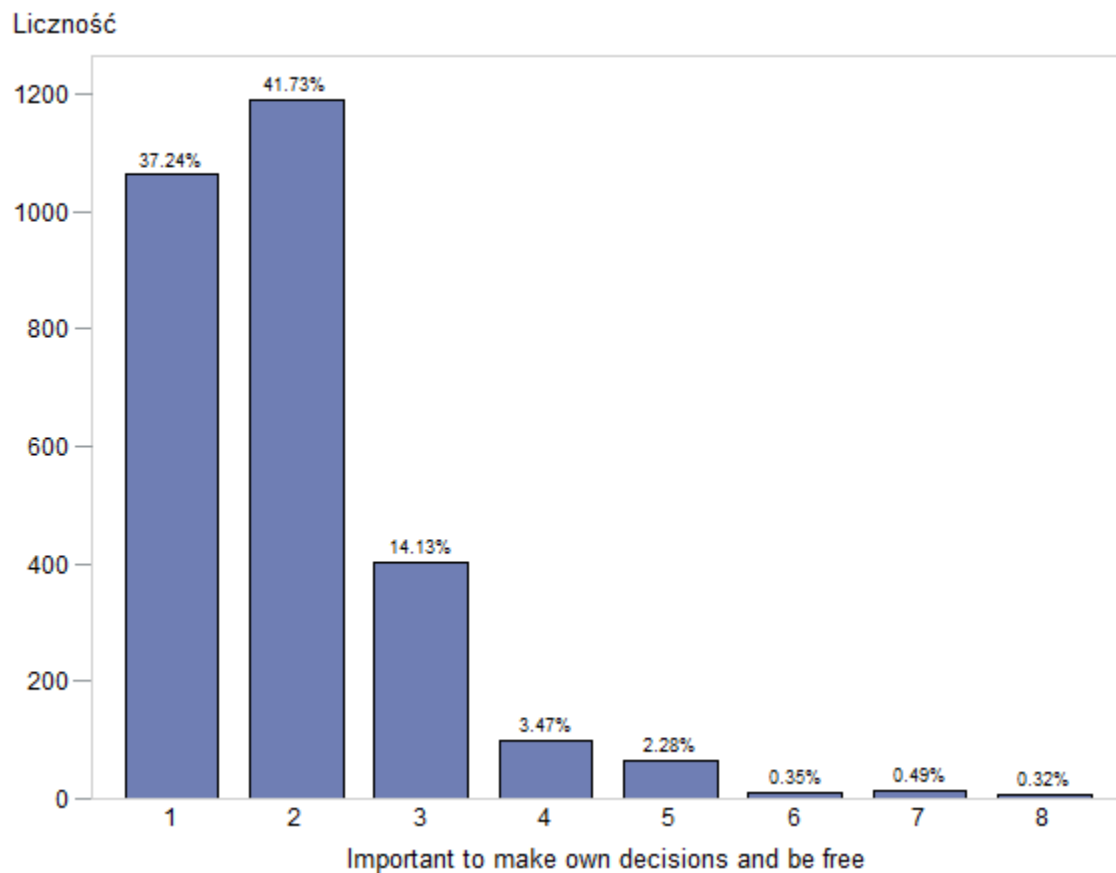
Bazowe miary statystyczne			
Polozenie		Zmienność	
Średnia	4.457923	Odchylenie std.	1.17114
Mediana	5.000000	Wariancja	1.37157
Moda	5.000000	Rozstęp	5.00000
		Rozstęp międzykwartkowy	1.00000

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



Important to make own decisions and be free				
impfree	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
1	1062	37.24	1062	37.24
2	1190	41.73	2252	78.96
3	403	14.13	2655	93.09
4	99	3.47	2754	96.56
5	65	2.28	2819	98.84
6	10	0.35	2829	99.19
7	14	0.49	2843	99.68
8	9	0.32	2852	100.00

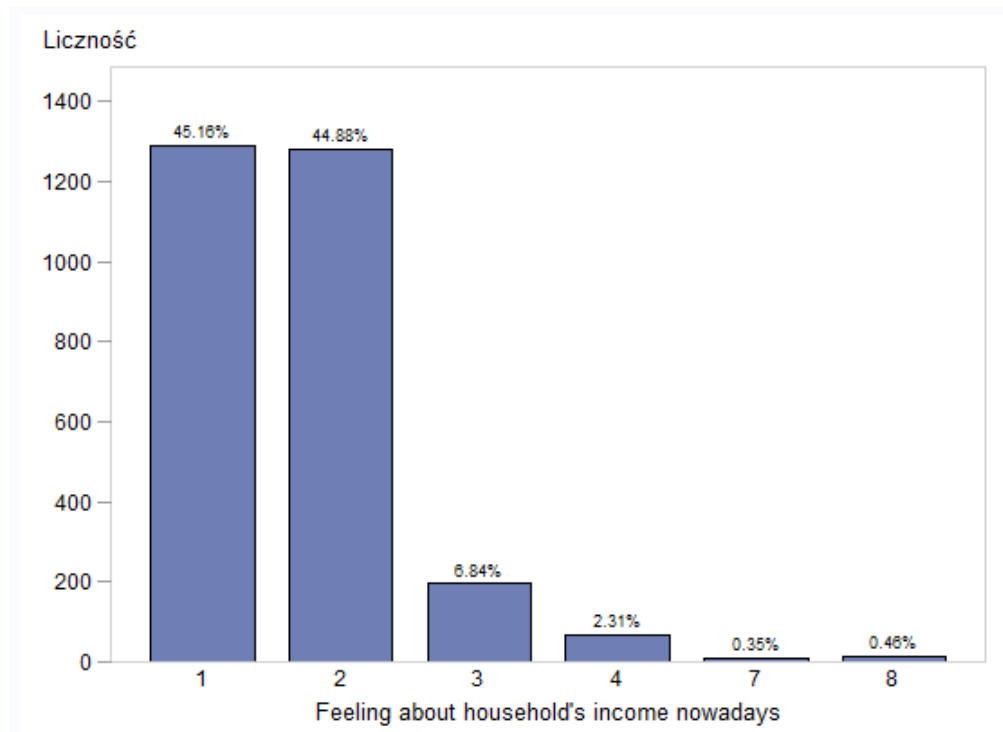
Momenty			
N	2638	Suma wag	2638
Średnia	1.89461713	Suma obserwacji	4998
Odchylenie std.	0.92056593	Wariancja	0.84744162
Skośność	1.15856603	Kurtoza	1.46061471
Niesk. suma kw.	11704	Skoryg. suma kw.	2234.70356
Wsp. zmienności	48.5884937	Błąd std. śr.	0.01792328

Bazowe miary statystyczne			
Polozenie		Zmienność	
Średnia	1.894617	Odchylenie std.	0.92057
Mediana	2.000000	Wariancja	0.84744
Moda	2.000000	Rozstęp	4.00000
		Rozstęp międzykwartkowy	1.00000

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



Feeling about household's income nowadays				
hincfel	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
1	1288	45.16	1288	45.16
2	1280	44.88	2568	90.04
3	195	6.84	2763	96.88
4	66	2.31	2829	99.19
7	10	0.35	2839	99.54
8	13	0.46	2852	100.00

Momenty			
N	2638	Suma wag	2638
Średnia	1.6459439	Suma obserwacji	4342
Odchylenie std.	0.70140214	Wariancja	0.49196496
Skośność	0.99977812	Kurtoza	1.0705197
Niesk. suma kw.	8444	Skoryg. suma kw.	1297.3116
Wsp. zmienności	42.6139761	Błąd std. śr.	0.01365619

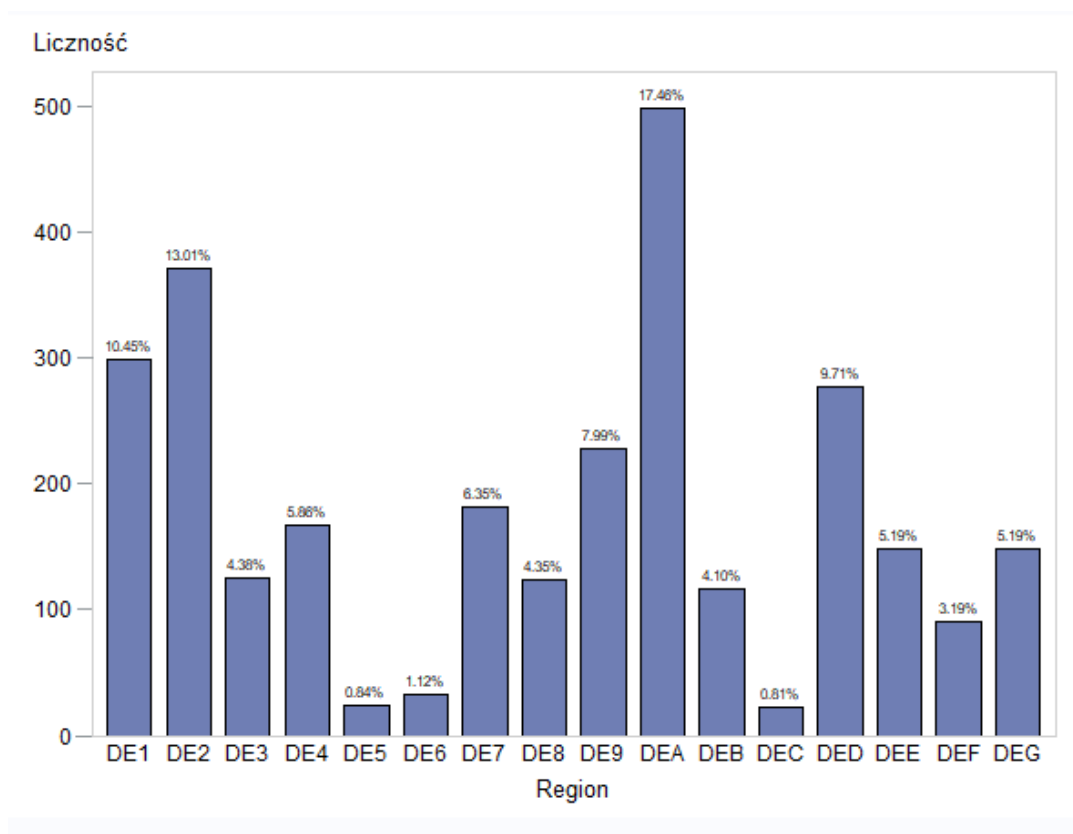
Bazowe miary statystyczne			
Polożenie		Zmienność	
Średnia	1.645944	Odchylenie std.	0.70140
Mediana	2.000000	Wariancja	0.49196
Moda	1.000000	Rozstęp	3.00000
		Rozstęp międzykwartkowy	1.00000

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 rozdzielanie 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

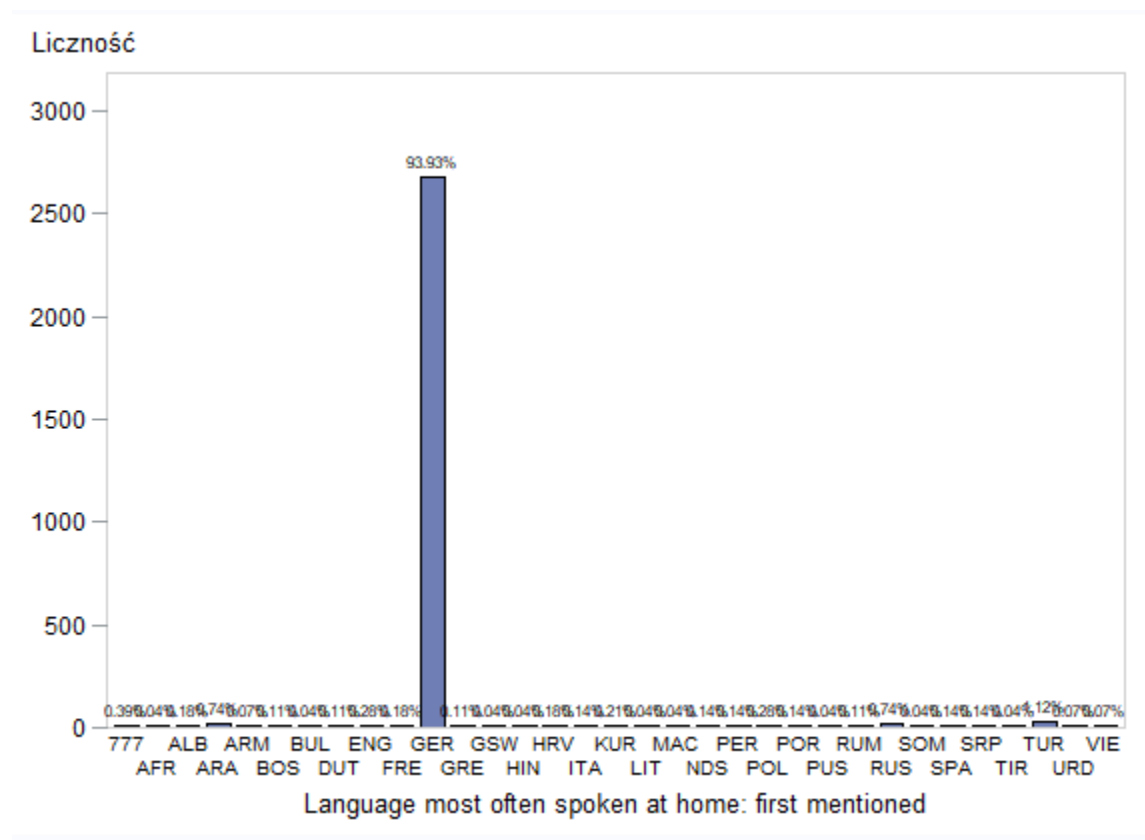


Region				
region	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
DE1	298	10.45	298	10.45
DE2	371	13.01	669	23.46
DE3	125	4.38	794	27.84
DE4	167	5.86	961	33.70
DE5	24	0.84	985	34.54
DE6	32	1.12	1017	35.66
DE7	181	6.35	1198	42.01
DE8	124	4.35	1322	46.35
DE9	228	7.99	1550	54.35
DEA	498	17.46	2048	71.81
DEB	117	4.10	2165	75.91
DEC	23	0.81	2188	76.72
DED	277	9.71	2465	86.43
DEE	148	5.19	2613	91.62
DEF	91	3.19	2704	94.81
DEG	148	5.19	2852	100.00

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

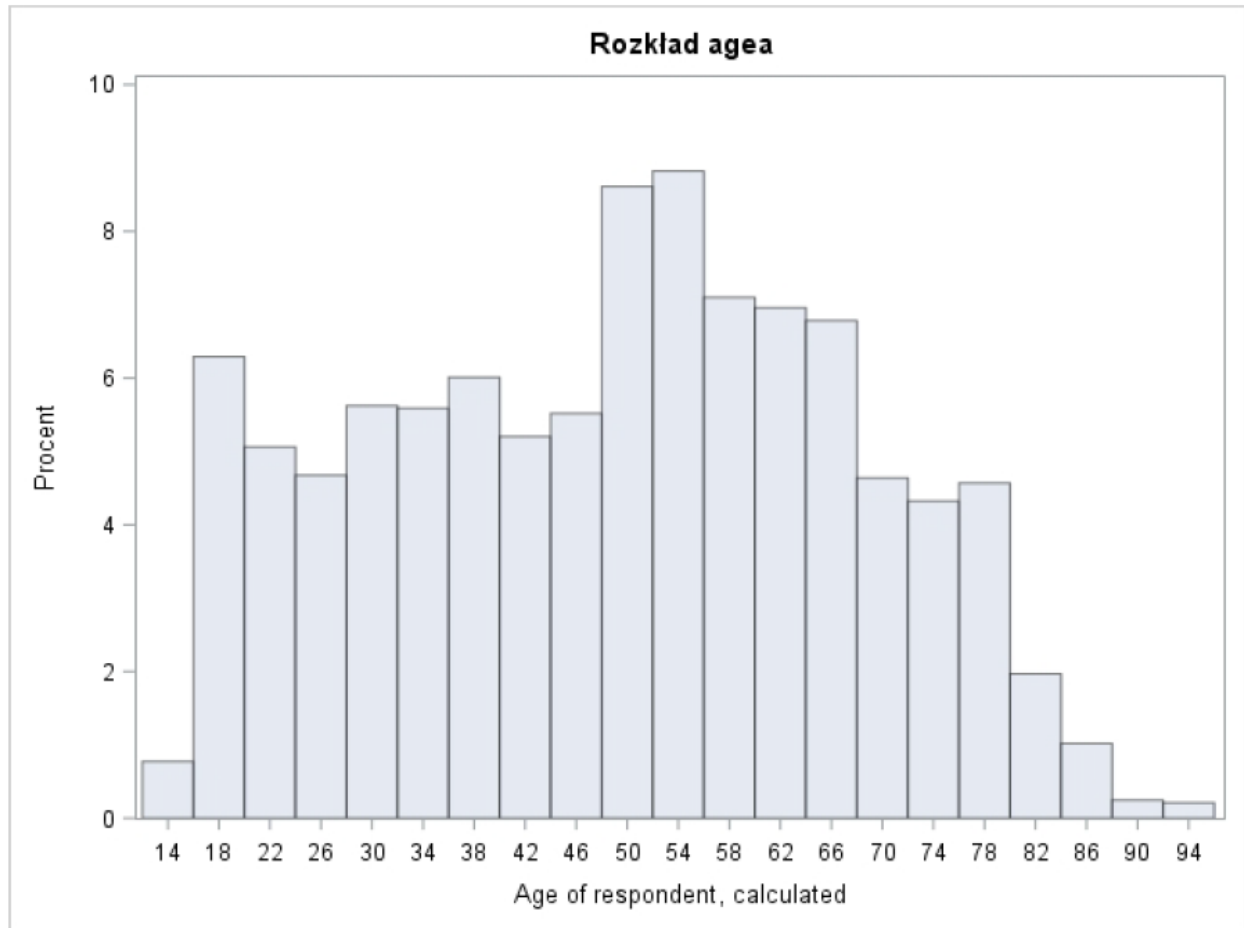


Jak możemy zaobserwować na wykresie, najwięcej osób mówi w domu w języku niemieckim - jest to aż 93.93%. Kolejnymi najliczniejszymi grupami są Turecki (TUR, 1.12%), Arabski (ARA) i Rosyjski (RUS) - po 0.74%.

8. Kategoryzacja zmiennych

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

Agea



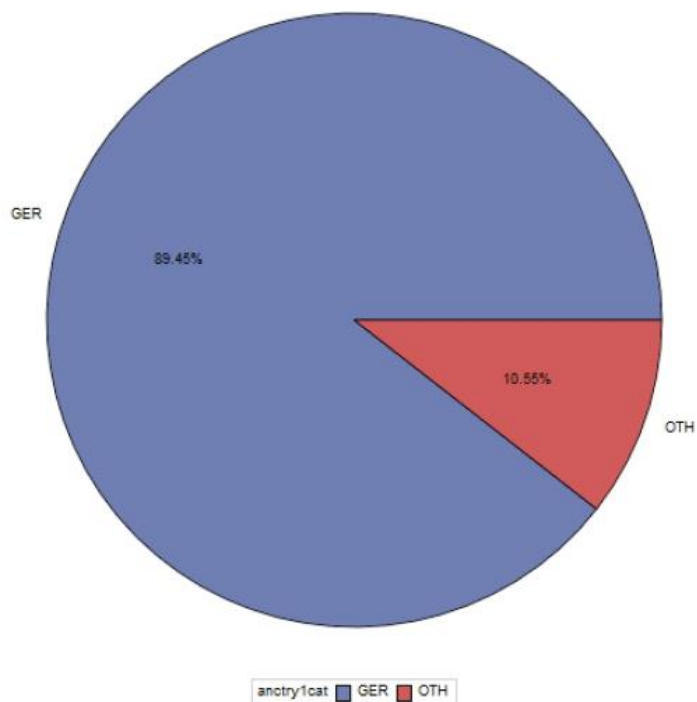
Najmłodszy uczestnik badania miał 15 lat, zaś najstarszy 94 lata.

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 oznaczmy odpowiednio jako “GER” i “OTH”.

Hhmb

Tabela hhmb od dvrcdeva					
		dvrcdeva(Ever been divorced/had civil union dissolved)			
		1	2	7	Razem
hhmb(Number of people living regularly as member of household)					
1	Liczebność	143	371	2	516
	Proc. wier.	27.71	71.90	0.39	
2	Liczebność	213	903	2	1118
	Proc. wier.	19.05	80.77	0.18	
3	Liczebność	85	455	0	540
	Proc. wier.	15.74	84.26	0.00	
4	Liczebność	38	425	1	464
	Proc. wier.	8.19	91.59	0.22	
5	Liczebność	7	134	1	142
	Proc. wier.	4.93	94.37	0.70	
6	Liczebność	4	40	0	44
	Proc. wier.	9.09	90.91	0.00	
7	Liczebność	0	18	0	18
	Proc. wier.	0.00	100.00	0.00	
8	Liczebność	1	5	0	6
	Proc. wier.	16.67	83.33	0.00	
9	Liczebność	0	1	0	1
	Proc. wier.	0.00	100.00	0.00	
10	Liczebność	0	2	0	2
	Proc. wier.	0.00	100.00	0.00	
77	Liczebność	0	0	1	1
	Proc. wier.	0.00	0.00	100.00	
Razem	Liczebność	491	2354	7	2852

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

Rlgatnd

Tabela rlgatnd od dvrcdeva					
		dvrcdeva(Ever been divorced/had civil union dissolved)			
rlgatnd(How often attend religious services apart from special occasions)		1	2	7	Razem
1	Liczebność	1	6	0	7
	Proc. wier.	14.29	85.71	0.00	
2	Liczebność	8	31	0	39
	Proc. wier.	20.51	79.49	0.00	
3	Liczebność	20	154	0	174
	Proc. wier.	11.49	88.51	0.00	
4	Liczebność	46	228	0	274
	Proc. wier.	16.79	83.21	0.00	
5	Liczebność	73	489	2	564
	Proc. wier.	12.94	86.70	0.35	
6	Liczebność	95	442	2	539
	Proc. wier.	17.63	82.00	0.37	
7	Liczebność	248	1003	1	1252
	Proc. wier.	19.81	80.11	0.08	
77	Liczebność	0	1	1	2
	Proc. wier.	0.00	50.00	50.00	
88	Liczebność	0	0	1	1
	Proc. wier.	0.00	0.00	100.00	
Razem	Liczebność	491	2354	7	2852

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

Sclmeet

Rezultaty analizy kontyngencji					
Procedura FREQ					
Tabela sclmeet od dvrcdeva					
		dvrcdeva(Ever been divorced/had civil union dissolved)			
		1	2	7	Razem
sclmeet(How often socially meet with friends, relatives or colleagues)					
1	Liczebność	9	14	0	23
	Proc. wier.	39.13	60.87	0.00	
2	Liczebność	43	114	0	157
	Proc. wier.	27.39	72.61	0.00	
3	Liczebność	58	210	0	268
	Proc. wier.	21.64	78.36	0.00	
4	Liczebność	136	577	2	715
	Proc. wier.	19.02	80.70	0.28	
5	Liczebność	101	428	0	529
	Proc. wier.	19.09	80.91	0.00	
6	Liczebność	104	760	1	865
	Proc. wier.	12.02	87.86	0.12	
7	Liczebność	39	251	3	293
	Proc. wier.	13.31	85.67	1.02	
77	Liczebność	1	0	1	2
	Proc. wier.	50.00	0.00	50.00	
Razem	Liczebność	491	2354	7	2852

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

Rezultaty analizy kontyngencji

Procedura FREQ

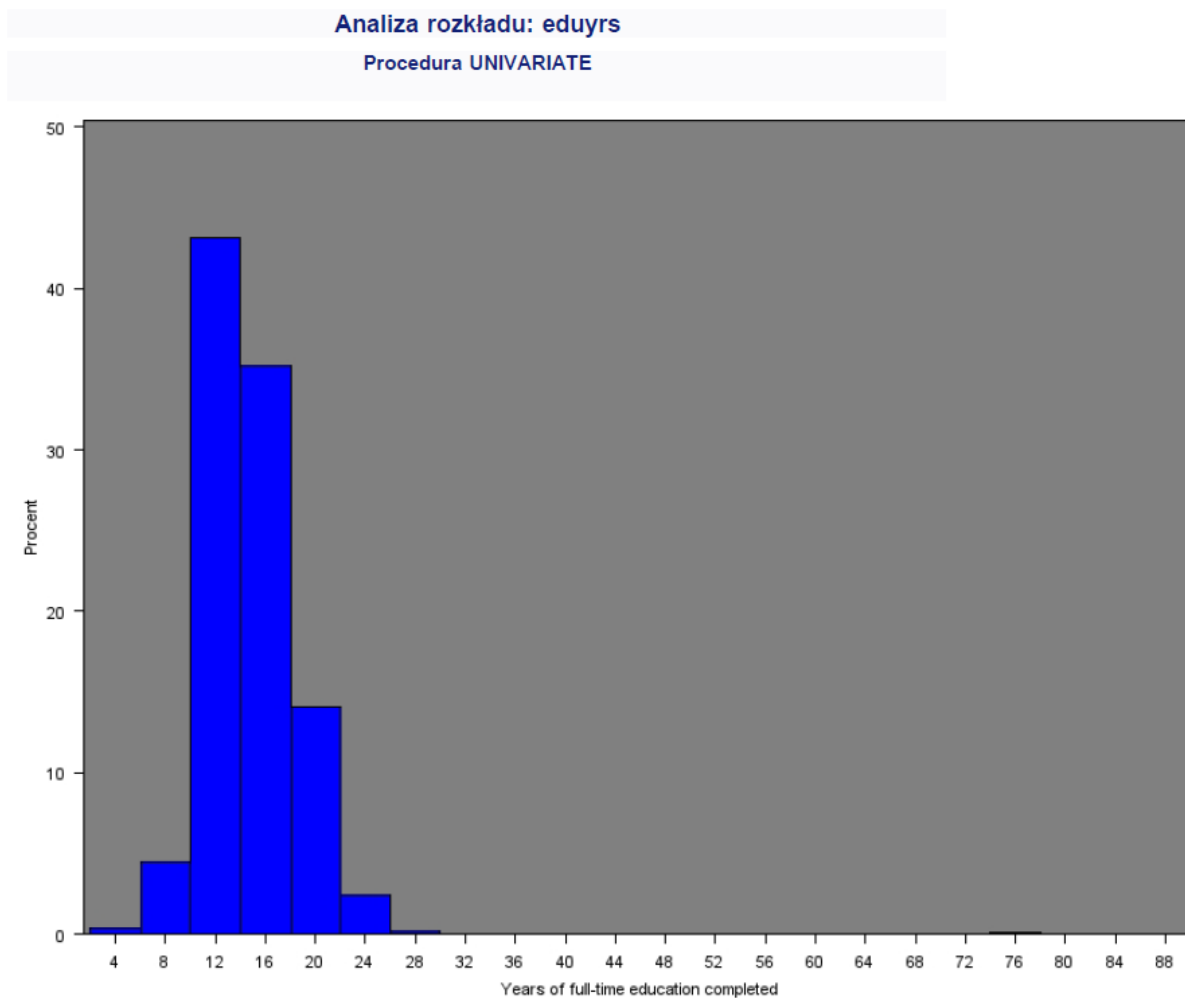
Tabela rlgdgr od dvrcdeva					
		dvrcdeva(Ever been divorced/had civil union dissolved)			
		1	2	7	Razem
rlgdgr(How religious are you)					
0	Liczebność	151	527	0	678
	Proc. wier.	22.27	77.73	0.00	
1	Liczebność	21	150	0	171
	Proc. wier.	12.28	87.72	0.00	
2	Liczebność	34	189	0	223
	Proc. wier.	15.25	84.75	0.00	
3	Liczebność	40	212	0	252
	Proc. wier.	15.87	84.13	0.00	
4	Liczebność	24	147	0	171
	Proc. wier.	14.04	85.96	0.00	
5	Liczebność	56	339	2	397
	Proc. wier.	14.11	85.39	0.50	
6	Liczebność	47	197	1	245
	Proc. wier.	19.18	80.41	0.41	
7	Liczebność	31	232	1	264
	Proc. wier.	11.74	87.88	0.38	
8	Liczebność	51	217	0	268
	Proc. wier.	19.03	80.97	0.00	
9	Liczebność	15	59	0	74
	Proc. wier.	20.27	79.73	0.00	
10	Liczebność	21	82	1	104
	Proc. wier.	20.19	78.85	0.96	
77	Liczebność	0	1	2	3
	Proc. wier.	0.00	33.33	66.67	
88	Liczebność	0	2	0	2
	Proc. wier.	0.00	100.00	0.00	
Razem	Liczebność	491	2354	7	2852

Wartości z zakresu od 0 - niereligijny do 10 - bardzo religijny. Analizując tablicę kontyngencji możemy wydzielić 3 kategorie - 0 jako osobę bardzo "niereligijną", 1-5 określa człowieka "umiarkowanie" religijną, a od 6 do 10 "religijną". 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.

eduysr



Analizując wykres rozkładu zdecydowaliśmy się podzielić zmienną eduysr 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

Tabela hlthhmp od dvrcdeva				
		dvrcdeva(Ever been divorced/had civil union dissolved)		
		1	2	7
hlthhmp(Hampered in daily activities by illness/disability/infirmary/mental problem)				
1	Liczebność	62	133	1
	Proc. wier.	31.63	67.86	0.51
2	Liczebność	153	597	0
	Proc. wier.	20.40	79.60	0.00
3	Liczebność	276	1622	5
	Proc. wier.	14.50	85.23	0.26
7	Liczebność	0	2	1
	Proc. wier.	0.00	66.67	33.33
Razem		Liczebność	491	2354
				7
				2852

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

Tabela stflife od dvrdeva					
		dvrdeva(Ever been divorced/had civil union dissolved)			Razem
		1	2	7	
stflife(How satisfied with life as a whole)					
0	Liczebność	10	12	0	22
	Proc. wier.	45.45	54.55	0.00	
1	Liczebność	5	11	0	16
	Proc. wier.	31.25	68.75	0.00	
2	Liczebność	5	33	0	38
	Proc. wier.	13.16	86.84	0.00	
3	Liczebność	31	51	0	82
	Proc. wier.	37.80	62.20	0.00	
4	Liczebność	16	62	0	78
	Proc. wier.	20.51	79.49	0.00	
5	Liczebność	54	163	0	217
	Proc. wier.	24.88	75.12	0.00	
6	Liczebność	38	147	1	186
	Proc. wier.	20.43	79.03	0.54	
7	Liczebność	71	368	1	440
	Proc. wier.	16.14	83.64	0.23	
8	Liczebność	130	691	1	822
	Proc. wier.	15.82	84.06	0.12	
9	Liczebność	80	465	1	546
	Proc. wier.	14.65	85.16	0.18	
10	Liczebność	50	347	2	399
	Proc. wier.	12.53	86.97	0.50	
77	Liczebność	1	3	1	5
	Proc. wier.	20.00	60.00	20.00	
88	Liczebność	0	1	0	1
	Proc. wier.	0.00	100.00	0.00	
Razem	Liczebność	491	2354	7	2852

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

Tabela edctn od dvrdeva					
		dvrdeva(Ever been divorced/had civil union dissolved)			Razem
		1	2	7	
edctn(Doing last 7 days: education)					
0	Liczebność	485	2043	6	2534
	Proc. wier.	19.14	80.62	0.24	
1	Liczebność	6	311	1	318
	Proc. wier.	1.89	97.80	0.31	
Razem	Liczebność	491	2354	7	2852

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

Jbspv

Tabela jbspv od dvrdeva					
		dvrdeva(Ever been divorced/had civil union dissolved)			Razem
		1	2	7	
jbspv(Responsible for supervising other employees)					
1	Liczebność	192	810	0	1002
	Proc. wier.	19.16	80.84	0.00	
2	Liczebność	293	1404	4	1701
	Proc. wier.	17.23	82.54	0.24	
6	Liczebność	4	134	2	140
	Proc. wier.	2.86	95.71	1.43	
7	Liczebność	2	4	1	7
	Proc. wier.	28.57	57.14	14.29	
8	Liczebność	0	2	0	2
	Proc. wier.	0.00	100.00	0.00	
Razem	Liczebność	491	2354	7	2852

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

Uemp3m

Tabela uemp3m od dvrdeva					
		dvrdeva(Ever been divorced/had civil union dissolved)			Razem
		1	2	7	
uemp3m(Ever unemployed and seeking work for a period more than three months)					
1	Liczebność	203	615	1	819
	Proc. wier.	24.79	75.09	0.12	
2	Liczebność	285	1733	4	2022
	Proc. wier.	14.09	85.71	0.20	
7	Liczebność	2	3	2	7
	Proc. wier.	28.57	42.86	28.57	
8	Liczebność	1	3	0	4
	Proc. wier.	25.00	75.00	0.00	
Razem	Liczebność	491	2354	7	2852

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

Tabela chldhhe od dvrdeva					
		dvrdeva(Ever been divorced/had civil union dissolved)			Razem
		1	2	7	
chldhhe(Ever had children living in household)					
1	Liczebność	295	690	0	985
	Proc. wier.	29.95	70.05	0.00	
2	Liczebność	58	918	1	977
	Proc. wier.	5.94	93.96	0.10	
6	Liczebność	138	745	1	884
	Proc. wier.	15.61	84.28	0.11	
7	Liczebność	0	1	5	6
	Proc. wier.	0.00	16.67	83.33	
Razem	Liczebność	491	2354	7	2852

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

Tabela emplrel od dvrcdeva					
		dvrcdeva(Ever been divorced/had civil union dissolved)			Razem
		1	2	7	
emplrel(Employment relation)					
1	Liczebność	412	1943	2	2357
	Proc. wier.	17.48	82.44	0.08	
2	Liczebność	61	233	1	295
	Proc. wier.	20.68	78.98	0.34	
3	Liczebność	11	40	1	52
	Proc. wier.	21.15	76.92	1.92	
6	Liczebność	4	134	2	140
	Proc. wier.	2.86	95.71	1.43	
7	Liczebność	2	3	1	6
	Proc. wier.	33.33	50.00	16.67	
8	Liczebność	1	1	0	2
	Proc. wier.	50.00	50.00	0.00	
Razem	Liczebność	491	2354	7	2852

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

Tabela bennent od dvrcdeva					
		dvrcdeva(Ever been divorced/had civil union dissolved)			Razem
		1	2	7	
bennent(Many manage to obtain benefits/services not entitled to)					
1	Liczebność	51	157	0	208
	Proc. wier.	24.52	75.48	0.00	
2	Liczebność	260	1126	2	1388
	Proc. wier.	18.73	81.12	0.14	
3	Liczebność	89	478	1	568
	Proc. wier.	15.67	84.15	0.18	
4	Liczebność	67	454	0	521
	Proc. wier.	12.86	87.14	0.00	
5	Liczebność	10	37	0	47
	Proc. wier.	21.28	78.72	0.00	
7	Liczebność	0	0	3	3
	Proc. wier.	0.00	0.00	100.00	
8	Liczebność	14	102	1	117
	Proc. wier.	11.97	87.18	0.85	
Razem	Liczebność	491	2354	7	2852

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.

lcwhct

Tabela lcwhct od dvrdeva					
		dvrdeva(Ever been divorced/had civil union dissolved)			
		1	2	7	Razem
lcwhct(Have a set 'basic' or contracted number of hours)					
1	Liczebność	437	2024	3	2464
	Proc. wier.	17.74	82.14	0.12	
2	Liczebność	50	196	2	248
	Proc. wier.	20.16	79.03	0.81	
6	Liczebność	4	134	2	140
	Proc. wier.	2.86	95.71	1.43	
Razem		Liczebność	491	2354	7 2852

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

Tabela imptrad od dvrdeva					
		dvrdeva(Ever been divorced/had civil union dissolved)			
		1	2	7	Razem
imptrad(Important to follow traditions and customs)					
1	Liczebność	97	356	0	453
	Proc. wier.	21.41	78.59	0.00	
2	Liczebność	147	746	2	895
	Proc. wier.	16.42	83.35	0.22	
3	Liczebność	92	500	1	593
	Proc. wier.	15.51	84.32	0.17	
4	Liczebność	58	306	1	365
	Proc. wier.	15.89	83.84	0.27	
5	Liczebność	68	302	0	370
	Proc. wier.	18.38	81.62	0.00	
6	Liczebność	23	128	0	151
	Proc. wier.	15.23	84.77	0.00	
7	Liczebność	6	7	3	16
	Proc. wier.	37.50	43.75	18.75	
8	Liczebność	0	9	0	9
	Proc. wier.	0.00	100.00	0.00	
Razem		Liczebność	491	2354	7 2852

Na podstawie analizy tablic kontyngencji widzimy, że zmienna Imptrad nie ma wpływu na zmienną dvrdeva - 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

Rezultaty analizy kontyngencji

Procedura FREQ

Tabela pplfair od dvrcdeva					
		dvrcdeva(Ever been divorced/had civil union dissolved)			
		1	2	7	Razem
pplfair(Most people try to take advantage of you, or try to be fair)					
0	Liczebność	13	32	1	46
	Proc. wier.	28.26	69.57	2.17	
1	Liczebność	7	15	0	22
	Proc. wier.	31.82	68.18	0.00	
2	Liczebność	23	62	0	85
	Proc. wier.	27.06	72.94	0.00	
3	Liczebność	21	139	0	160
	Proc. wier.	13.13	86.88	0.00	
4	Liczebność	52	175	0	227
	Proc. wier.	22.91	77.09	0.00	
5	Liczebność	94	431	2	527
	Proc. wier.	17.84	81.78	0.38	
6	Liczebność	41	323	1	365
	Proc. wier.	11.23	88.49	0.27	
7	Liczebność	96	547	0	643
	Proc. wier.	14.93	85.07	0.00	
8	Liczebność	91	416	2	509
	Proc. wier.	17.88	81.73	0.39	
9	Liczebność	35	147	1	183
	Proc. wier.	19.13	80.33	0.55	
10	Liczebność	18	62	0	80
	Proc. wier.	22.50	77.50	0.00	
88	Liczebność	0	5	0	5
	Proc. wier.	0.00	100.00	0.00	
Razem	Liczebność	491	2354	7	2852

Wygenerowane przez System SAS ('Local', X64_8HOME) dnia 23 maja 2019 o godz. 11:20:34 PM

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

Rezultaty analizy kontyngencji					
Procedura FREQ					
Tabela imprich od dvrcdeva					
		dvrcdeva(Ever been divorced/had civil union dissolved)			Razem
		1	2	7	
imprich(Important to be rich, have money and expensive things)					
1	Liczebność	4	32	0	36
	Proc. wier.	11.11	88.89	0.00	
2	Liczebność	21	149	0	170
	Proc. wier.	12.35	87.65	0.00	
3	Liczebność	50	365	0	415
	Proc. wier.	12.05	87.95	0.00	
4	Liczebność	91	454	1	546
	Proc. wier.	16.67	83.15	0.18	
5	Liczebność	224	960	3	1187
	Proc. wier.	18.87	80.88	0.25	
6	Liczebność	96	381	0	477
	Proc. wier.	20.13	79.87	0.00	
7	Liczebność	5	6	3	14
	Proc. wier.	35.71	42.86	21.43	
8	Liczebność	0	7	0	7
	Proc. wier.	0.00	100.00	0.00	
Razem	Liczebność	491	2354	7	2852

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" przyjmujemy osoby, dla których bogactwo jest istotne (1-3), a za "no" tych, których pieniądze nie interesują (odpowiedzi 4-6)

Impfree

Rezultaty analizy kontyngencji

Procedura FREQ

Tabela impfree od dvrcdeva					
		dvrcdeva(Ever been divorced/had civil union dissolved)			Razem
		1	2	7	
impfree(Important to make own decisions and be free)					
1	Liczebność	229	831	2	1062
	Proc. wier.	21.56	78.25	0.19	
2	Liczebność	167	1021	2	1190
	Proc. wier.	14.03	85.80	0.17	
3	Liczebność	57	346	0	403
	Proc. wier.	14.14	85.86	0.00	
4	Liczebność	16	83	0	99
	Proc. wier.	16.16	83.84	0.00	
5	Liczebność	16	49	0	65
	Proc. wier.	24.62	75.38	0.00	
6	Liczebność	1	9	0	10
	Proc. wier.	10.00	90.00	0.00	
7	Liczebność	5	6	3	14
	Proc. wier.	35.71	42.86	21.43	
8	Liczebność	0	9	0	9
	Proc. wier.	0.00	100.00	0.00	
Razem	Liczebność	491	2354	7	2852

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

Rezultaty analizy kontyngencji

Procedura FREQ

Tabela hincfel od dvrcdeva						
		dvrcdeva(Ever been divorced/had civil union dissolved)				
		1	2	7		
hincfel(Feeling about household's income nowadays)					Razem	
1	Liczebność	180	1105	3	1288	
	Proc. wier.	13.98	85.79	0.23		
2	Liczebność	241	1038	1	1280	
	Proc. wier.	18.83	81.09	0.08		
3	Liczebność	49	145	1	195	
	Proc. wier.	25.13	74.36	0.51		
4	Liczebność	19	47	0	66	
	Proc. wier.	28.79	71.21	0.00		
7	Liczebność	1	7	2	10	
	Proc. wier.	10.00	70.00	20.00		
8	Liczebność	1	12	0	13	
	Proc. wier.	7.69	92.31	0.00		
Razem		Liczebność	491	2354	7	2852

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

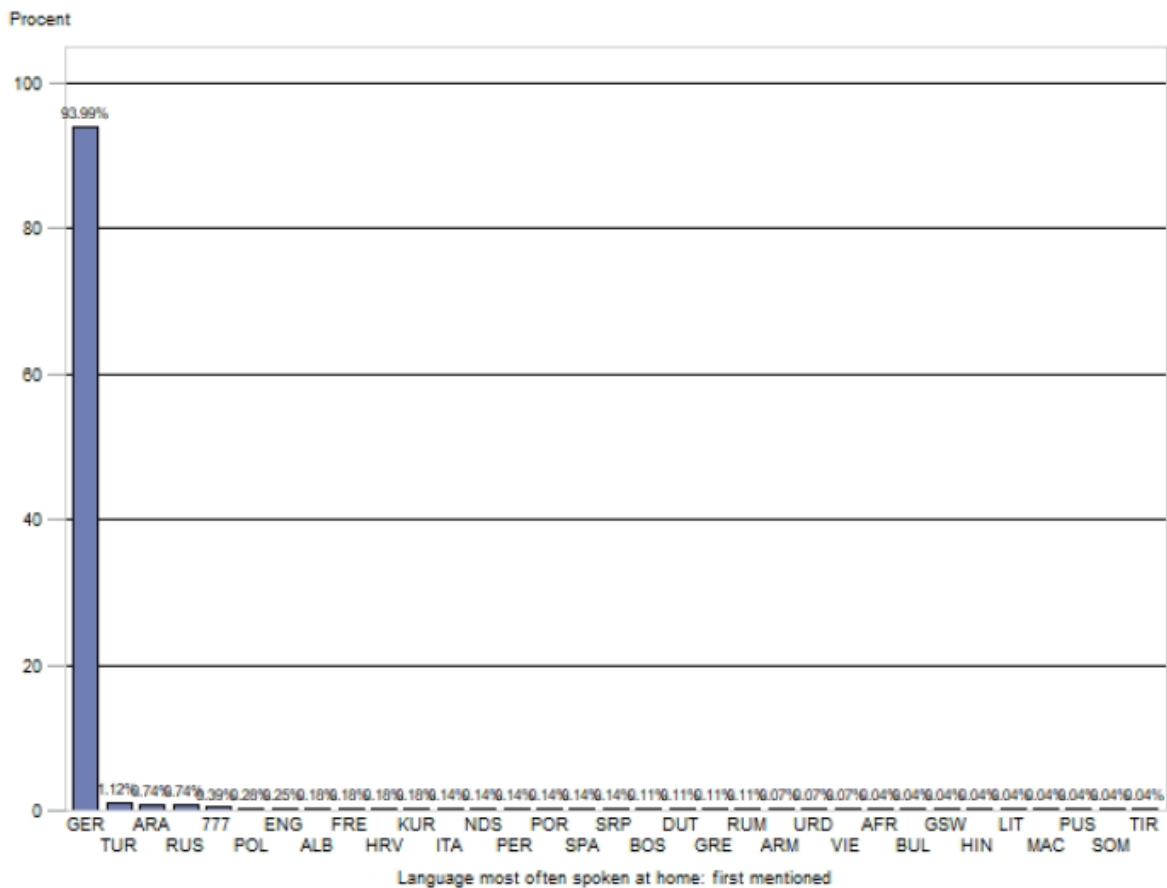
Jedną z naszych tez badawczych jest sprawdzenie, czy miejsce zamieszkania *ceteris paribus* ma wpływ na wzrost lub spadek liczby rozwodów. W szczególności interesuje nas, czy osoby zamieszkujące tereny dawnego RFN rozwodzą się częściej, niż zamieszkujący obszar dawnego NRD. W związku z tym musimy skategoryzować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

Wykres słupkowy



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 **dvrccdeva**, 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.
- **Hhmmbcac** - skategoryzowana liczba mieszkańców gospodarstwa domowego. Jeśli osoba badana założyła rodzinę bądź mieszka sama, ma ona wpływ na jej wartość.
- **Eduyrscac** - 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.
- **Jbspvcac** - 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 miesiące. Na zmienną tą można było wpłynąć swoim poglądem na pracę (czy ktoś jest lekkoduchem czy też karierowiczem).
- **Imptrad** - opinia dotycząca przywiązania do tradycji

Rezultaty regresji logistycznej		
Procedura LOGISTIC		
Informacje o modelu		
Zbiór	WORK.SORTTEMPTABLESORTED	
Zmienna objaśniana	dvrdeva	Ever been divorced/had civil union dissolved
Liczba poziomów odpowiedzi	2	
Model	logit binarny	
Technika optymalizacji	Ocena Fishera	
Wczytano obserwacji		2638
Użyto obserwacji		2638
Profil odpowiedzi		
Wartość uporządkowana	dvrdeva	Całkowita liczebność
1	1	463
2	2	2175
Modelowane prawdopodobieństwo wynosi dvrdeva=1.		

Źródło: wydruk z programu SAS

Powyżej został wygenerowany raport opisujący model regresji logistycznej. Do opisanie 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.

Informacje o poziomach klasyfikacji		
Klasa	Wartość	Zmienne planu
imptrad	1	1 0 0 0 0
	2	0 1 0 0 0
	3	0 0 1 0 0
	4	0 0 0 1 0
	5	0 0 0 0 1
	6	0 0 0 0 0
rlgatnd	1	1 0 0 0 0 0
	2	0 1 0 0 0 0
	3	0 0 1 0 0 0
	4	0 0 0 1 0 0
	5	0 0 0 0 1 0
	6	0 0 0 0 0 1
uemp3m	1	1
	2	0
chldhhe	1	1 0
	2	0 1
	6	0 0
hhmmbcat	high	1
	low	0
eduyrscat	basic	1 0
	higher	0 1
	medium	0 0

Informacje o poziomach klasyfikacji				
Klasa	Wartość	Zmienne planu		
stflifecat	no	1		
	yes	0		
Jbspvcat	1	1		
	2	0		
Bennentcat	Extreme	1		
	Middle	0		
Empleirecat	Applicable	1		
	Not_applicable	0		
lcwhctcat	answer	1		
	no answer	0		
Scimeetcat	high	1	0	
	low	0	1	
	medium	0	0	
Rlgdgrcat	high	1	0	
	medium	0	1	
	no	0	0	
Pplfaircat	cagey	1		
	trustful	0		
Imprichcat	no	1		
	yes	0		
Impfreecat	high	1	0	
	low	0	1	
	medium	0	0	
Nwspolcat	medium	1	0	
	often	0	1	
	rarely	0	0	
regioncat	NRD	1		
	RFN	0		

Źródło: wydruk z programu SAS

W tabelach powyżej przedstawiono zmienne objaśniające użyte do modelu.

Status zbieżności				
Kryterium zbieżności (GCONV=1E-8) spełnione.				
Statystyki dobroci dopasowania dewiencji i Pearsona				
Kryterium	Wartość	DF	Wartość/DF	Pr. > chi-kw.
Dewiencja	2066.3442	2540	0.8135	1.0000
Pearson	2449.1885	2540	0.9642	0.8998
Liczba unikatowych profilów: 2574				

Źródło: wydruk z programu SAS

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

Statystyki dopasowania		
Kryterium	Tylko wyraz wolny	Wyraz wolny i współzmiennie
AIC	2452.803	2170.388
SC	2458.680	2370.232
-2 log L	2450.803	2102.388

R-kwadrat	0.1237	Maksymalnie przeskalowane R-kwadrat	0.2045
-----------	--------	-------------------------------------	--------

Testowanie globalnej hipotezy zerowej: BETA=0			
Test	Chi-kwadrat	DF	Pr. > chi-kw.
Iloraz wiarygod.	348.4147	33	<.0001
Wynik punktowy	315.5712	33	<.0001
Wald	261.3706	33	<.0001

Źródło: wydruk z programu SAS

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

Analiza efektów typu 3			
Efekt	DF	Chi-kwadrat	Pr. > chi-kw.
		Walda	
imptrad	5	2.5741	0.7653
rlgatnd	6	10.2373	0.1150
uemp3m	1	18.2931	<.0001
chldhhe	2	99.2284	<.0001
hhmmbcat	1	21.5017	<.0001
eduyrscat	2	10.3491	0.0057
stflifecat	1	8.9051	0.0028
Jbspvcacat	1	0.7056	0.4009
Bennentcat	1	0.5924	0.4415
Emplrelcat	1	3.7268	0.0535
lcwhctcat	1	0.3228	0.5699
ScImeetcat	2	8.8282	0.0121
Rlgdgrcat	2	9.8385	0.0073
Pplfaircat	1	1.3602	0.2435
Imprichcat	1	5.0727	0.0243
Impfreecat	2	17.2357	0.0002
Nwspolcat	2	0.1652	0.9207
regioncat	1	1.8529	0.1734

Źródło: wydruk z programu SAS

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

Test zgodności Hosmera i Lemeshowa		
Chi-kwadrat	DF	Pr. > chi-kw.
10.6363	8	0.2232

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 częściowo wynika z własnego 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.
- **hmmmbcat** - 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 sytuację, 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.
- **Emprelcat** - 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

Informacje o modelu		
Zbiór	WORK.SORTTEMPTABLESORTED	
Zmienna objaśniana	dvrdeva	Ever been divorced/had civil union dissolved
Liczba poziomów odpowiedzi	2	
Model	logit binarny	
Technika optymalizacji	Ocena Fishera	

Źródło: wydruk z programu SAS

Wczytano obserwacji	2638
Użyto obserwacji	2638

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:

Profil odpowiedzi		
Wartość uporządkowana	dvrdeva	Całkowita liczebność
1	1	463
2	2	2175

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:

Informacje o poziomach klasyfikacji							
Klasa	Wartość	Zmienne planu					
imptrad	1	1	0	0	0	0	
	2	0	1	0	0	0	
	3	0	0	1	0	0	
	4	0	0	0	1	0	
	5	0	0	0	0	1	
	6	0	0	0	0	0	
rlgatnd	1	1	0	0	0	0	0
	2	0	1	0	0	0	0
	3	0	0	1	0	0	0
	4	0	0	0	1	0	0
	5	0	0	0	0	1	0
	6	0	0	0	0	0	1
	7	0	0	0	0	0	0
hlthhmp	1	1	0				
	2	0	1				
	3	0	0				
chldhhe	1	1	0				
	2	0	1				
	6	0	0				
Inghom1cat	GER	1					
	OTH	0					
regioncat	NRD	1					
	RFN	0					
anctry1cat	GER	1					
	OTH	0					
hhmmbcac	high	1					
	low	0					
eduyscat	basic	1	0				
	higher	0	1				
	medium	0	0				
stflifecat	no	1					
	yes	0					
Rlgdgrcat	high	1	0				
	medium	0	1				
	no	0	0				

Źródło: wydruk z programu SAS

Powyższa tabela prezentuje sposób kodowania zmiennych referencyjnych wykorzystanych w modelu.

Status zbieżności:

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

Źródło: wydruk z programu SAS

Statystyki dopasowania dewiancji i Pearsona:

Statystyki dobroci dopasowania dewiancji i Pearsona				
Kryterium	Wartość	DF	Wartość/DF	Pr. > chi-kw.
Dewiancja	1324.8208	1517	0.8733	0.9999
Pearson	1583.4583	1517	1.0438	0.1148

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 oceny dobroci dopasowania modelu.

Statystyki dopasowania:

Statystyki dopasowania		
Kryterium	Tylko wyraz wolny	Wyraz wolny i współzmiennie
AIC	2452.803	2205.286
SC	2458.680	2352.231
-2 log L	2450.803	2155.286

Źródło: wydruk z programu SAS

Statystyka R kwadrat:

R-kwadrat	0.1060	Maksymalnie przeskalowane R-kwadrat	0.1751
-----------	--------	-------------------------------------	--------

Źródło: wydruk z programu SAS

Testowanie globalnej hipotezy zerowej BETA=0

Testowanie globalnej hipotezy zerowej: BETA=0			
Test	Chi-kwadrat	DF	Pr. > chi-kw.
Iloraz wiarygod.	295.5163	24	<.0001
Wynik punktowy	274.8015	24	<.0001
Wald	234.0312	24	<.0001

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:

Analiza efektów typu 3			
Efekt	DF	Chi-kwadrat Walda	Pr. > chi-kw.
imptrad	5	4.3821	0.4958
rlgatnd	6	14.7553	0.0222
hlthhmp	2	9.3176	0.0095
chldhhe	2	115.2278	<.0001
Inghom1cat	1	0.4378	0.5082
regioncat	1	1.0650	0.3021
anctry1cat	1	0.0161	0.8991
hhmmbcat	1	24.8969	<.0001
eduyrscat	2	12.1182	0.0023
stflifecat	1	11.3513	0.0008
Rlgdgrcat	2	10.5539	0.0051

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, Inghom1cat, regioncat, anctry1cat). Powyższe wyniki mogą stanowić podstawę wykluczenia tych zmiennych z modelu.

Analiza ocen maksymalnej wiarygodności:

Analiza ocen maksymalnej wiarygodności						
Parametr		DF	Ocena	Błąd standardowy	Chi-kwadrat Walda	Pr. > chi-kw.
Intercept		1	-1.0277	0.4000	6.6020	0.0102
imptrad	1	1	0.3136	0.2879	1.1861	0.2761
imptrad	2	1	0.0691	0.2754	0.0630	0.8018
imptrad	3	1	0.1574	0.2844	0.3064	0.5799
imptrad	4	1	0.2629	0.2984	0.7767	0.3782
imptrad	5	1	0.3426	0.2892	1.4037	0.2361
rigatnd	1	1	-0.3195	1.2311	0.0674	0.7952
rigatnd	2	1	-0.5870	0.4959	1.4012	0.2365
rigatnd	3	1	-1.0997	0.3177	11.9801	0.0005
rigatnd	4	1	-0.3507	0.2262	2.4030	0.1211
rigatnd	5	1	-0.2569	0.1724	2.2204	0.1362
rigatnd	6	1	-0.00993	0.1601	0.0038	0.9506
hlthhmp	1	1	0.5898	0.1936	9.2804	0.0023
hlthhmp	2	1	0.0786	0.1258	0.3906	0.5320
chldhhe	1	1	0.4261	0.1434	8.8258	0.0030
chldhhe	2	1	-1.2873	0.1797	51.3453	<.0001
Inghom1cat	GER	1	-0.2118	0.3201	0.4378	0.5082
regioncat	NRD	1	-0.1297	0.1257	1.0650	0.3021
anctry1cat	GER	1	0.0314	0.2479	0.0161	0.8991
hhmmbcat	high	1	-0.9622	0.1928	24.8969	<.0001
eduyrscat	basic	1	-0.4806	0.2177	4.8719	0.0273
eduyrscat	higher	1	-0.6031	0.2130	8.0194	0.0046
stflifecat	no	1	0.4350	0.1291	11.3513	0.0008
Rlgdgrcat	high	1	-0.0107	0.1808	0.0035	0.9530
Rlgdgrcat	medium	1	-0.3866	0.1518	6.4865	0.0109

Źródło: wydruk z programu SAS

Oceny parametrów i przedziały ufności Walda:

Oceny parametrów i przedziały ufności Walda				
Parametr		Ocena	Przedział ufności 95%	
Intercept		-1.0277	-1.8116	-0.2438
imptrad	1	0.3136	-0.2507	0.8779
imptrad	2	0.0691	-0.4706	0.6089
imptrad	3	0.1574	-0.4000	0.7149
imptrad	4	0.2629	-0.3218	0.8477
imptrad	5	0.3426	-0.2241	0.9093
rlgatnd	1	-0.3195	-2.7325	2.0934
rlgatnd	2	-0.5870	-1.5589	0.3849
rlgatnd	3	-1.0997	-1.7225	-0.4770
rlgatnd	4	-0.3507	-0.7940	0.0927
rlgatnd	5	-0.2569	-0.5949	0.0810
rlgatnd	6	-0.00993	-0.3238	0.3039
hlthhmp	1	0.5898	0.2103	0.9692
hlthhmp	2	0.0786	-0.1679	0.3251
chldhhe	1	0.4261	0.1450	0.7072
chldhhe	2	-1.2873	-1.6394	-0.9352
Inghom1cat	GER	-0.2118	-0.8392	0.4156
regioncat	NRD	-0.1297	-0.3761	0.1167
anctry1cat	GER	0.0314	-0.4544	0.5172
hhmmbcat	high	-0.9622	-1.3401	-0.5842
eduyrscat	basic	-0.4806	-0.9073	-0.0538
eduyrscat	higher	-0.6031	-1.0204	-0.1857
stflifecat	no	0.4350	0.1819	0.6881
Rlgdgrcat	high	-0.0107	-0.3650	0.3437
Rlgdgrcat	medium	-0.3866	-0.6841	-0.0891

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 $\exp(\beta)$ z współczynnikiem $1-\alpha$ zawiera liczbę 1. Jak można zauważyć, w modelu występuje 95% przedział ufności dla zmiennych ($\alpha=0,05$). 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), Inghom1cat, regioncat, anctry1cat i obie zmienne rlgdgrcat.

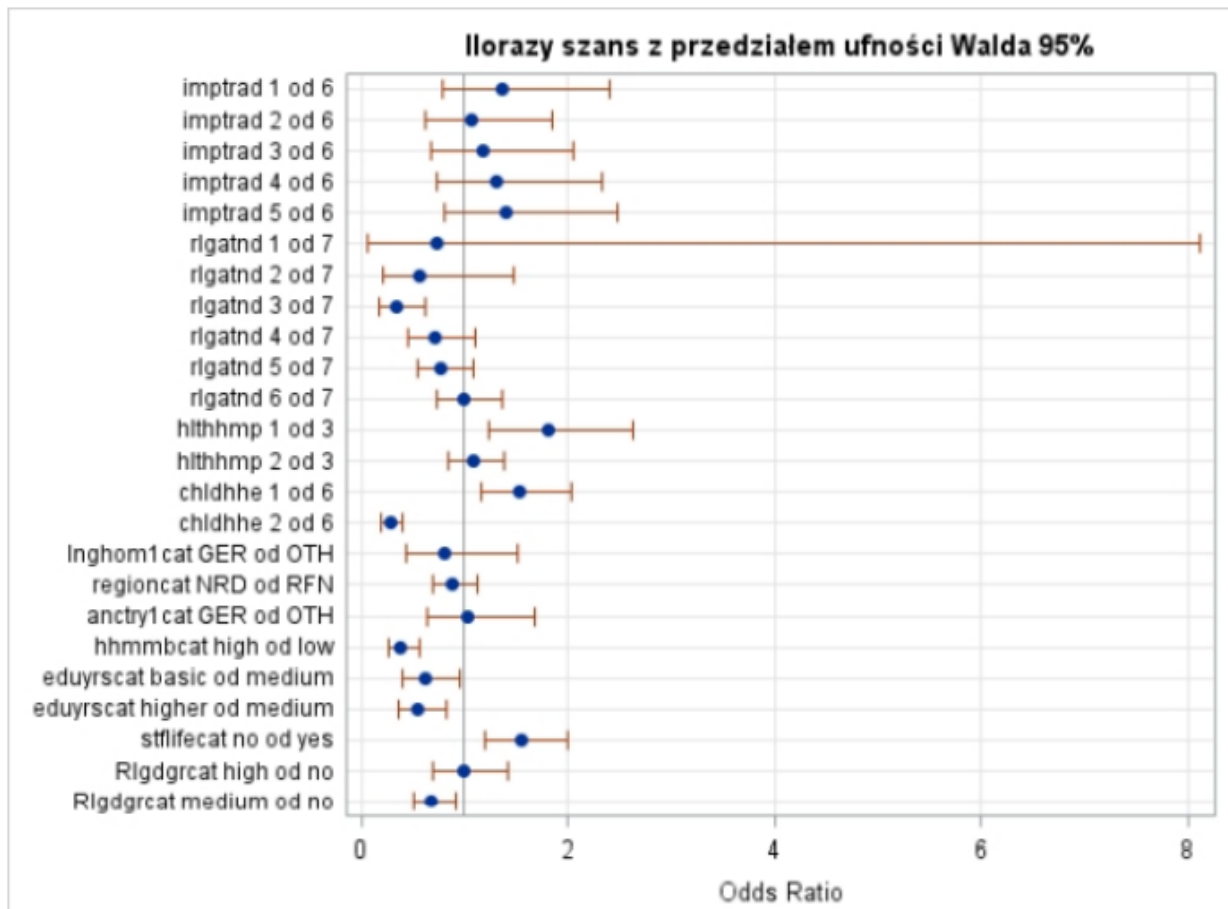
Oceny ilorazów szans i przedziału ufności Walda

Oceny ilorazów szans i przedziały ufności Walda					
Efekt		Jednostka	Ocena	Przedział ufności 95%	
imptrad	1 od 6	1.0000	1.368	0.778	2.406
imptrad	2 od 6	1.0000	1.072	0.625	1.838
imptrad	3 od 6	1.0000	1.171	0.670	2.044
imptrad	4 od 6	1.0000	1.301	0.725	2.334
imptrad	5 od 6	1.0000	1.409	0.799	2.483
rlgatnd	1 od 7	1.0000	0.726	0.065	8.113
rlgatnd	2 od 7	1.0000	0.556	0.210	1.470
rlgatnd	3 od 7	1.0000	0.333	0.179	0.621
rlgatnd	4 od 7	1.0000	0.704	0.452	1.097
rlgatnd	5 od 7	1.0000	0.773	0.552	1.084
rlgatnd	6 od 7	1.0000	0.990	0.723	1.355
hlthhmp	1 od 3	1.0000	1.804	1.234	2.636
hlthhmp	2 od 3	1.0000	1.082	0.845	1.384
chldhhe	1 od 6	1.0000	1.531	1.156	2.028
chldhhe	2 od 6	1.0000	0.276	0.194	0.393
Inghom1cat	GER od OTH	1.0000	0.809	0.432	1.515
regioncat	NRD od RFN	1.0000	0.878	0.687	1.124
anctry1cat	GER od OTH	1.0000	1.032	0.635	1.677
hhmmbcat	high od low	1.0000	0.382	0.262	0.558
eduyrscat	basic od medium	1.0000	0.618	0.404	0.948
eduyrscat	higher od medium	1.0000	0.547	0.360	0.831
stflifecat	no od yes	1.0000	1.545	1.200	1.990
Rlgdgrcat	high od no	1.0000	0.989	0.694	1.410
Rlgdgrcat	medium od no	1.0000	0.679	0.505	0.915

Źródło: wydruk z programu SAS

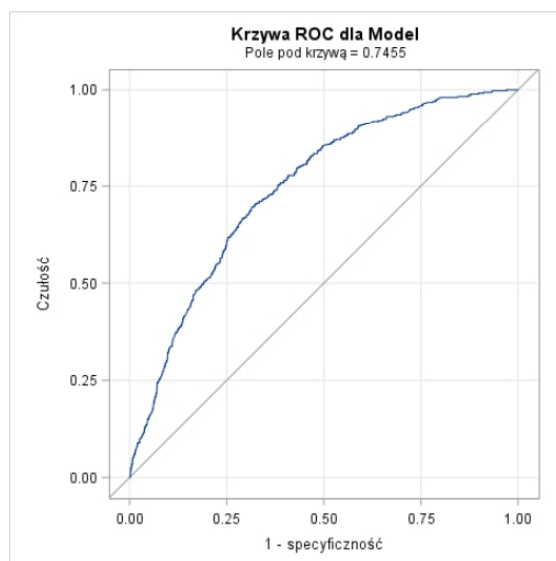
W powyżej tabeli zostały przedstawione oceny ilorazów szans i przedziałów ufności Walda. W tym przypadku można mówić o nieistotności, jeśli przedział zawiera liczbę 1. Wyniki zamieszczone w tabeli powyżej prezentują występowanie istotności dla zmiennych: rlfatnd (3 od 7), hlthhmp (1 od 3), chldhhe (obie 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 $<0,1>$. Oczekiwana 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 decyzję, który z nich wybrać do objaśnienia zmiennej *dvrdeva* (rozwoju). W tym celu wykorzystamy wartości kryterium informacyjnego Akaike'a oraz kryterium Schwarz'a by uzyskać model, który charakteryzuje się największą wiarygodnością.

Model zmiennych światopoglądu:

Statystyki dopasowania		
Kryterium	Tylko wyraz wolny	Wyraz wolny i współzmiennie
AIC	2452.803	2171.634
SC	2458.680	2377.357
-2 log L	2450.803	2101.634

Źródło: wydruk z programu SAS

Model zmiennych środowiska:

Statystyki dopasowania		
Kryterium	Tylko wyraz wolny	Wyraz wolny i współzmiennie
AIC	2452.803	2205.286
SC	2458.680	2352.231
-2 log L	2450.803	2155.286

Źródło: wydruk z programu SAS

Wartości kryterium Schwarz'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:

Podsumowanie selekcji postępującej						
Krok	Efekt wstawiony	DF	Liczba w	Chi-kwadrat punktacji	Pr. > chi-kw.	Etykieta zmiennej
1	chldhhe	2	1	183.9949	<.0001	Ever had children living in household
2	hhmmbcat	1	2	32.4904	<.0001	
3	stflifecat	1	3	19.9131	<.0001	
4	Impfreecat	2	4	23.2500	<.0001	
5	Sclmeetcat	2	5	14.4387	0.0007	
6	eduyrscat	2	6	11.7861	0.0028	
7	Rlgdgrcat	2	7	8.1520	0.0170	
8	Imprichcat	1	8	4.1114	0.0426	
9	Emplrelcat	2	9	5.9927	0.0500	

Źródło: wydruk z programu SAS

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

Testowanie globalnej hipotezy zerowej: BETA=0			
Test	Chi-kwadrat	DF	Pr. > chi-kw.
Iloraz wiarygod.	311.0553	15	<.0001
Wynik punktowy	281.9490	15	<.0001
Wald	237.2904	15	<.0001

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 $\alpha=0.05$.

10.1 Wynik testu zgodności Hosmera i Lemeshowa

Test zgodności Hosmera i Lemeshowa		
Chi-kwadrat	DF	Pr. > chi-kw.
7.1549	8	0.5200

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

Statystyki dobroci dopasowania dewiancji i Pearsona				
Kryterium	Wartość	DF	Wartość/DF	Pr. > chi-kw.
Dewiancja	1864.3443	2247	0.8297	1.0000
Pearson	2211.4982	2247	0.9842	0.6993

Liczba unikatowych profili: 2263

Źródło: wydruk z programu SAS

Zarówno dla statystyk dewiancji jak i Pearsona wartość krytyczna prawdopodobieństwa jest wyższa niż poziom istotności $\alpha = 0.05$. 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

Analiza ocen maksymalnej wiarygodności						
Parametr		DF	Ocena	Błąd standardowy	Chi-kwadrat Walda	Pr. > chi-kw.
Intercept		1	-1.5195	0.2098	52.4515	<.0001
chldhhe	1	1	0.4110	0.1426	8.3043	0.0040
chldhhe	2	1	-1.1594	0.1816	40.7786	<.0001
hhmmbcat	high	1	-0.9500	0.1934	24.1358	<.0001
eduyrscat	basic	1	-0.4301	0.2175	3.9103	0.0480
eduyrscat	higher	1	-0.5932	0.2124	7.8022	0.0052
stflifecat	no	1	0.4833	0.1248	15.0043	0.0001
Emplrelcat	Business	1	0.1567	0.1561	1.0076	0.3155
Emplrelcat	Not_applicable	1	-1.1209	0.5366	4.3631	0.0367
ScImeetcat	high	1	-0.2906	0.1229	5.5905	0.0181
ScImeetcat	low	1	0.3706	0.1989	3.4724	0.0624
Rlgdgrcat	high	1	-0.1828	0.1406	1.6903	0.1936
Rlgdgrcat	medium	1	-0.3902	0.1350	8.3573	0.0038
Imprichcat	no	1	0.3085	0.1482	4.3348	0.0373
Impfreecat	high	1	0.4918	0.1128	19.0204	<.0001
Impfreecat	low	1	0.7685	0.3228	5.6671	0.0173

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.

Oceny ilorazów szans i przedziały ufności Walda				
Efekt	Jednostka	Ocena	Przedział ufności 95%	
chldhhe 1 od 6	1.0000	1.508	1.141	1.995
chldhhe 2 od 6	1.0000	0.314	0.220	0.448
hhmmbcat high od low	1.0000	0.387	0.265	0.565
eduysrcat basic od medium	1.0000	0.650	0.425	0.996
eduysrcat higher od medium	1.0000	0.553	0.364	0.838
stflifecat no od yes	1.0000	1.621	1.270	2.071
Emplrelcat Business od Position	1.0000	1.170	0.861	1.588
Emplrelcat Not_applicable od Position	1.0000	0.326	0.114	0.933
Scmeetcat high od medium	1.0000	0.748	0.588	0.952
Scmeetcat low od medium	1.0000	1.449	0.981	2.139
Rlgdgrcat high od no	1.0000	0.833	0.632	1.097
Rlgdgrcat medium od no	1.0000	0.677	0.520	0.882
Imprichcat no od yes	1.0000	1.361	1.018	1.820
Impfreecat high od medium	1.0000	1.635	1.311	2.040
Impfreecat low od medium	1.0000	2.157	1.145	4.060

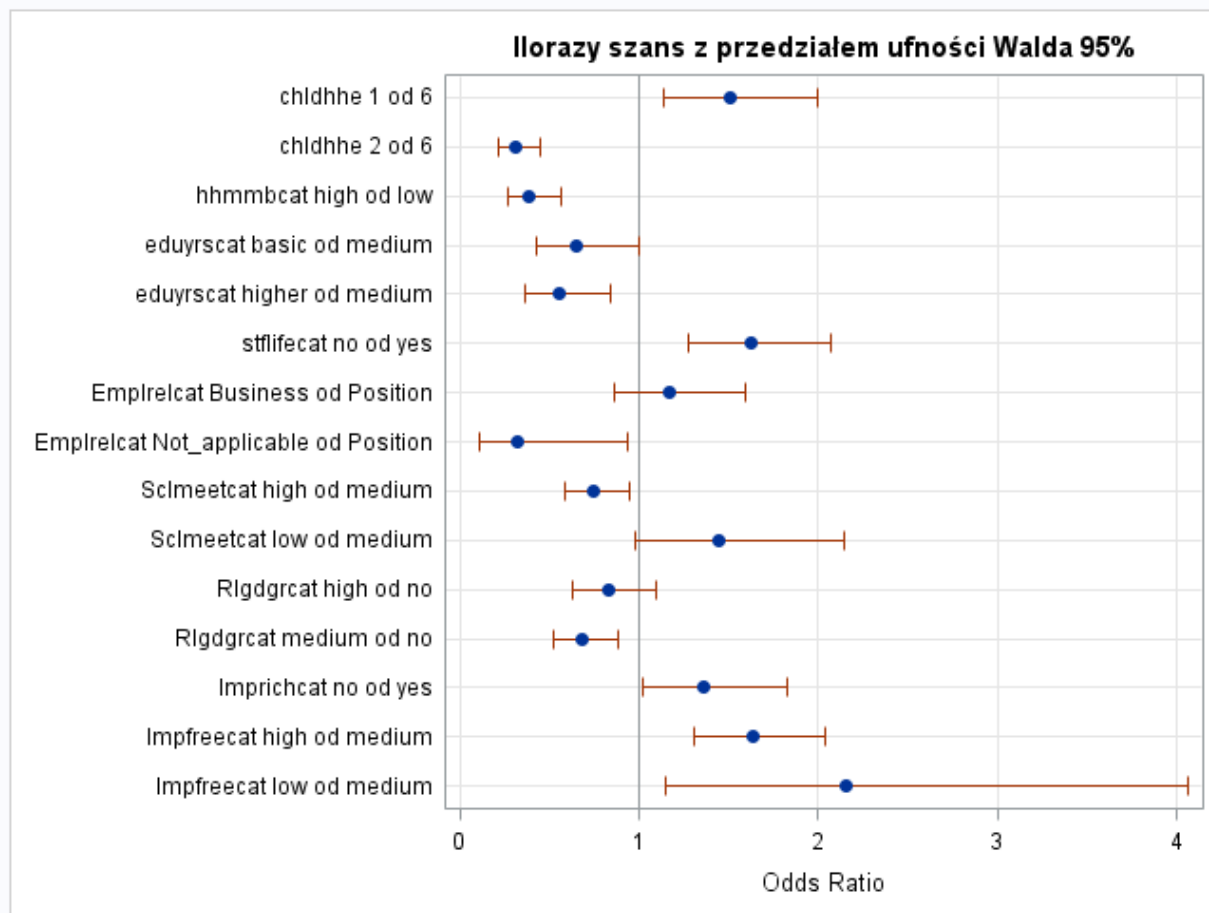
Źródło: wydruk z programu SAS

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.

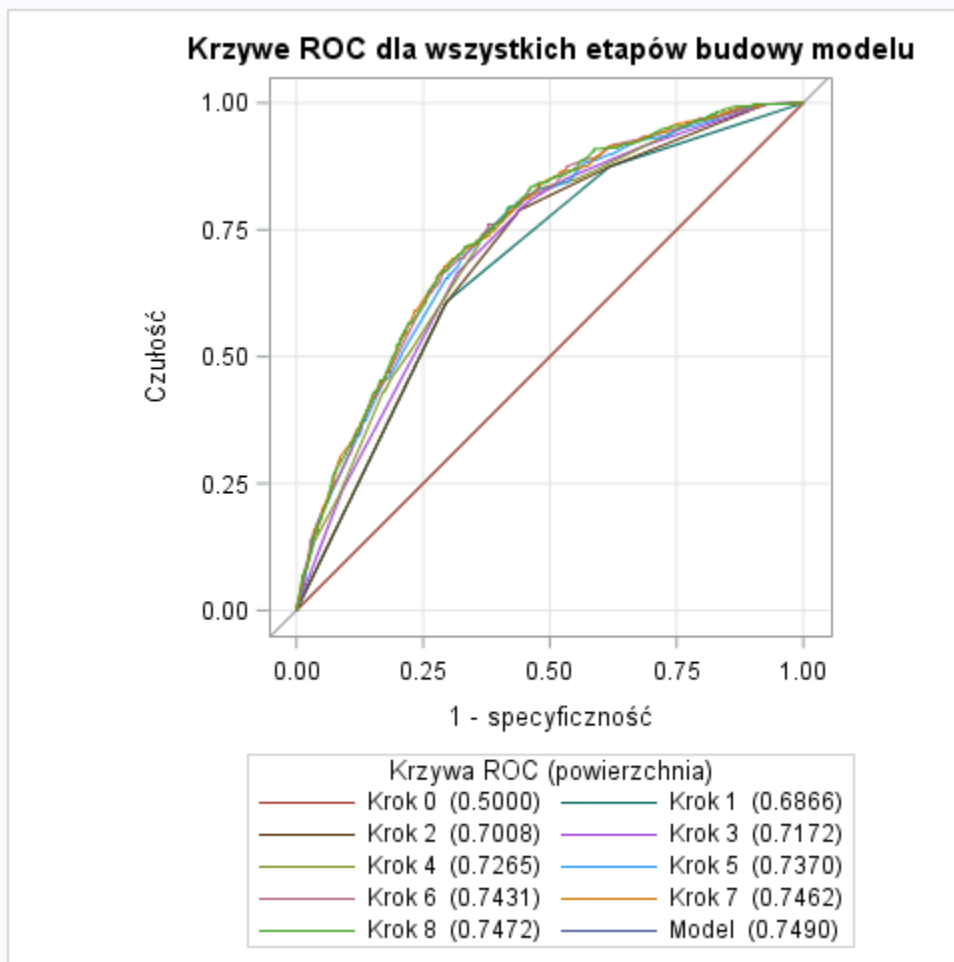
- **Czy osoby pochodzące z obszaru dawnej Republiki Federalnej Niemiec (RFN) rozwodzą się częściej, niż ludzie zamieszkujący dawną Niemiecką Republikę Demokratyczną (NRD)?**

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 osoby zadowolone rozwodzą się znacznie rzadziej. Przy poziomie istotności $\alpha = 0.05$ wartość ta waha się pomiędzy 27% a 107.1%.

10.5 Krzywa ROC

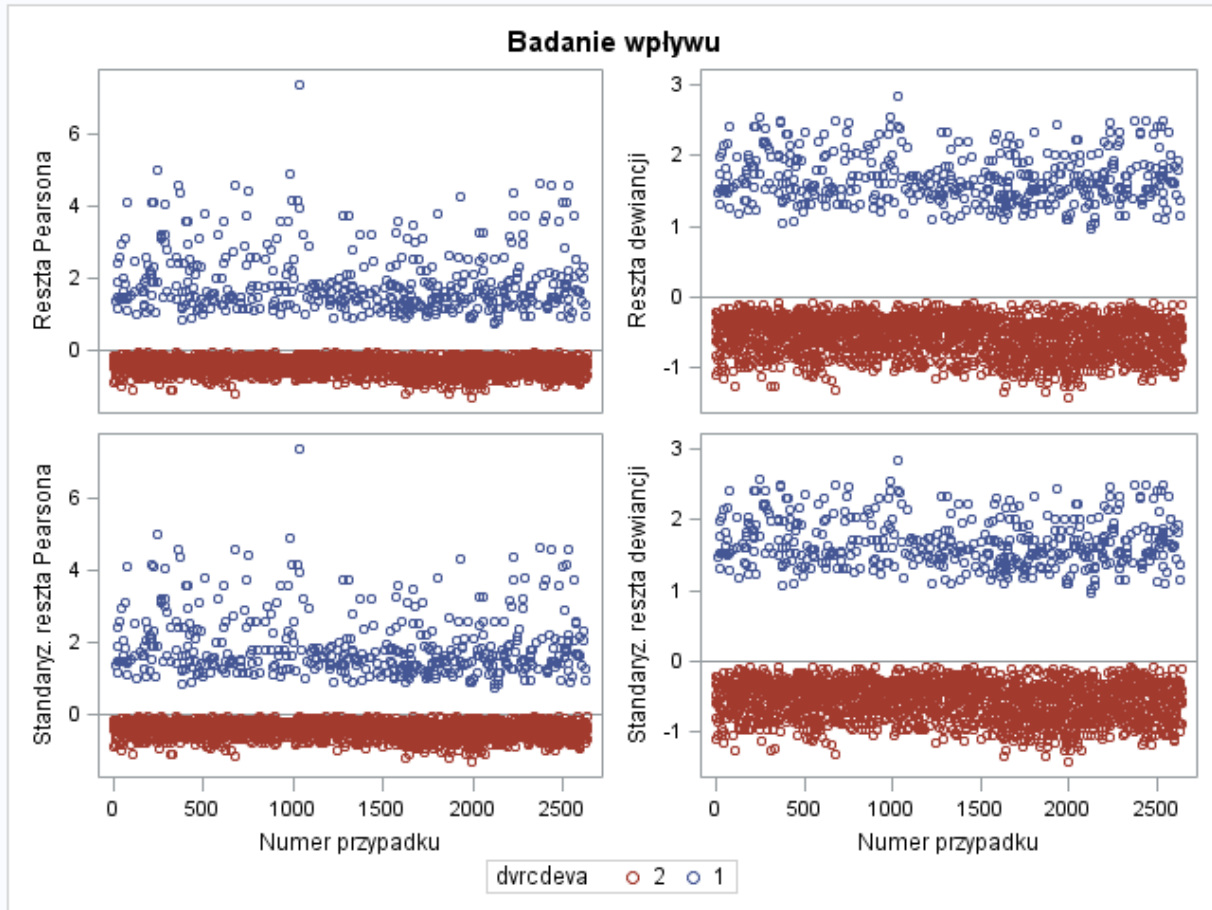


Źródło: wydruk z programu SAS

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 zbytnej 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.

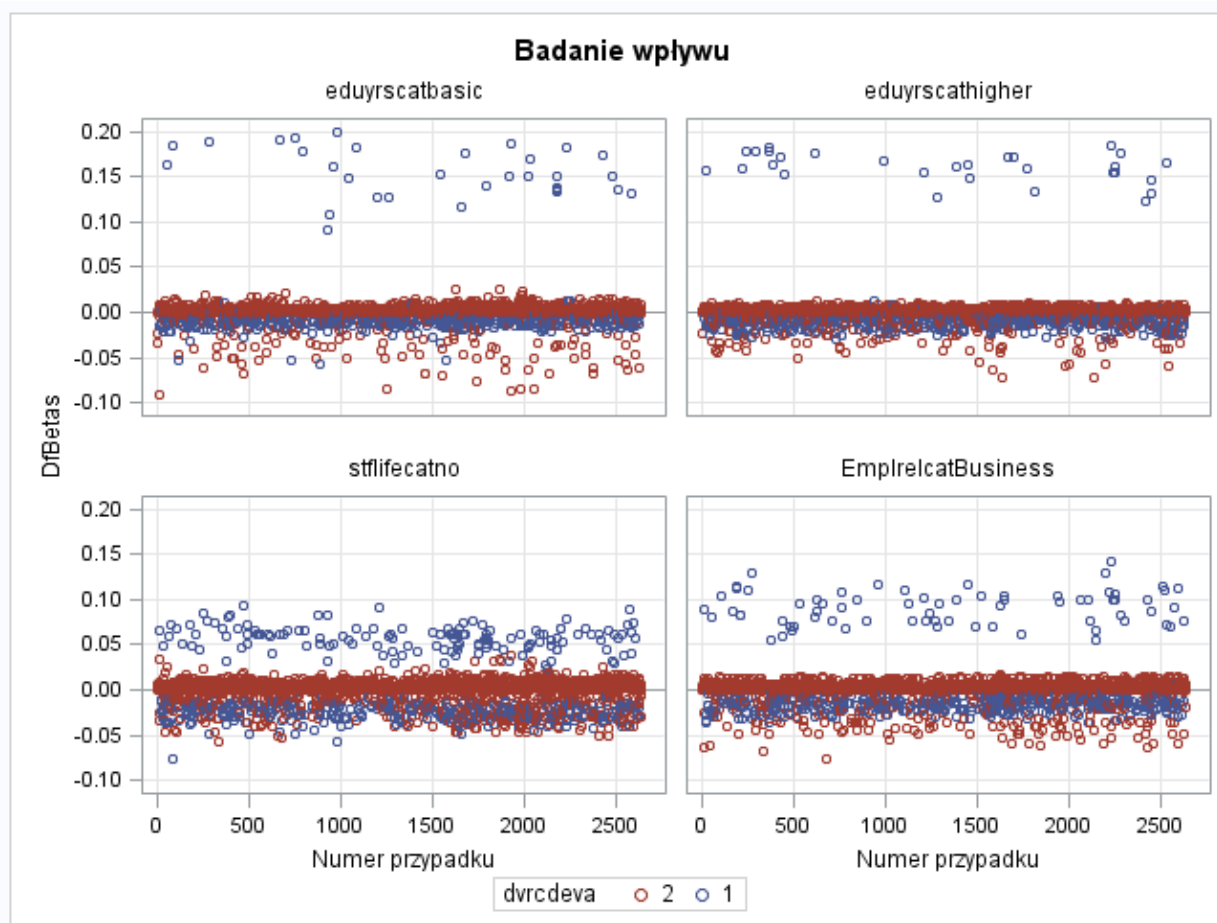


Źródło: wydruk z programu SAS

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 objaśniana 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, zadbaliliśmy 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 w 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, największy wpływ zaobserwowaliśmy dla zmiennej *imfree*, 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 first, then childbearing: Women's strategy in post-socialist Poland. *Population Studies*, 63 (3), 253–276
- [6] <http://kolegia.sgh.waw.pl/pl/KAE/struktura/ISiD/Documents/KSIAZKA%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/30721993-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_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](http://www.nationalcouncilonfamilyrelations.org/)
- [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] <https://link.springer.com/article/10.1186/s40711-015-0003-0> Qi Xu, Jianning Yu, Zeqi 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 demograficzne w Polsce w latach 90. w świetle koncepcji drugiego przejścia demograficznego*, Monografie 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/>

13. Kody SAS

Filtrowanie danych

```
LIBNAME TMP00001 "D:\Michał\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.lmptrad,
           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 lcwhct = 2 then 'no answer'
              else 'answer' end as lcwhctcat
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(%QSYFUNC(DATE(), NLDATE20.)) o godz. %TRIM(%SYFUNC(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.rlгатnd, T.chldhhe, T.hhmmbcat, T.eduyrscat, T.stflifecat,
T.Jbspvcат, T.Bennentcat, T.Emplrelcat, T.lcwhctcat, 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(%QSYFUNC(DATE(), NLDATE20.)) o godz. %TRIM(%SYFUNC(TIME(),
NLTIMAP20.))";

PROC LOGISTIC DATA=WORK.SORTTempTableSorted

PLOTS(ONLY)=ALL

;

CLASS rlгатnd (PARAM=REF) chldhhe (PARAM=REF) hhmmbcat
(PARAM=REF) eduyrscат (PARAM=REF) stflifecat (PARAM=REF) Jbspvcат
(PARAM=REF) Bennentcat (PARAM=REF) Emplrelcat (PARAM=REF) lcwhctcat
(PARAM=REF) Sclmeetcat (PARAM=REF)

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

MODEL dvrcdeva (Event = '1')=rlгатnd chldhhe hhmmbcat eduyrscат stflifecat Jbspvcат
Bennentcat Emplrelcat lcwhctcat 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;