

**BADANIE ZALEŻNOŚCI POMIĘDZY  
POCZUCIEM SZCZĘŚCIA A POSIADANIEM  
DZIECI MIESZKAŃCÓW BUŁGARII NA  
PODSTAWIE EUROPEAN SOCIAL SURVEY –  
RUNDA 9**

Alicja Kuś 82529

Joanna Olech 116711

Jan Puzio 81738

Justyna Zbiegień 107961

Mikołaj Szalek 72337

Tytus Panecki 77797

Zuzanna Przybylska 115979

## Spis treści

Wstęp.....	3
Przegląd zmiennych w literaturze .....	4
Wybrane zmienne.....	4
Wstępna analiza zbioru danych wybranych .....	6
Kodowanie i kategoryzacja zmiennych.....	11
Kodowanie zmiennej objaśnianej.....	11
Kodowanie zmiennych objaśniających .....	13
Analiza doboru zmiennych.....	38
Badanie korelacji .....	38
Badanie współliniowości - VIF .....	41
Współczynnik V-Cramera.....	42
Estymacja binarnego modelu regresji logistycznej .....	45
Etap I .....	47
Modele z interakcjami .....	50
Etap II .....	50
Model ze wszystkimi interakcjami.....	50
Model metodą selekcji wstecznej.....	50
Model metodą selekcji krokowej .....	51
Model zredukowany .....	52
Etap III.....	52
Wybór modelu.....	54
Podsumowanie.....	57
Bibliografia.....	58
Kod .....	58
Kodowanie zmiennych.....	68
Analiza korelacji.....	78

## Wstęp

Celem niniejszej pracy jest utworzenie i opisanie modelu tłumaczącego relacje pomiędzy odpowiedziami w ankiecie European Social Survey, a subiektywnym poziomem szczęścia w życiu. Ponadto wysunięto i badano hipotezę badawczą dotyczącą interakcji pomiędzy wiekiem ankietowanego i ilością dzieci, które posiada, a jego szczęściem. Niniejsza praca poszukuje odpowiedzi na pytania “Czy interakcja pomiędzy wiekiem a ilością posiadanych dzieci jest istotna statystycznie?” oraz “Czy powyżej pewnego wieku brak dzieci czyni ludzi bardziej nieszczęśliwymi?”

Tematyka szczęścia, a raczej subiektywnego odczucia szczęścia przez ankietowanych często pojawia się w artykułach dotyczących psychologii czy socjologii. Bardzo szybko okazało się, że czynniki takie jak PKB per capita nie są wystarczającym predyktorem określenia dobrostanu danej osoby (np. UN [1954]). W związku z tym rozpoczęły się poszukiwania indywidualnych zmiennych tłumaczących szczęście danej osoby. Zależności pomiędzy zadowoleniem z życia a zdrowiem, poziomem edukacji, bezrobociem czy zanieczyszczeniem pojawiały się w literaturze wielokrotnie. Artykuły takie jak Rehdanz i Maddison [2005], pokazały, że istotny wpływ na poziom szczęścia ma m. in. Klimat, a ludność w krajach, w których najzimniejszy miesiąc roku jest ciepły osiąga średnio wyższe poziomy szczęścia.

W związku z mnogością zjawisk, które wpływają na szczęście pośrednio i bezpośrednio ciężko jest zdefiniować model, który jednoznacznie opisuje wszystkie przyczyny szczęścia nie pomijając zmiennych współprzyczynowych. Niemniej jednak na przestrzeni wielu badań pewnego rodzaju zmienne powtarzają się i osiągają rezultaty istotne statystyczne w wielu badaniach. Stanowią one tak dużą część życia społecznego danej osoby, że nieuwzględnienie ich bez wątpienia spowodowałoby błąd pominiętych zmiennych. Należą do nich m. in. wiek, płeć, stan cywilny, subiektywny stan zdrowia, czy też pewnego rodzaju dochód na osobę. Niemniej jednak została wysunięta przez nas hipoteza dotycząca interakcji wieku i posiadania dzieci a subiektywnego stanu zdrowia. W artykułach naukowych kwestia ta nie została poruszona, a jednak istnieją uzasadnione przypuszczenia, że postrzeganie rodzicielstwa i jego wpływ na życie zmienia się wraz z wiekiem.

Dane pochodzą z badania *European Social Survey*. Do grupy badawczej wchodzi osoby w wieku od 15 lat, będącymi członkami gospodarstw domowych w każdym z 30 ankietowanych krajów. Niniejsze badanie dotyczy Bułgarii. Każda edycja ankiety zawiera pytania *core*, które powtarzają się ale także część rotacyjną, która zmienia się z badania na

badanie. Minimalna próba badawcza dla każdego kraju to 1,500 osób lub 800 jeżeli populacja kraju ma mniej niż 2 miliony osób. Obecna edycja to edycja 9, a moduły rotacyjne to “*Justice and Fairness in Europe*” oraz “*Timing of Life*” (który powtarza się częściowo z edycji 3 z 2006 roku).

## Przegląd zmiennych w literaturze

Dobór zmiennych został zainspirowany dwoma artykułami naukowymi. Pierwszy to “*Does Education Affect Happiness? Evidence for Spain*” Cunado, Gracia [2012]. Zmienne dobrane przez badaczy to wiek, płeć, subiektywny stan zdrowia, poziom dochodów, stan cywilny, główne zajęcie, posiadanie dzieci oraz oczywiście poziom edukacji. Najistotniejsze w tej analizie okazał się subiektywny stan zdrowia, dochód oraz wykształcenie. We wszystkich tych zmiennych można było zauważyć monotoniczne zależności występujące zgodnie z naszymi oczekiwaniami, tj. ludzie zdrowi okazali się szczęśliwsi od w złym stanie, wykształceni od mniej wykształconych. W danym badaniu bycie rodzicem pozytywnie wpływa na dobrostan ankietowanej osoby, niemniej jednak nie rozważono żadnych interakcji z tym związanych.

W kolejnym badaniu “*Happiness and Childbearing Across Europe*” Aassve, Goisis, Sironi [2012] temat wpływu rodzicielstwa został poruszony bardziej dogłębnie, niemniej jednak zmienne użyte są podobne. Badacze zdecydowali się na użycie wieku, ilości dzieci, statusie związku, podjęciu stałej pracy, dochodowi gospodarstwa domowego, ilości lat edukacji. Ponadto, użyto tam oddzielnych modeli dla każdej płci, co oznacza, że są to zmienne występujące w integracji ze wszystkimi innymi zmiennymi. Wyniki przez nich prezentowane są dość niejednoznaczne. Pomimo, że większa ilość dzieci jest skorelowana z wyższymi poziomami dobrostanu, to jednak w modelu, w którym użyto zmiennej binarnej “*przynajmniej jedno dziecko*” parametr okazał się negatywny.

## Wybrane zmienne

Zmienną modelowaną jest zmienna *happy* odpowiadająca na pytanie “*Taking all things together, how happy would you say you are?*”. Ankietowani po zadaniu pytania byli następnie proszeni o wyrażenie swojego szczęścia liczbą od 0 do 10, gdzie 10 stanowi wartość najwyższą. Zmienna ta następnie została uzależniona z od:

1. *inprdsc* - "How many people with whom you can discuss intimate and personal matters?" Z jak dużą ilością możesz rozmawiać o rzeczach osobistych i intymnych. Przyjmująca wartości od "0" dla braku takich osób do "6" dla 10 osób i więcej.
2. *mocntr* - "Was your mother born in [country]?" Czy twoja matka było urodzona w danym kraju? będącą zmienną binarną, "0" dla odpowiedzi negatywnej, "1" dla odpowiedzi pozytywnej.
3. *health* - "How is your health in general? Would you say it is ..." Jaki jest ogólnie twój stan zdrowia? dla której wartość "1" oznacza bardzo dobry subiektywny stan zdrowia, natomiast "7" oznacza negatywny.
4. *nbthcld* - "How many children have you ever given birth to/fathered?" Jak dużo dzieci masz? będącą zmienną ciągłą, naturalną.
5. *hmmmb* - "Including yourself, how many people - including children - live here regularly as members of this household?" Jak dużo osób żyje w regularnie w twoim gospodarstwie domowym? również będące zmienną ciągłą, naturalną.
6. *gndr* - zmienna pokazująca płeć danej osoby, "1" dla mężczyzny, "2" dla kobiety.
7. *agea* - wiek osoby ankietowanej.
8. *marsts* - Status cywilny. Zmienna kategoryczna nieuporządkowana, dla której różne poziomy oznaczają różne stany cywilne.
9. *estsz* - "Including yourself, about how many people are/were employed at the place where you usually work/worked..." Jak dużo osób pracuje w Twoim miejscu pracy? - Zmienna kategoryczna uporządkowana, dla której "1" oznacza miejsce pracy poniżej 10 osób, natomiast "5" oznacza 500 osób i więcej.
10. *jbspv* - "In your main job, do/did you have any responsibility for supervising the work of other employees?" - Czy w Twoim miejscu pracy, nadzorujesz prace innych ludzi? Zmienna binarna, dla której "1" oznacza, że osoba była na stanowisku kierowniczym, przeciwnie dla "2".
11. *hinctnta* - łączny dochód gospodarstwa domowego. Zmienna kategoryczna uporządkowana, w której kolejne wartości oznaczają kolejne decyle prawdopodobieństwa, uporządkowane rosnąco.

## Wstępna analiza zbioru danych wybranych

W poniższej części przeanalizowano braki w danych dla zmiennych wybranych w poprzedniej części. Wstępna analiza braków danych wykazała brak pustych obserwacji dla wszystkich 11 zmiennych (tabela poniżej).

**The MEANS Procedure**

Variable	Label	N Miss
happy	How happy are you	0
inprdsc	How many people with whom you can discuss intimate and personal matters	0
health	Subjective general health	0
moctr	Mother born in country	0
nbthld	Number of children ever given birth to/ fathered	0
hhmb	Number of people living regularly as member of household	0
gndr	Gender	0
agea	Age of respondent, calculated	0
marsts	Legal marital status	0
estsz	Establishment size	0
njbasp	Number of people responsible for in job	0
hinctnta	Household's total net income, all sources	0

*Tabela 1. Braki danych w zmiennych użytych do budowy modelu*

Należy jednak zwrócić uwagę na to w jaki sposób zmienne zostały skategoryzowane. Dla części zmiennych brak odpowiedzi, odmowa odpowiedzi czy odpowiedź „nie wiem” zostały zakodowane wartościami takimi jak na przykład: 77, 88, 99 czy 999. Tabela 2 przedstawia oryginalne skategoryzowanie wszystkich 11 zmiennych.

Zmienna	Oryginalne wartości i kategorie zmiennej																														
happy	<table> <tr> <th>Values</th><th>Categories</th></tr> <tr> <td>0</td><td>Extremely unhappy</td></tr> <tr> <td>1</td><td>1</td></tr> <tr> <td>2</td><td>2</td></tr> <tr> <td>3</td><td>3</td></tr> <tr> <td>4</td><td>4</td></tr> <tr> <td>5</td><td>5</td></tr> <tr> <td>6</td><td>6</td></tr> <tr> <td>7</td><td>7</td></tr> <tr> <td>8</td><td>8</td></tr> <tr> <td>9</td><td>9</td></tr> <tr> <td>10</td><td>Extremely happy</td></tr> <tr> <td>77</td><td>Refusal</td></tr> <tr> <td>88</td><td>Don't know</td></tr> <tr> <td>99</td><td>No answer</td></tr> </table>	Values	Categories	0	Extremely unhappy	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	Extremely happy	77	Refusal	88	Don't know	99	No answer
Values	Categories																														
0	Extremely unhappy																														
1	1																														
2	2																														
3	3																														
4	4																														
5	5																														
6	6																														
7	7																														
8	8																														
9	9																														
10	Extremely happy																														
77	Refusal																														
88	Don't know																														
99	No answer																														

inprdsc	<b>Values</b> <b>Categories</b> 0       None 1       1 2       2 3       3 4       4-6 5       7-9 6       10 or more 77      Refusal 88      Don't know 99      No answer
moctr	<b>Values</b> <b>Categories</b> 1       Yes 2       No 7       Refusal 8       Don't know 9       No answer
health	<b>Values</b> <b>Categories</b> 1       Very good 2       Good 3       Fair 4       Bad 5       Very bad 7       Refusal 8       Don't know 9       No answer
nbthld	<b>Values</b> <b>Categories</b> 66      Not applicable 77      Refusal 88      Don't know 99      No answer
hmmmb	<b>Values</b> <b>Categories</b> 77      Refusal 88      Don't know 99      No answer
gndr	<b>Values</b> <b>Categories</b> 1       Male 2       Female 9       No answer
agea	<b>Values</b> <b>Categories</b> 999     Not available

marsts	<b>Values</b> <b>Categories</b> 1   Legally married 2   In a legally registered civil union 3   Legally separated 4   Legally divorced/Civil union dissolved 5   Widowed/Civil partner died 6   None of these (NEVER married or in legally registered civil union) 66   Not applicable 77   Refusal 88   Don't know 99   No answer
estsz	<b>Values</b> <b>Categories</b> 1   Under 10 2   10 to 24 3   25 to 99 4   100 to 499 5   500 or more 6   Not applicable 7   Refusal 8   Don't know 9   No answer
njbospv	<b>Values</b> <b>Categories</b> 66666   Not applicable 77777   Refusal 88888   Don't know 99999   No answer
hinctnta	<b>Values</b> <b>Categories</b> 1   J - 1st decile 2   R - 2nd decile 3   C - 3rd decile 4   M - 4th decile 5   F - 5th decile 6   S - 6th decile 7   K - 7th decile 8   P - 8th decile 9   D - 9th decile 10   H - 10th decile 77   Refusal 88   Don't know 99   No answer

*Tabela 2. Oryginalna kategoryzacja zmiennych*

Po przeanalizowaniu oryginalnej kategoryzacji zmiennych postanowiono zakwalifikować wartości przedstawione w tabeli poniżej (tabela 3) jako braki danych.



Zmienna	Oryginalne wartości i kategorie zmiennej - braki	
happy	<b>Values</b>	<b>Categories</b>
	77	Refusal
	88	Don't know
	99	No answer
inprdsc	77	Refusal
	88	Don't know
	99	No answer
mocntr	7	Refusal
	8	Don't know
	9	No answer
health	7	Refusal
	8	Don't know
	9	No answer
nbthld	77	Refusal
	88	Don't know
hhmmb	77	Refusal
	88	Don't know
	99	No answer
gndr	9	No answer
agea	999	Not available
marsts	77	Refusal
	88	Don't know
estsz	6	Not applicable
	7	Refusal
	8	Don't know
	9	No answer
njbspv	77777	Refusal
	88888	Don't know
	99999	No answer

hinctnta	77 88 99	Refusal Don't know No answer
----------	----------------	------------------------------------

*Tabela 3. Klasyfikacja braków danych*

Następnie, ponownie przeprowadzono analizę braków danych. Wyniki zaprezentowano w tabeli nr 4. Zmienną z największym odsetkiem brakujących obserwacji była zmienna estsz (wielkość zakładu pracy) 29,94% z 2189 obserwacji. Odsetek braków danych dla zmiennych: happy, inprdsc, mocntr, agea, marsts, njbspv i hinctnta mieścił się w przedziale od 0,05% do 13,97%. Natomiast zmienne health, nbtheld, hhmmb oraz gndr posiadały 100% uzupełnionych obserwacji. Z uwagi na fakt, że odsetek brakujących obserwacji nie przekroczył progu 33% w żadnej ze zmiennych, nie podjęto działań usuwania zmiennych. Po usunięciu braków, w zbiorze danych pozostało 1269 obserwacji.

Zmienna	Liczba brakujących obserwacji	Liczba pełnych obserwacji	% braków danych	% uzupełnionych danych
happy	16	2182	0,73%	99,27%
inprdsc	204	1994	9,28%	90,72%
health	0	2198	0,00%	100,00%
mocntr	4	2194	0,18%	99,82%
nbtheld	0	2198	0,00%	100,00%
hhmmb	0	2198	0,00%	100,00%
gndr	0	2198	0,00%	100,00%
agea	1	2197	0,05%	99,95%

marsts	19	2179	0,86%	99,14%
estsz	658	1540	29,94%	70,06%
njbospv	41	2157	1,87%	98,13%
hinctnta	307	1891	13,97%	86,03%

*Tabela 4. Analiza braków danych*

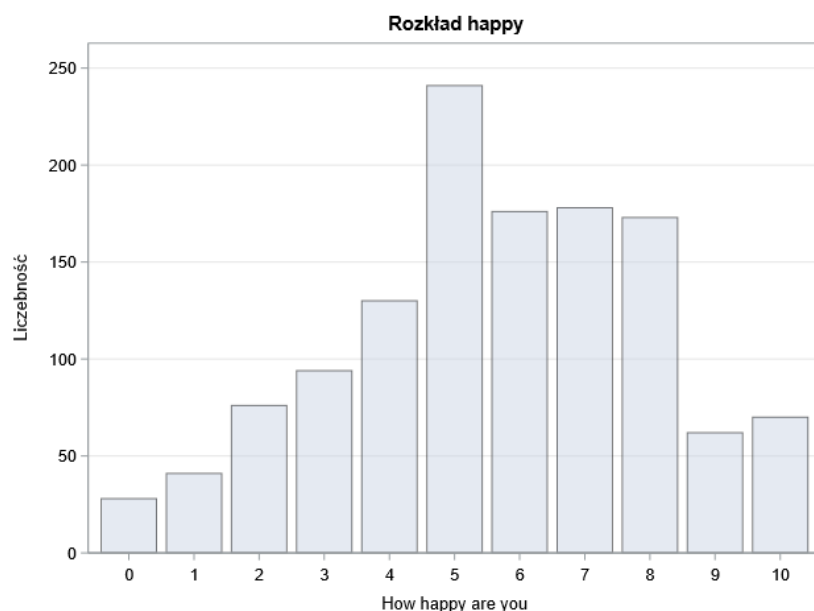
## Kodowanie i kategoryzacja zmiennych

### Kodowanie zmiennej objaśnianej

Zmienną celu jest kolumna *happy*. Jest to zmienna jakościowa, składająca się z 11 poziomów, gdzie 0 oznacza „Bardzo nieszczęśliwy”, a 10 „Bardzo szczęśliwy”. Stanowi ona odpowiedź na pytanie „Jak szczęśliwy jesteś?”. Po usunięciu obserwacji zawierających odmowę odpowiedzi na pytanie kodowanie zmiennej objaśnianej oraz jej rozkład prezentuje się następująco:

How happy are you				
happy	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
0	28	2.21	28	2.21
1	41	3.23	69	5.44
2	76	5.99	145	11.43
3	94	7.41	239	18.83
4	130	10.24	369	29.08
5	241	18.99	610	48.07
6	176	13.87	786	61.94
7	178	14.03	964	75.97
8	173	13.63	1137	89.60
9	62	4.89	1199	94.48
10	70	5.52	1269	100.00

*Tabela 5. Rozkład zmiennej happy – przed kategoryzacją*



Wykres 1. Rozkład zmiennej happy – przed kategoryzacją

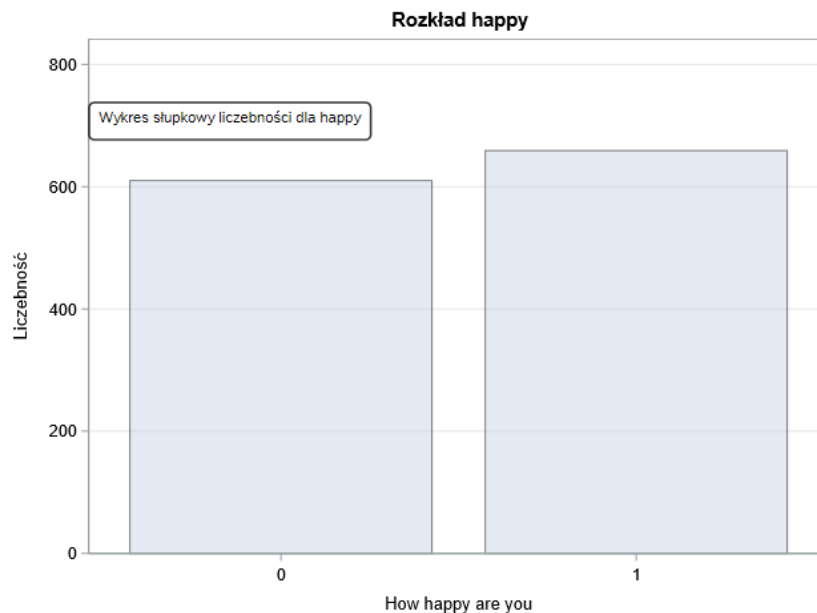
Widoczne jest, że dominującą kategorią jest kategoria „5”. Stanowi ona 19% wszystkich odpowiedzi respondentów. Skrajne wartości wybierane były najrzadziej. Ze względu na to, że celem analizy jest budowa modelu regresji logistycznej, wybraną zmienną celu należy poddać dychotomizacji. Polega to na przekodowaniu zmiennej do dwóch poziomów. Nowe kodowanie zmiennej zostało przedstawione poniżej.

- 0 – „Nieszczęśliwy” (kategorie 0, 1, 2, 3, 4, 5)
- 1 – „Szczęśliwy” (kategorie 6, 7, 8, 9, 10)

Po przekodowaniu rozkład zmiennej happy wygląda następująco:

How happy are you				
happy	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
0	610	48.07	610	48.07
1	659	51.93	1269	100.00

Tabela 6. Rozkład zmiennej happy – po kategoryzacji



Wykres 2. Rozkład zmiennej happy – po kategoryzacji

Liczebność odpowiedzi w obydwóch kategoriach jest do siebie zbliżona, 48,07% respondentów określiło się jako osoby nieszczęśliwe, natomiast 51,93% jako osoby szczęśliwe.

## Kodowanie zmiennych objaśniających

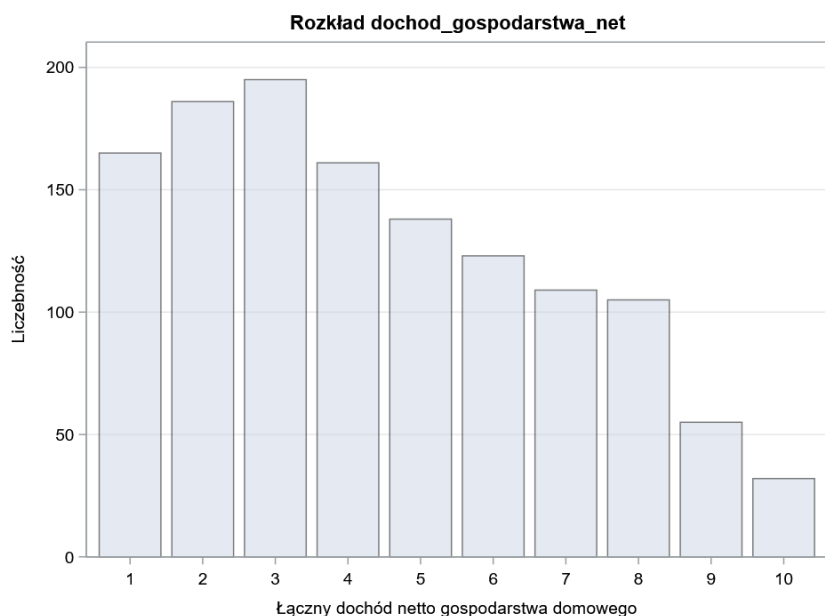
### 1. hinctnta - dochod\_gospodarstwa\_net

Zmienna hinctnta określa całkowity dochód netto gospodarstwa domowego ze wszystkich źródeł, przedstawiony w decylach. Zmienna obejmuje 10 kategorii, odpowiednio dla każdego kolejnego decyla, 1 – decyl pierwszy, 2- decyl drugi itd. Tabela oraz wykres przedstawione poniżej przedstawiają rozkład zmiennej przed kategoryzacją.

Liczebność  
Proc. wier.  
Proc. kol.

Tabela dochod_gospodarstwa_net od Szczęśliwy			
dochod_gospodarstwa_net(Łączny dochód netto gospodarstwa domowego)	Szczęśliwy(Jak bardzo jesteś szczęśliwy)		
	0	1	Suma
1	120 72.73 19.67	45 27.27 6.83	165
2	102 54.84 16.72	84 45.16 12.75	186
3	114 58.46 18.69	81 41.54 12.29	195
4	83 51.55 13.61	78 48.45 11.84	161
5	58 42.03 9.51	80 57.97 12.14	138
6	44 35.77 7.21	79 64.23 11.99	123
7	35 32.11 5.74	74 67.89 11.23	109
8	31 29.52 5.08	74 70.48 11.23	105
9	16 29.09 2.62	39 70.91 5.92	55
10	7 21.88 1.15	25 78.13 3.79	32
Suma	610	659	1269

Tabela 7. Tablica kontyngencji zmiennych *hinctnta* i *happy* – przed kategoryzacją



Wykres 3. Rozkład zmiennej *hinctnta* – przed kategoryzacją

Pierwsze 4 decyle stanowią największy udział wszystkich odpowiedzi, jest to aż 55%. Najbardziej liczną kategorią jest decyl 3 (ok.15,4%), natomiast najmniej liczną decyl 10, znajduje się tam jedynie ok. 2,5% obserwacji.

Zmienną *hinctnta* podzielono na następujące kategorie:

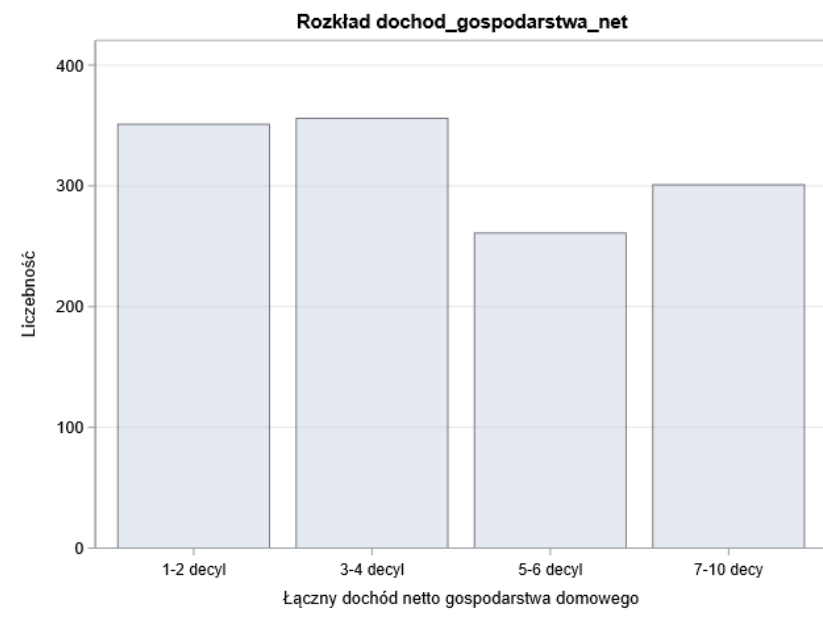
- 0 – 1-2 decyl (kategorie 1 i 2)
- 1 – 3-4 decyl (kategorie 3 i 4)

- 2 – 5-7 decyl (kategorie 5, 6 i 7)
- 3 – 8-10 decyl (kategorie 8, 9 ,10)

Rozkład zmiennej po kategoryzacji i przekodowaniu przedstawiony został na wykresie oraz w tabeli poniżej.

Frequency Row Percent Column Percent	Łączny dochód netto gospodarstwa domowego			
	Jak bardzo jesteś szczęśliwy			
	Łączny dochód netto gospodarstwa domowego	Jak bardzo jesteś szczęśliwy		
		0	1	Suma
	1-2 decyl	222 63.25 36.39	129 36.75 19.58	351
	3-4 decyl	197 55.34 32.30	159 44.66 24.13	356
	5-6 decyl	102 39.08 16.72	159 60.92 24.13	261
	7-10 decy	89 29.57 14.59	212 70.43 32.17	301
	Suma	610	659	1269

Tabela 8. Tablica kontyngencji zmiennych *hinctnta* i *happy* – po kategoryzacji

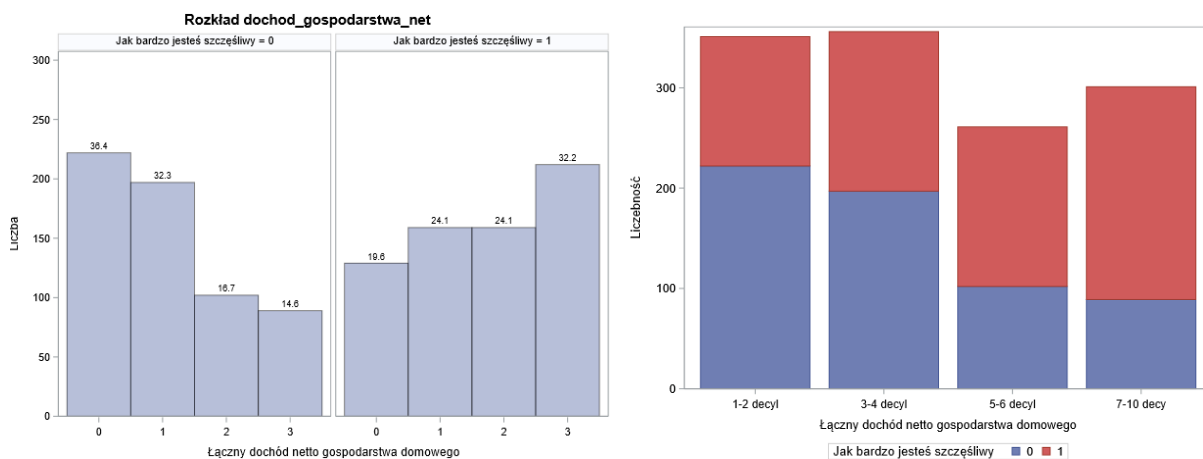


Wykres 4. Rozkład zmiennej *hinctnta* – po kategoryzacji

Po kategoryzacji zmiennej liczebności obserwacji w poszczególnych kategoriach są do siebie zbliżone. Najmniej liczna jest obecnie kategoria „5-6 decyl”, stanowi ona ok. 20,5% wszystkich obserwacji.

Dodatkowo na wykresach poniżej przedstawiono wykres rozkładu zmiennej *hinctnta* ze względu na zmienną objaśnianą *happy*. Można zauważyć, że najmniej szczęśliwi respondenci

to osoby posiadające dochód gospodarstwa domowego znajdującego się w kategoriach „1 – 2 decyl” oraz „3 – 4 decyl”. Natomiast 32,2% osób szczęśliwych posiada dochód netto w najwyższej kategorii - „7 – 10 decyl”. W tej kategorii znacznie przeważa liczba osób szczęśliwych w porównaniu do kategorii pozostałych.



Wykres 5. Rozkład zmiennej hincntcta względem zmiennej happy (udział 0 i 1)

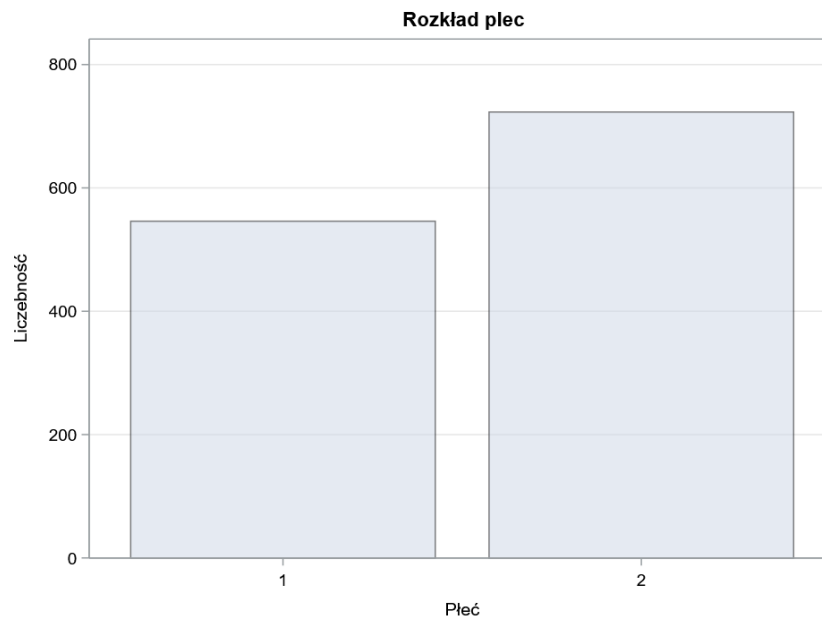
## 2. gndr – płeć

Zmienna gndr opisuje płeć respondenta, wartość 1 oznacza mężczyznę, natomiast wartość 2 kobietę. Jest to zmienna nominalna. Analizując poniższą tabelę oraz wykres można zauważyć, że kobiety stanowią ponad połowę respondentów w badanej próbie, jest to ok. 57%. Dodatkowo można zauważyć, że rozkład pomiędzy osobami szczęśliwymi oraz nieszczęśliwymi jest do siebie bardzo zbliżony dla obydwóch płci (48,5%/51,5% dla mężczyzn oraz 47,7%/52,3% dla kobiet).

Liczebność Proc. wier. Proc. kol.	Tabela plec od Szczęśliwy		
	Szczęśliwy(Jak bardzo jesteś szczęśliwy)		
	0	1	Suma
plec(Płeć)			
1	265 48.53 43.44	281 51.47 42.64	546
2	345 47.72 56.56	378 52.28 57.36	723
Suma	610	659	1269

Tabela 9. Tablica kontyngencji zmiennych gndr i happy



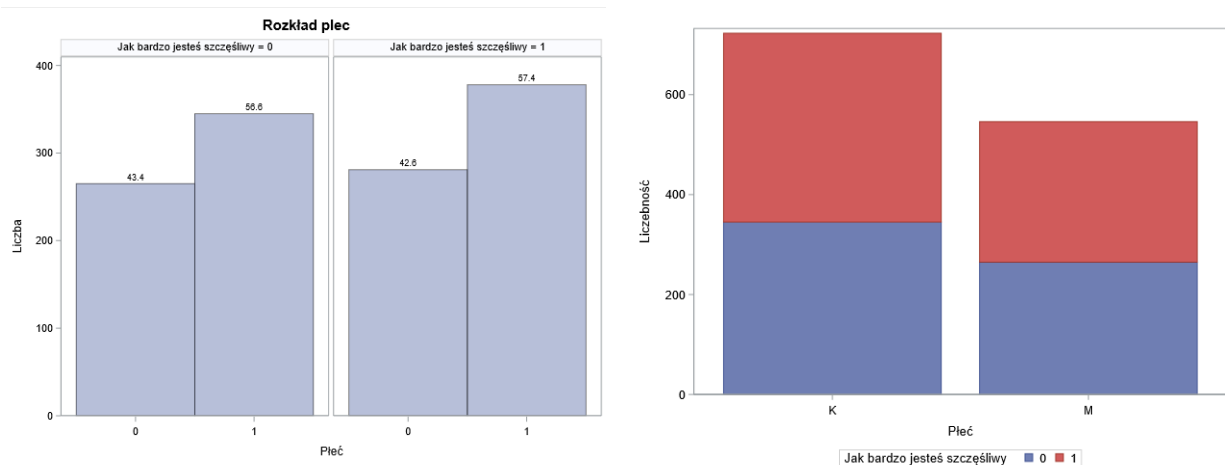


Wykres 6. Rozkład zmiennej *gndr*

Zmienną *gndr* zakodowano w następujący sposób:

- 0 – 1 (kategoria mężczyzn)
- 1 – 2 (kategoria kobiet)

Wykresy przedstawione poniżej ukazują rozkład zmiennej *gndr* ze względu na zmienną objaśnianą, czyli bycie szczęśliwym oraz udział osób szczęśliwych/nieszczęśliwych w poszczególnych kategoriach. Widoczne jest, że zarówno w przypadku osób szczęśliwych jak i nieszczęśliwych przeważa liczba kobiet, wynika to jednak ze specyfikacji próby i faktu, że dominują w niej kobiety.



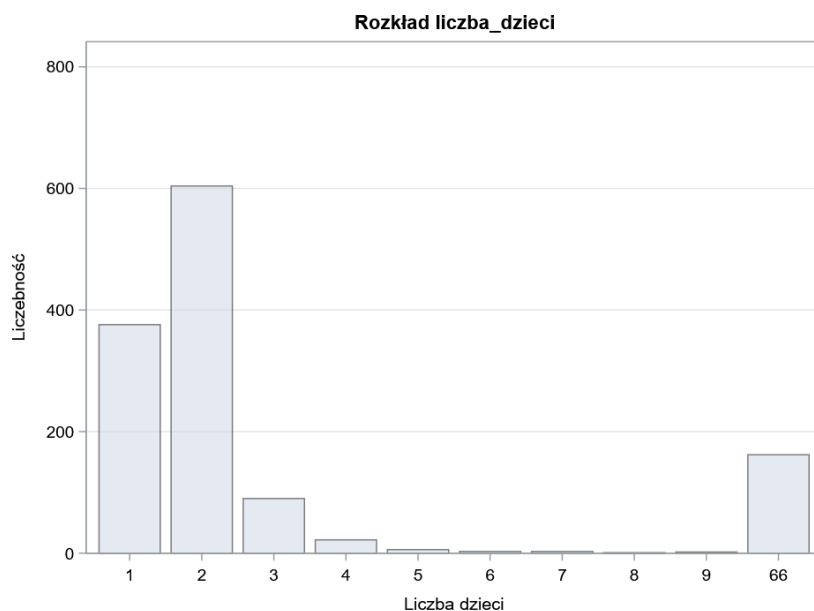
Wykres 7. Rozkład zmiennej *gndr* względem zmiennej *happy* (udział 0 i 1)

### 3. nbthclد – liczba\_dzieci

Zmienna nbthclد opisuje liczbę dzieci posiadaną przez respondentów. W przypadku tej zmiennej, odpowiedź „66 – nie dotyczy” potraktowano jako odpowiedź mówiącą o nieposiadaniu żadnego dziecka. Maksymalna liczba dzieci, którą posiada respondent to 9, w całej próbie znajdują się jedynie dwie takie osoby. Poniższy wykres oraz tabela w widoczny sposób pokazują, że najwięcej osób posiada 2 dzieci, jest to aż 47,6%. Mniej liczną grupą są osoby posiadające 1 dziecko, a następnie nieposiadające żadnego dziecka.

Liczebność Proc. wier. Proc. kol.	Tabela liczba_dzieci od Szczęśliwy		
	liczba_dzieci(Liczba dzieci)	Szczęśliwy(Jak bardzo jesteś szczęśliwy)	
		0	1
			Suma
	1	168 44.68 27.54	208 55.32 31.56
	2	296 49.01 48.52	308 50.99 46.74
	3	52 57.78 8.52	38 42.22 5.77
	4	13 59.09 2.13	9 40.91 1.37
	5	5 83.33 0.82	1 16.67 0.15
	6	1 33.33 0.16	2 66.67 0.30
	7	2 66.67 0.33	1 33.33 0.15
	8	1 100.00 0.16	0 0.00 0.00
	9	1 50.00 0.16	1 50.00 0.15
	66	71 43.83 11.64	91 56.17 13.81
	Suma	610	659
			1269

Tabela 10. Tablica kontyngencji zmiennych nbthclد i happy – przed kategoryzacją



Wykres 8. Rozkład zmiennej nbthclد – przed kategoryzacją

Zdecydowano się na następujące przekodowanie zmiennej nbthclد ze względu na małe liczebności w kategoriach powyżej kategorii „2”.

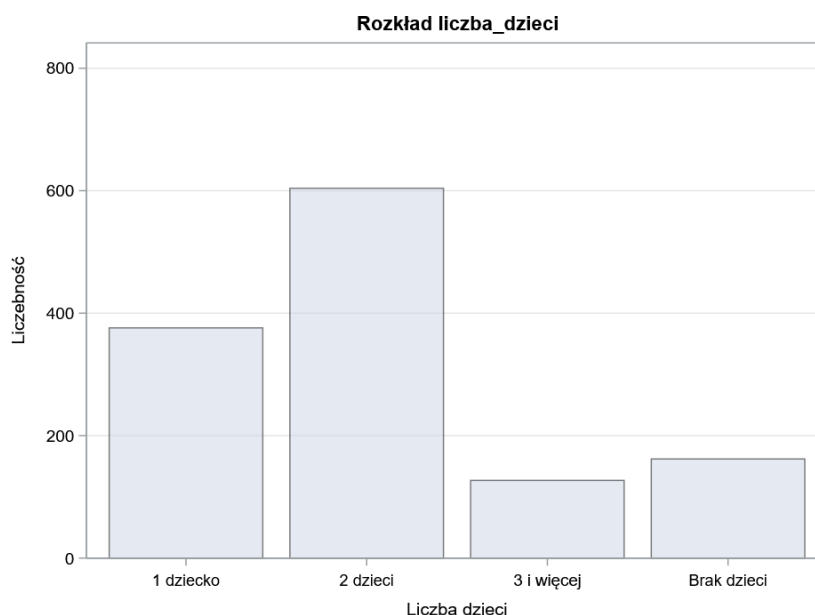
- 0 – Brak dzieci (kategoria 66)
- 1 – Jedno dziecko (kategoria 1)
- 2 – Dwoje dzieci (kategoria 2)
- 3 – 3 i więcej dzieci (kategoria 3, 4, 5, 6, 7, 8, 9)

Nowy rozkład zmiennej po skategoryzowaniu został przedstawiony poniżej w tabeli oraz na wykresie.

Liczebność  
Proc. wier.  
Proc. kol.

Tabela liczba_dzieci od Szczęśliwy			
liczba_dzieci(Liczba dzieci)	Szczęśliwy(Jak bardzo jesteś szczęśliwy)		
	0	1	Suma
1 dziecko	168	208	376
	44.68	55.32	
	27.54	31.56	
2 dzieci	296	308	604
	49.01	50.99	
	48.52	46.74	
3 i więcej	75	52	127
	59.06	40.94	
	12.30	7.89	
Brak dzieci	71	91	162
	43.83	56.17	
	11.64	13.81	
Suma	610	659	1269

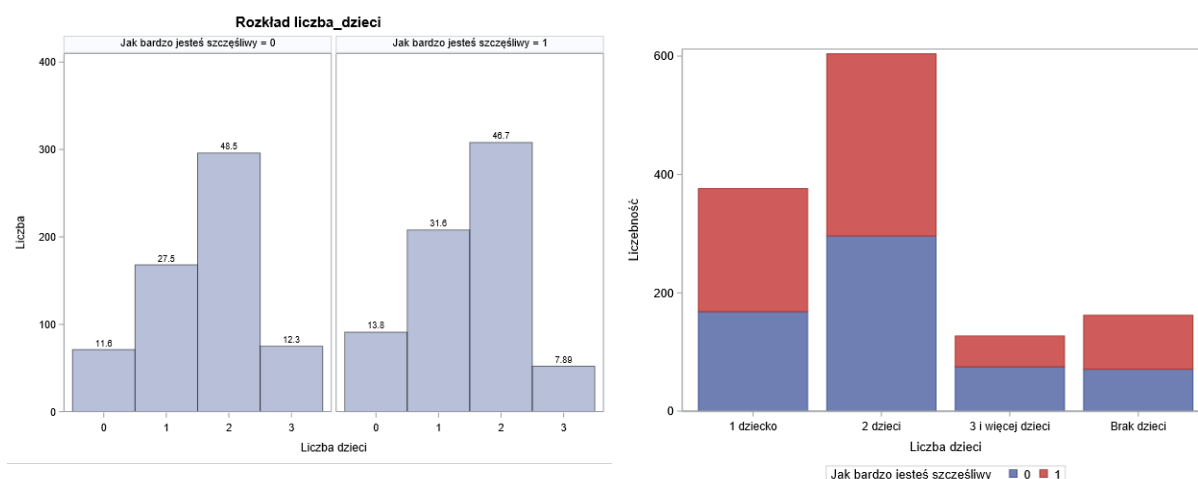
Tabela 11. Tablica kontyngencji zmiennych nbthclد i happy – po kategoryzacji



Wykres 9. Rozkład zmiennej nbthcld – po kategoryzacji

Po kategoryzacji nadal najbardziej liczną kategorią jest ta dotycząca posiadania 2 dzieci, a najmniej liczną kategoria obejmująca posiadanie 3 i więcej dzieci.

Poniższe wykresy przedstawiają udział poszczególnych kategorii w podziale ze względu na zmienną zależną *happy*. Można zauważyć, że osoby nieposiadające dzieci lub posiadające jedynie jedno dziecko są szczęśliwsze od tych, które posiadają 2 i więcej dzieci. W kategorii „3 i więcej dzieci” największa część respondentów jest nieszczęśliwa. Najbardziej zrównoważona liczebność odpowiedzi pod względem bycia szczęśliwym występuje w kategorii „2 dzieci”.



Wykres 10. Rozkład zmiennej nbthcld względem zmiennej happy (udział 0 i 1)

#### 4. njbspv – liczba\_podwładnych

Zmienna *njbspv* informuje o liczbie osób, za które respondent jest odpowiedzialny w pracy. Jest to zmienna ilościowa. Jej rozkład został przedstawiony w tabeli poniżej. Zdecydowano się na potraktowanie odpowiedzi „66666 – Nie dotyczy” jako informację o nieposiadaniu żadnych podwładnych w pracy, czyli nieponoszeniu odpowiedzialności za nikogo. Tabela ukazuje, że zdecydowana większość osób należy do tej właśnie grupy, jest to aż 87% badanej próby.

Jakiego typu stanowisko respondent obejmuje w pracy zwykłe/kierownicze	
liczba_podwładnych	Liczebność
1	15
2	19
3	10
4	5
5	15
6	4
7	7
8	4
9	2
10	17
11	1
12	4
14	1
15	10
16	1
19	1
20	10
22	2
24	1
25	4
26	1
27	1
29	1
30	6
35	1
39	1
40	1
42	1
43	1
50	5
70	1
86	1
66666	1115

Tabela 12. Rozkład zmiennej *njbspv* – przed kategoryzacją

Ze względu na duże rozbieżności w ilości podwładnych zdecydowano się na skategoryzowanie tej zmiennej na dwie kategorie oraz zrobienie z niej zmiennej jakościowej.

- 0 – Stanowisko zwykłe (kategoria 66666)
- 1 – Stanowisko kierownicze (wszystkie pozostałe kategorie)

Rozkład zmiennej po tej zmianie został przedstawiony poniżej.

Liczebność Proc. wier. Proc. kol.	Tabela liczba_podwładnych od Szczęśliwy		
	Szczęśliwy(Jak bardzo jesteś szczęśliwy)		
liczba_podwładnych(Jakiego typu stanowisko respondent obejmuje w pracy zwykłe/kierownicze)	0	1	Suma
Stanowisko kierow	52 33.77 8.52	102 66.23 15.48	154
Zwykłe stanowisko	558 50.04 91.48	557 49.96 84.52	1115
Suma	610	659	1269

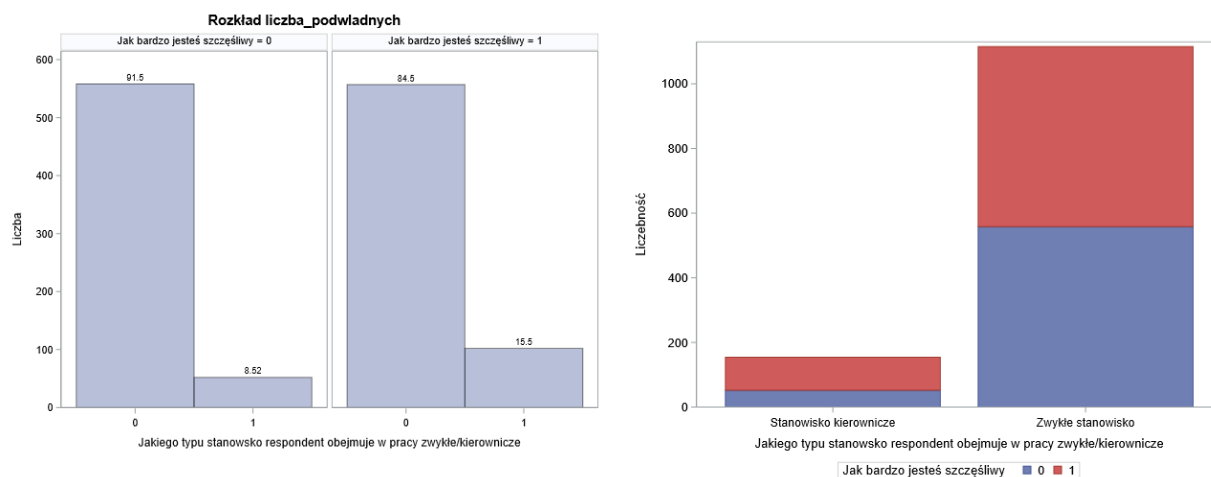
Tabela 13. Tablica kontyngencji zmiennych njbspv i happy – po kategoryzacji



Wykres 11. Rozkład zmiennej njbspv – po kategoryzacji

Można zaobserwować, że w obrębie tych dwóch kategorii, to w kategorii mówiącej obejmowaniu stanowiska kierowniczego jest większy udział osób szczęśliwych (66%).

Dodatkowo przygotowano wykres rozkładu zmiennej względem poziomu szczęścia. Tutaj również wyraźnie widoczne jest, że osoby na stanowiskach kierowniczych są szczęśliwsze od osób na zwykłych stanowiskach pracy. Natomiast w przypadku obejmowania zwykłego stanowiska pracy rozkład zmiennej *happy* jest bardzo wyrównany.



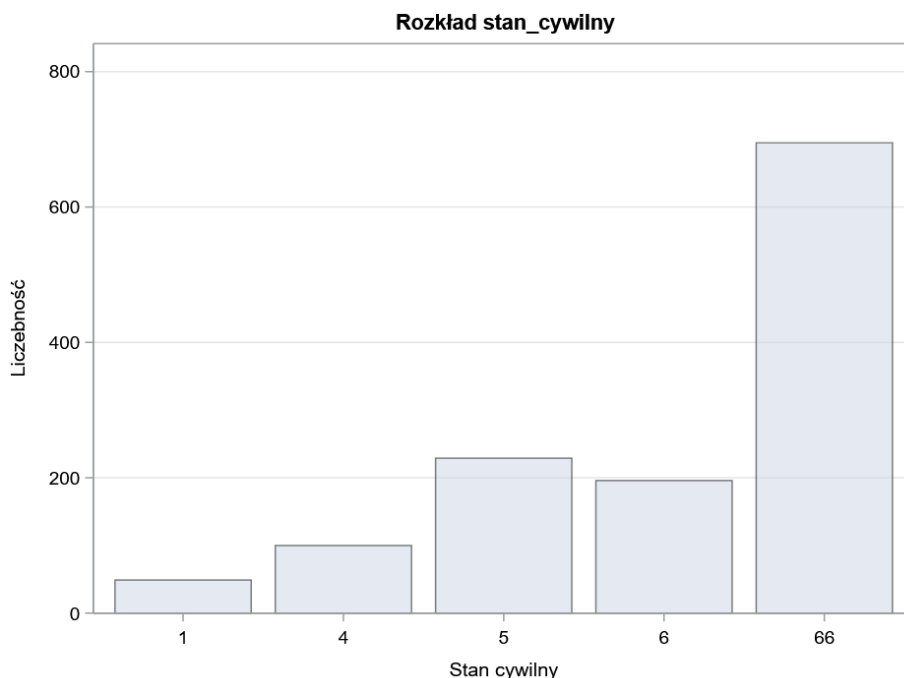
Wykres 12. Rozkład zmiennej njbspv względem zmiennej happy (udział 0 i 1)

## 5. marsts – stan\_cywilny

Zmienna *marsts* opsiuje stan cywilny respondenta. Jej rozkład został przedstawiony w tabeli poniżej. Zdecydowano się na potraktowanie odpowiedzi „66 – Nie dotyczy” jako nie bycie w żadnym związku lub bycie w związku nieformalnym. Tabela poniżej pokazuje, że najwięcej osób, bo aż 54% znajduje się w kategorii 66 - potraktowanych przez nas jako osoby niebędące w związku lub będące w związku nieformalnym. Pozostałe kategorie nie różnią się już od siebie liczebnością w tak znaczącym stopniu.

Liczebność Proc. wier. Proc. kol.	Tabela stan_cywilny od Szczęśliwy		
	Szczęśliwy(Jak bardzo jesteś szczęśliwy)		
stan_cywilny(Stan cywilny)	0	1	Suma
1	32 65.31 5.25	17 34.69 2.58	49
4	54 54.00 8.85	46 46.00 6.98	100
5	154 67.25 25.25	75 32.75 11.38	229
6	84 42.86 13.77	112 57.14 17.00	196
66	286 41.15 48.89	409 58.85 62.06	695
Suma	610	659	1269

Tabela 14. Tablica kontyngencji zmiennych *marsts* i *happy* – przed kategoryzacją



Wykres 13. Rozkład zmiennej *marsts* – przed kategoryzacją

Nowo powstałe kategorie prezentują się w następujący sposób:

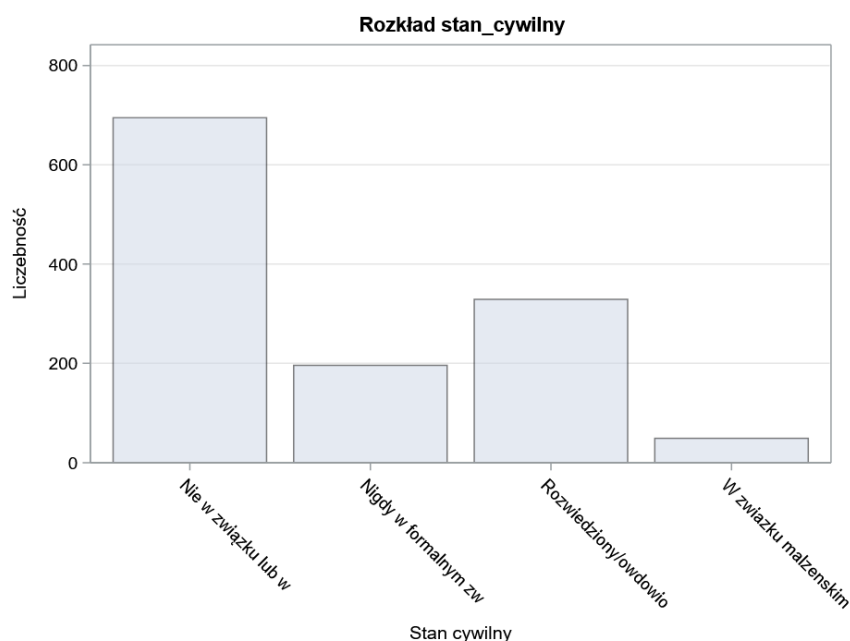
- 0 – „W związku małżeńskim” (kategoria 1)
- 1 – „Rozwiedziony/owdowiały” (kategoria 4,5)
- 2 – „Nigdy w formalnym związku” (kategoria 6)
- 3 – “Nie w związku lub w związku nieformalnym” (kategoria 66)

Po tej niewielkiej zmianie rozkład zmiennej prezentuje się następująco:

Liczebność Proc. wier. Proc. kol.	Tabela stan_cywilny od Szczęśliwy			
	stan_cywilny(Stan cywilny)	Szczęśliwy(Jak bardzo jesteś szczęśliwy)		
		0	1	Suma
	Nie w związku lub w	286	409	695
		41.15	58.85	
		46.89	62.06	
	Nigdy w formalnym zw	84	112	196
		42.86	57.14	
		13.77	17.00	
	Rozwiedziony/owdowio	208	121	329
		63.22	36.78	
		34.10	18.36	
	W związku małżeńskim	32	17	49
		65.31	34.69	
		5.25	2.58	
	Suma	610	659	1269

Tabela 15. Tablica kontyngencji zmiennych *marsts* i *happy* – po kategoryzacji

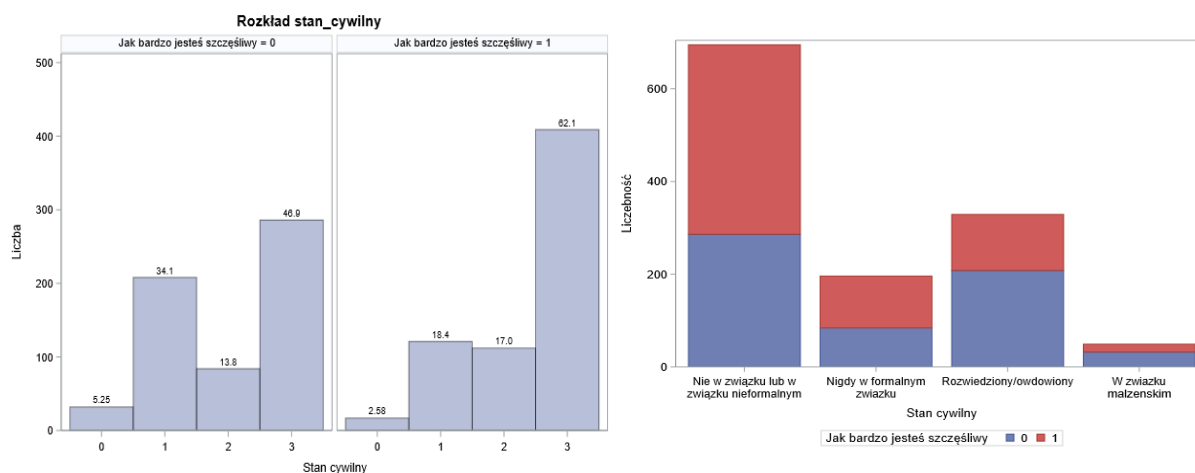




Wykres 14. Rozkład zmiennej *marsts* – po kategoryzacji

Najmniejszą pod względem liczebności jest kategoria, mówiąca o tym, że respondent jest obecnie w związku małżeńskim, a następnie ta mówiąca o tym, że nigdy nie był on w związku formalnym.

Poniższe wykresy pokazują, że w grupie osób szczęśliwych najwięcej jest osób niebędących w związkach lub będących w związkach nieformalnych. Ta grupa osób stanowi również spory odsetek wśród osób nieszczęśliwych, tutaj na drugim miejscu jako osoby nieszczęśliwe są osoby owdowiałe lub rozwiedzione. Również wśród osób będących w związku małżeńskim znaczna większość jest nieszczęśliwa.



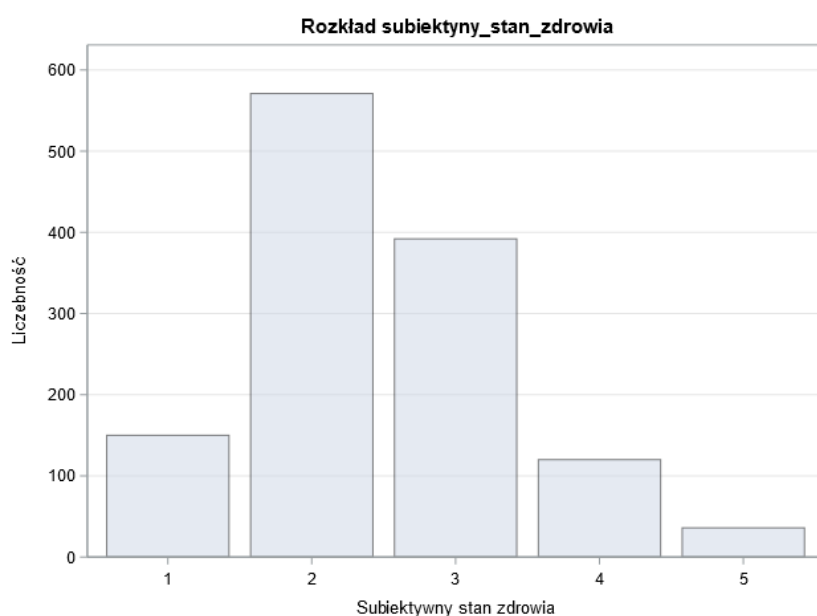
Wykres 15. Rozkład zmiennej *marsts* względem zmiennej *happy* (udział 0 i 1)

## 5. health – subiektywny\_stan\_zdrowia

Zmienna health odnosi się do subiektywnej oceny stanu zdrowia przez respondenta. Przyjmuje ona pięć wartości, od 1 – „Bardzo dobre” do 5- „Bardzo złe”. Im wyższa wartość zmiennej tym respondent określa swój stan zdrowia gorzej. Rozkład zmiennej przed kodowaniem i kategoryzacją prezentuje się następująco:

Liczebność Proc. wier. Proc. kol.	Tabela subiektywny_stan_zdrowia od Szczęśliwy		
	Szczęśliwy(Jak bardzo jesteś szczęśliwy)		
subiektywny_stan_zdrowia(Subiektywny stan zdrowia)	0	1	Suma
1	35 23.33 5.74	115 76.67 17.45	150
2	223 39.05 36.56	348 60.95 52.81	571
3	234 59.69 38.36	158 40.31 23.98	392
4	84 70.00 13.77	36 30.00 5.46	120
5	34 94.44 5.57	2 5.56 0.30	36
Suma	610	659	1269

Tabela 16. Tablica kontyngencji zmiennych health i happy – przed kategoryzacją



Wykres 16. Rozkład zmiennej health – przed kategoryzacją

Około 45% respondentów określa swój stan zdrowia jako „Dobry”, a kolejne 31% uznaje, że ich stan zdrowia jest „Średni”. Jedynie 36 osób z 1269 respondentów określiło swój stan zdrowia jako „Bardzo zły”, zdecydowana większość z nich uznaje się za osoby nieszczęśliwe.

Podjęto decyzję o następującym kodowaniu i kategoryzacji zmiennej *health*:

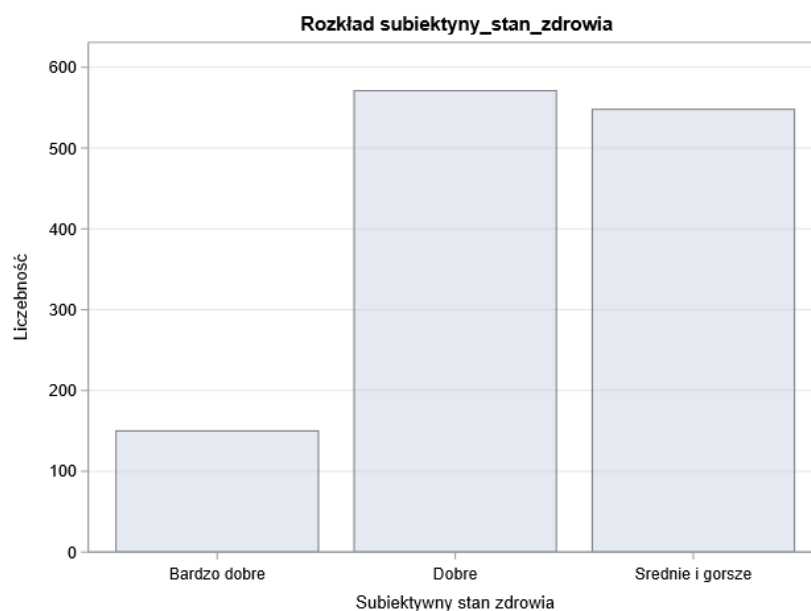
- 0 – „Bardzo dobre” (kategoria 1)
- 1 – „Dobre” (kategoria 2)

- 2 – „Średnie i gorsze” (kategoria 3, 4, 5)

Rozkład zmiennej po kategoryzacji został przedstawiony w tabeli oraz na wykresie poniżej.

Frequency Row Percent Column Percent	Subiektywny stan zdrowia			
	Jak bardzo jesteś szczęśliwy			
	Subiektywny stan zdrowia	Jak bardzo jesteś szczęśliwy		
		0	1	Suma
	Bardzo dobre	35 23.33 5.74	115 76.67 17.45	150
	Dobre	223 39.05 36.56	348 60.95 52.81	571
	Srednie i gorsze	352 64.23 57.70	196 35.77 29.74	548
	Suma	610	659	1269

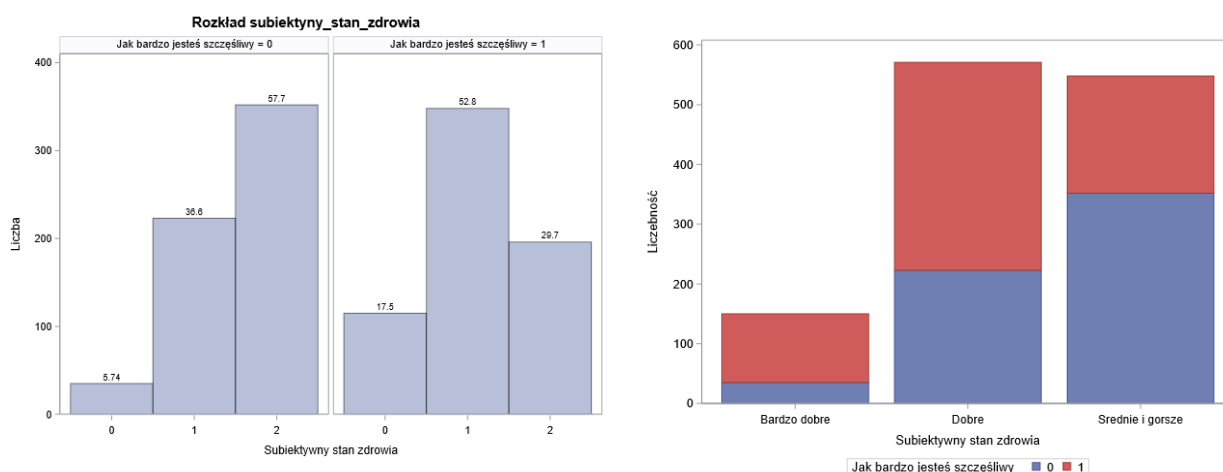
Tabela 17. Tablica kontyngencji zmiennych health i happy – po kategoryzacji



Wykres 17. Rozkład zmiennej health – po kategoryzacji

Liczebność kategorii „Dobre” oraz „Średnie i gorsze” są do siebie bardzo zbliżone. Najmniej liczną kategorią jest kategoria „Bardzo dobre” (ok. 12%).

Poniżej przygotowano wykres przedstawiający rozkład liczebności zmiennej health w zależności od zmiennej zależnej happy.



Wykres 18. Rozkład zmiennej health względem zmiennej happy (udział 0 i 1)

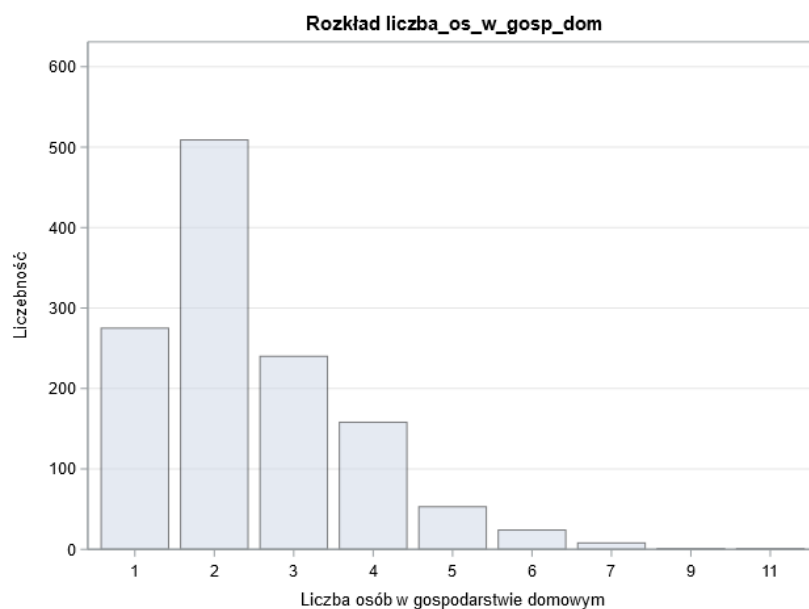
Można zauważyć, że rozkłady zmiennej health są bardzo zróżnicowane w zależności od zmiennej happy. Widoczne jest, że im lepszy subiektywny poziom zdrowia (kategoria 0 i 1) tym więcej osób szczęśliwych. W grupie osób szczęśliwych 70% osób uważa, że ich stan zdrowia jest dobry bądź bardzo dobry. Dodatkowo można wyciągnąć wniosek, że im lepsza ocena zdrowia przez respondenta tym większe prawdopodobieństwo, że jest on szczęśliwy.

## 6. hhmmb – liczba\_os\_w\_gosp\_dom

Kolejną zmienną jest zmienna hhmmb, która informuje o liczbie osób w gospodarstwie domowym. Jest to zmienna ilościowa. Jej rozkład jest zróżnicowany, maksymalna liczba osób jaka została podana przez jednego respondenta to 11. Liczebności grup maleją wraz ze wzrostem liczby osób w gospodarstwie domowym. Najbardziej liczną grupą jest odpowiedź, że gospodarstwo domowe składa się z dwóch osób. Rozkład zmiennej został przedstawiony poniżej.

Liczebność Proc. wier. Proc. kol.	Tabela liczba_os_w_gosp_dom od Szczęśliwy		
	liczba_os_w_gosp_dom(Liczba osób w gospodarstwie domowym)	Szczęśliwy(Jak bardzo jesteś szczęśliwy)	
		0	1
	1	172 62.55 28.20	103 37.45 15.63
	2	249 48.92 40.82	260 51.08 39.45
	3	92 38.33 15.08	148 61.67 22.46
	4	55 34.81 9.02	103 65.19 15.63
	5	27 50.94 4.43	26 49.08 3.95
	6	10 41.67 1.64	14 58.33 2.12
	7	4 50.00 0.66	4 50.00 0.61
	9	1 100.00 0.16	0 0.00 0.00
	11	0 0.00 0.00	1 100.00 0.15
	Suma	610	659
			1269

Tabela 18. Tablica kontyngencji zmiennych hhmmb i happy – przed kategoryzacją



Wykres 19. Rozkład zmiennej hhmmb – przed kategoryzacją

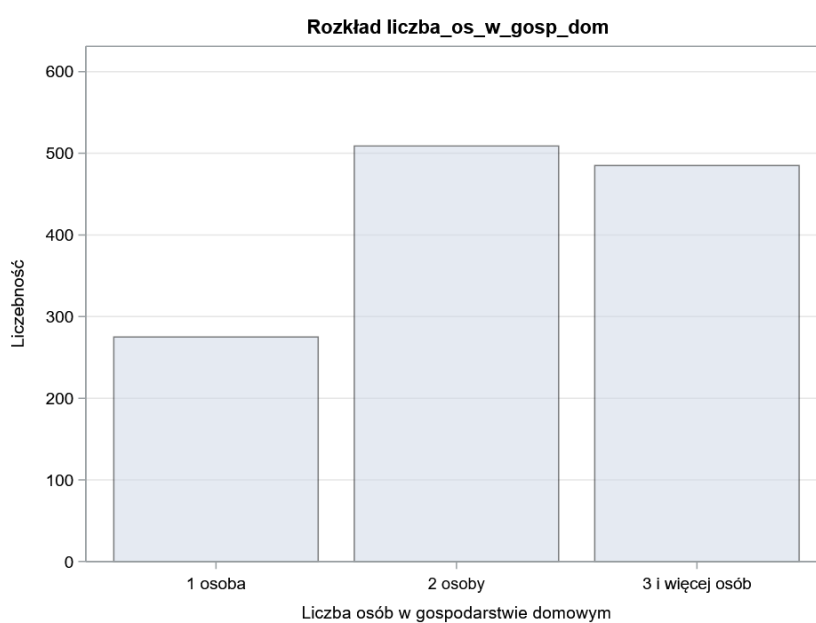
Zdecydowana większość wartości skupia się w czterech pierwszych kategoriach, dlatego podjęto decyzję o kategoryzacji zmiennej na 3 kategorie i zrobienie z niej zmiennej jakościowej.

- 0 – 1 osobowe gospodarstwo domowe (kategoria 1)
- 1 – 2 osobowe gospodarstwo domowe (kategoria 2)
- 2 – 3 i więcej osobowe gospodarstwo domowe (kategoria 3, 4, 5, 6, 7, 9, 11)

Rozkład zmiennej po powyższej zmianie prezentuje się następująco:

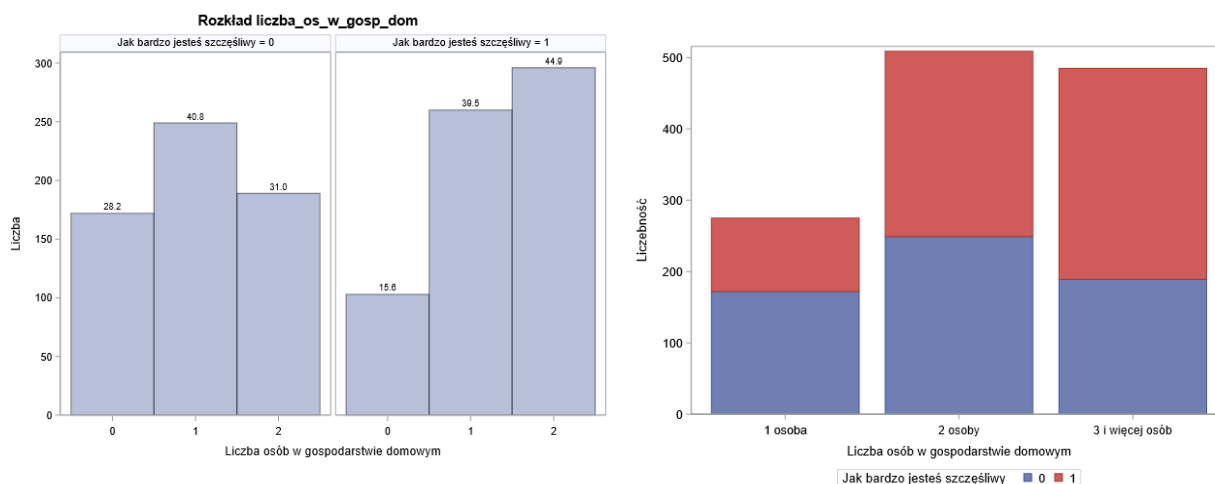
Frequency Row Percent Column Percent	Liczba osób w gospodarstwie domowym			
	Jak bardzo jesteś szczęśliwy			
	Liczba osób w gospodarstwie domowym	Jak bardzo jesteś szczęśliwy		
		0	1	Suma
	1 osoba	172	103	275
		62.55	37.45	
		28.20	15.63	
	2 osoby	249	260	509
48.92		51.08		
40.82		39.45		
3 i więcej osób	189	296	485	
	38.97	61.03		
	30.98	44.92		
Suma	610	659	1269	

Tabela 19. Tablica kontyngencji zmiennych hhmb i happy – po kategoryzacji



Wykres 20. Rozkład zmiennej hhmb – po kategoryzacji

Nadal najliczniejszą kategorią jest ta dotycząca gospodarstw dwuosobowych, jednak po połączeniu pozostałych kategorii w jedną rozkład został wyrównany. Obecnie najmniej respondentów znajduje się w kategorii mówiącej o jednoosobowym gospodarstwie domowym (ok. 22%).



Wykres 21. Rozkład zmiennej *hhmb* względem zmiennej *happy* (udział 0 i 1)

Powyższe wykresy przedstawiają rozkład zmiennej *hhmb* w zależności od poziomu szczęścia. Można zaobserwować, że liczne gospodarstwa domowe (3 i więcej osób) są szczęśliwsze od tych składających się tylko z jednej osoby. Natomiast osoby z dwuosobowych gospodarstw domowych stanowią bardzo zbliżony udział w obydwu grupach, osób szczęśliwych i nieszczęśliwych. Można również zauważyć, że im większe gospodarstwo domowe, tym większy udział osób szczęśliwych wśród respondentów.

## 7. estsz – liczba\_os\_w\_pracy

Zmienna *estsz* dotyczy liczby osób w pracy, wielkości zakładu pracy. Zmienna składa się z 5 kategorii:

- 1 – Poniżej 10
- 2 – od 10 do 24
- 3 – od 25 do 99
- 4 – od 100 do 499
- 5 – 500 lub więcej

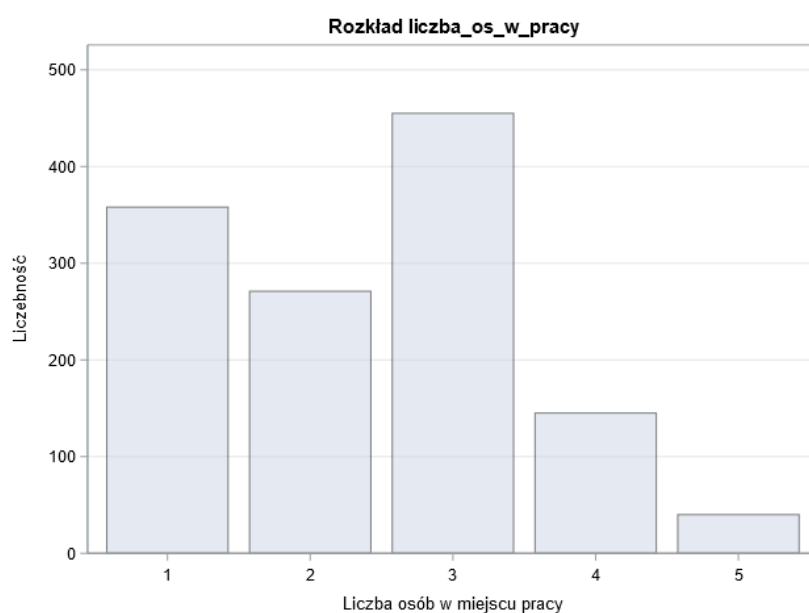
Największa liczba odpowiedzi należy do kategorii 3, jest to ok. 36% badanej próby. Najmniej liczna jest natomiast kategoria 5, obejmująca największe zakłady pracy, powyżej 500 osób.

Rozkład omawianej zmiennej jest zwizualizowany poniżej.

Liczebność  
Proc. wier.  
Proc. kol.

Tabela liczba_os_w_pracy od Szczęśliwy			
liczba_os_w_pracy(Liczba osób w miejscu pracy)	Szczęśliwy(Jak bardzo jesteś szczęśliwy)		
	0	1	Suma
1	166 46.37 27.21	192 53.63 29.14	358
2	141 52.03 23.11	130 47.97 19.73	271
3	221 48.57 36.23	234 51.43 35.51	455
4	66 45.52 10.82	79 54.48 11.99	145
5	16 40.00 2.62	24 60.00 3.64	40
Suma	610	659	1269

Tabela 20. Tablica kontyngencji zmiennych estsz i happy – przed kategoryzacją



Wykres 22. Rozkład zmiennej estsz – przed kategoryzacją

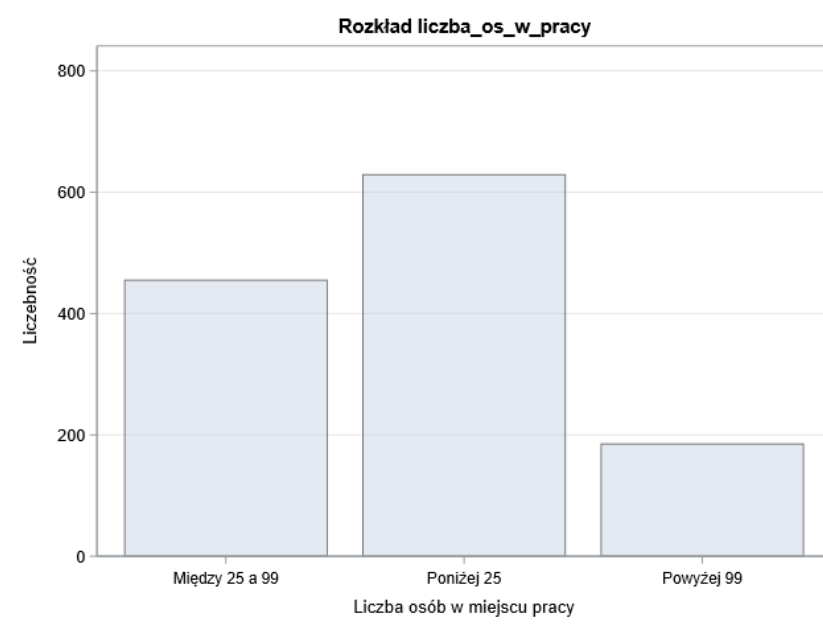
Ze względu na stosunkowo mało liczne kategorie 4 i 5 zdecydowano się na zmienienie kategorii zmiennych na następujące oraz kodowanie ich względem dominanty, czyli kategorii 3:

- 0 – poniżej 25 (kategoria 1, 2)
- 1 – między 25 a 99 (kategoria 3)
- 2 – powyżej 99 (kategoria 4, 5)



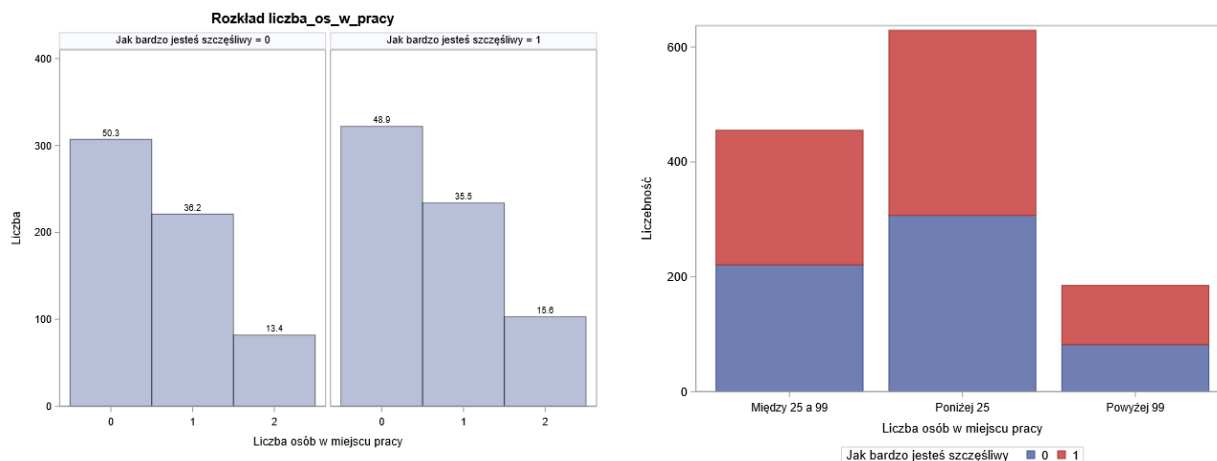
Frequency Row Percent Column Percent	Liczba osób w miejscu pracy			
	Jak bardzo jesteś szczęśliwy			
	Liczba osób w miejscu pracy	Jak bardzo jesteś szczęśliwy		
		0	1	Suma
		Między 25 a 99	221 48.57 36.23	234 51.43 35.51
Poniżej 25		307 48.81 50.33	322 51.19 48.86	629
Powyżej 99	82 44.32 13.44	103 55.68 15.63	185	
Suma	610	659	1269	

Tabela 21. Tablica kontyngencji zmiennych *estsz* i *happy* – po kategoryzacji



Wykres 23. Rozkład zmiennej *estsz* – po kategoryzacji

Powyższy wykres oraz tabela przedstawiają liczebności dla skategoryzowanej zmiennej *estsz*. Nadal najmniejszą liczebność ma kategoria ostatnia „Powyżej 99”, stanowi ona ok. 15%. Najliczniejszą kategorią nie jest już teraz kategoria „Między 25 a 99” (36%). Jest nią aktualnie nowo stworzona kategoria „Poniżej 25” (49%), powstała poprzez połączenie wcześniejszych kategorii 1 oraz 2.



Wykres 24. Rozkład zmiennej estsz względem zmiennej happy (udział 0 i 1)

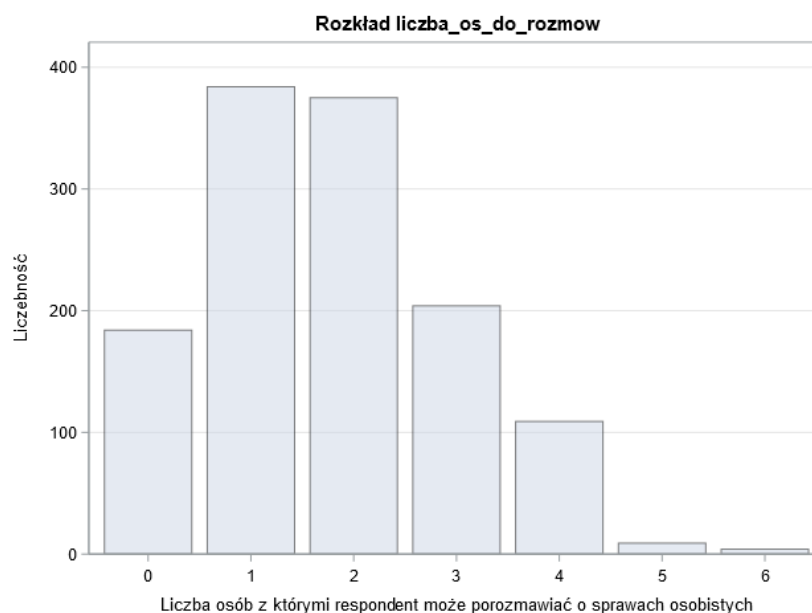
Patrząc na powyższy wykres można zauważyć, że rozkład zmiennej estsz w zależności od bycia szczęśliwym lub nieszczęśliwym jest prawie identyczny. Trochę większy udział w grupie osób szczęśliwych niż nieszczęśliwych stanowią respondenci pracujący w bardzo dużych zakładach pracy, Powyżej 500 osób. Natomiast w przypadku dwóch pozostałych kategorii udział w grupie osób szczęśliwych jest niższy o ok. 1 punkt procentowy niż w grupie osób nieszczęśliwych.

## 8. inprdsc – liczba\_os\_do\_rozmow

Kolejna zmienna to *inprdsc*, jest to liczba osób z którymi respondent może porozmawiać o sprawach prywatnych. Analizując rozkład zmiennej przedstawiony w tabeli oraz na wykresie poniżej można zauważyć, że zdecydowana większość osób z badanej próby ma w swoim otoczeniu 1 lub 2 osoby, z którymi może rozmawiać na tematy prywatne (57%). Dużo jest także respondentów, którzy nie mają ani jednej takiej osoby (14,5%).

Liczebność Proc. wier. Proc. kol.	Tabela liczba_os_do_rozmow od Szczęśliwy		
	Szczęśliwy(Jak bardzo jesteś szczęśliwy)		
liczba_os_do_rozmow(Liczba osób z którymi respondent może porozmawiać o sprawach osobistych)	0	1	Suma
0	114 61.96 18.69	70 38.04 10.62	184
1	196 51.04 32.13	188 48.96 28.53	384
2	170 45.33 27.87	205 54.67 31.11	375
3	88 43.14 14.43	116 56.86 17.60	204
4	37 33.94 6.07	72 66.06 10.93	109
5	4 44.44 0.66	5 55.56 0.76	9
6	1 25.00 0.16	3 75.00 0.46	4
Suma	610	659	1269

Tabela 22. Tablica kontyngencji zmiennych inprdsc i happy – przed kategoryzacją



Wykres 25. Rozkład zmiennej *inprdsc* – przed kategoryzacją

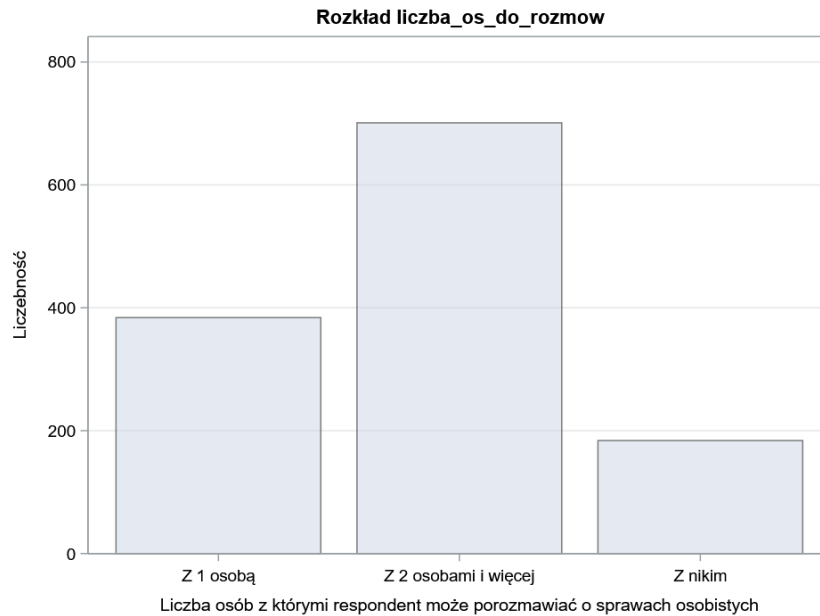
Kategorie 4, 5 oraz 6 są bardzo mało liczne, dlatego podjęto decyzję o przekodowaniu i kategoryzacji zmiennych. Nowe kodowanie przedstawione zostało poniżej.

- 0 – „Z nikim” (kategoria 0)
- 1 – „Z 1 osobą” (kategoria 1)
- 2 – „Z 2 osobami i więcej” (kategoria 2, 3, 4, 5, 6)

Po takim działaniu najbardziej liczną kategorią jest kategoria „Z 2 osobami i więcej”, stanowi ona ponad połowę obserwacji. Najmniej liczna kategoria to „Z nikim” (14,5%).

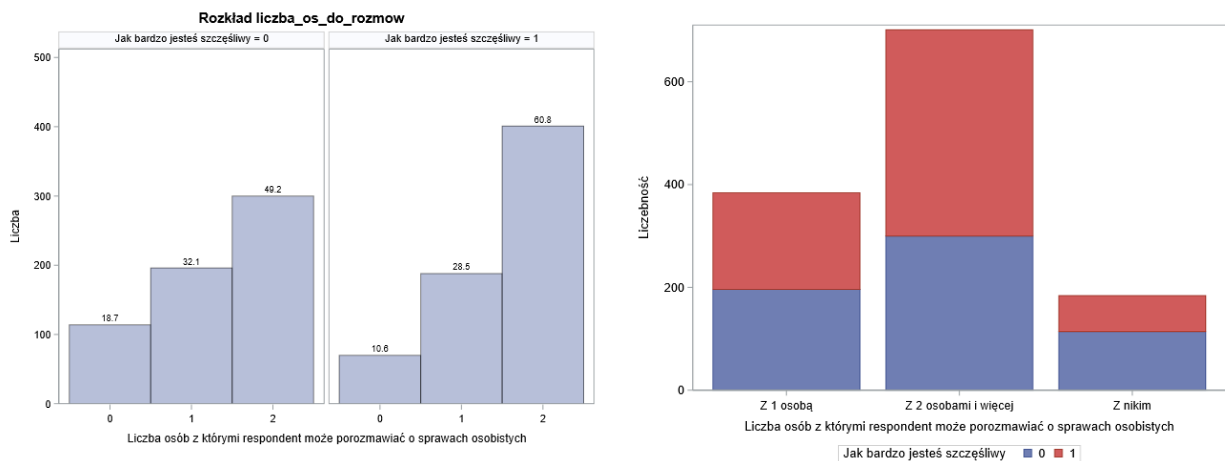
Frequency Row Percent Column Percent	Liczba osób z którymi respondent może porozmawiać o sprawach osobistych			
	Jak bardzo jesteś szczęśliwy			
	Jak bardzo jesteś szczęśliwy			
	0	1	Suma	
	Liczba osób z którymi respondent może porozmawiać o sprawach osobistych			
	Z 1 osobą	196 51.04 32.13	188 48.96 28.53	384
	Z 2 osobami i więcej	300 42.80 49.18	401 57.20 60.85	701
	Z nikim	114 61.96 18.69	70 38.04 10.62	184
	Suma	610	659	1269

Tabela 23. Tablica kontyngencji zmiennych *inpdsc* i *happy* – po kategoryzacji



*Wykres 26. Rozkład zmiennej inprdsc – po kategoryzacji*

Poniższe wykresy przedstawiają rozkład zmiennej *inprdsc* ze względu na zmienną objaśnianą *happy* oraz udział osób szczęśliwych/nieszczęśliwych w poszczególnych kategoriach. Największy udział osób szczęśliwych znajduje się w kategorii ostatniej, mówiącej że respondent posiada 2 lub więcej osób, z którymi może porozmawiać o swoich prywatnych sprawach. Widoczna jest także zależność, że im więcej takich osób posiada respondent tym większe prawdopodobieństwo, że jest on szczęśliwy. Wśród osób nieszczęśliwych dominująca kategoria to również „Z 2 osobami lub więcej”, jest to najliczniejsza kategoria, dlatego rozkład prezentuje się jak poniżej.



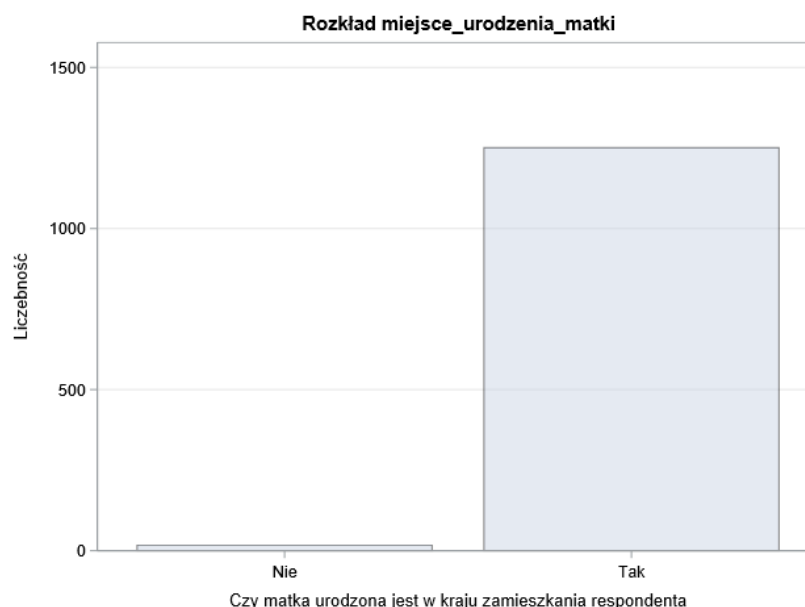
*Wykres 27. Rozkład zmiennej inprdsc względem zmiennej happy (udział 0 i 1)*

## 9. mocntr – miejsce\_urodzenia\_matki

Zmienna *mocntr* informuje o tym, czy matka respondenta urodzona jest w kraju zamieszkania respondenta. Składa się ona z dwóch kategorii „Tak” oraz „Nie” (1 i 2). Zdecydowana większość odpowiedzi to „Tak”, jest to ponad 90% analizowanej próby. Jej rozkład został przedstawiony poniżej.

Frequency Row Percent Column Percent	Czy matka urodzona jest w kraju zamieszkania respondenta			
	Jak bardzo jesteś szczęśliwy			
	Czy matka urodzona jest w kraju zamieszkania respondenta	Jak bardzo jesteś szczęśliwy		Suma
		0	1	
	Nie	8 44.44 1.31	10 55.56 1.52	18
	Tak	602 48.12 98.69	649 51.88 98.48	1251
	Suma	610	659	1269

Tabela 24. Tablica kontyngencji zmiennych *mocntr* i *happy*

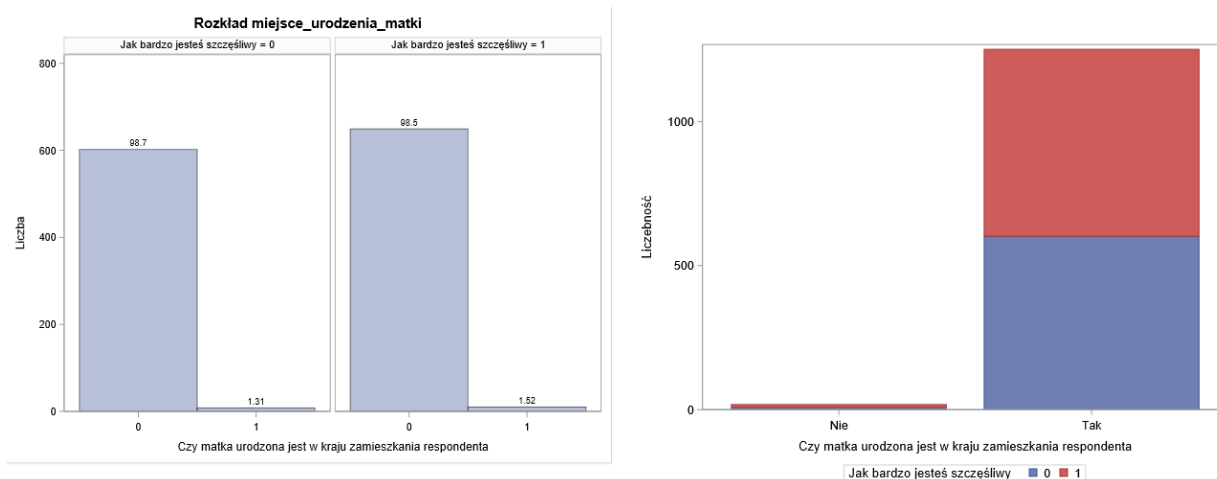


Wykres 28. Rozkład zmiennej *mocntr*

Kategorie zmiennej prezentują się następująco:

- 0 – „Tak” (kategoria 1)
- 1 – „Nie” (kategoria 2)

Wykresy przedstawione poniżej ukazują rozkład zmiennej *mocntr* ze względu na zmienną objaśnianą. W obydwóch kategoriach około 50% respondentów jest szczęśliwych.



Wykres 29. Rozkład zmiennej mocntr względem zmiennej happy (udział 0 i 1)

## Analiza doboru zmiennych

### Badanie korelacji

W modelu ekonometrycznym jednym z podstawowych wyznaczników przyjmowania zmiennych do modelu jest korelacja zmiennych. Korelację opisuje się jako liniową współzależność zmiennych i obiektów. Korelacja może przyjmować wartość między -1 a 1, przy czym im korelacja jest bliżej 0, tym związek między obserwowanymi zmiennymi jest mniejszy. W praktyce przyjmuje się, że wartość korelacji od 0 do 0,3 oznacza słabą zależność, od 0,3 do 0,5 umiarkowaną zależność, a od 0,5 do 1,0 informuje o związku silnym lub bardzo silnym.

Pożądanym jest uzyskanie jak najwyższej korelacji zmiennych niezależnych ze zmienną zależną i zarazem jak najmniejszej korelacji pomiędzy zmiennymi niezależnymi. Do zbadania korelacji pomiędzy zmiennymi stosuje się najczęściej jeden z dwóch wzorów: współczynnik korelacji rang Spearmana oraz współczynnik korelacji Pearsona.

Użyty w pracy współczynnik korelacji Pearsona wylicza się ze wzoru:

$$r_{xy} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

Oznacza to, że liczy on iloraz kowariancji zmiennych i iloczyn ich odchyleń standardowych. Wzór ten można rozpisać i przedstawić w następującej postaci:

$$\hat{r}_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

gdzie:

- X, Y to badane cechy,
- $\text{cov}(X, Y)$  jest nieobciążonym estymatorem kowariancji X i Y,
- $\sigma_x, \sigma_y$  to nieobciążone odchylenia standardowe zmiennych X i Y.

Jest to wówczas estymator współczynnika korelacji Pearsona. Współczynnik korelacji Pearsona wykorzystywany jest również w testowaniu statystycznym. Hipotezą zerową bazującego testu jest brak istotnej zależności między badanymi cechami.

Test wykorzystuje poniższą testującą o rozkładzie t-Studenta z  $n-2$  stopniami swobody, z wykorzystaniem poniższego wzoru:

$$t = \sqrt{n-2} * \sqrt{\frac{r^2}{1-r^2}}$$

Poniższe tabele przedstawiają macierz korelacji badanych zmiennych:

TYPE		wiek	plec	stan_cywilny	dochod_gospodarstwa_net	liczba_dzieci	liczba_os_w_gosp_dom	liczba_os_w_pracy	liczba_podwladnych	liczba_os_do_rozmow	subiektywny_stan_zdrowia	miejsce_urodzenia_matki
MEAN		57.06	0.57	2.21	1.4	1.55	1.17	0.65	0.12	1.41	1.31	0.01
STD		15.69	0.5	0.96	1.13	0.84	0.76	0.72	0.33	0.73	0.67	0.12
N		1269	1269	1269	1269	1269	1269	1269	1269	1269	1269	1269
CORR	wiek	1	0.06	-0.16	-0.41	0.34	-0.43	0.2	-0.01	-0.18	0.58	0.05
CORR	plec	0.06	1	-0.13	-0.06	0.11	-0.05	-0.08	-0.05	0	0.1	0.04
CORR	stan_cywilny	-0.16	-0.13	1	0.32	0.01	0.5	0.02	0.06	0.12	-0.12	0.03
CORR	dochod_gospodarstwa_net	-0.41	-0.06	0.32	1	-0.16	0.51	0.04	0.15	0.19	-0.36	-0.01
CORR	liczba_dzieci	0.34	0.11	0.01	-0.16	1	-0.01	0.02	-0.03	-0.08	0.23	0.02
CORR	liczba_os_w_gosp_dom	-0.43	-0.05	0.5	0.51	-0.01	1	-0.08	0.03	0.12	-0.24	-0.02
CORR	liczba_os_w_pracy	0.2	-0.08	0.02	0.04	0.02	-0.08	1	0.03	0	0.09	0
CORR	liczba_podwladnych	-0.01	-0.05	0.06	0.15	-0.03	0.03	0.03	1	0.05	-0.08	-0.04
CORR	liczba_os_do_rozmow	-0.18	0	0.12	0.19	-0.08	0.12	0	0.05	1	-0.18	0.02
CORR	subiektywny_stan_zdrowia	0.58	0.1	-0.12	-0.36	0.23	-0.24	0.09	-0.08	-0.18	1	0.02

Tabela 25. Korelacja zmiennych

TYPE		wiek	plec	stan_cywilny	dochod_gospodarstwa_net	liczba_dzieci	liczba_os_w_gosp_dom	liczba_os_w_pracy	liczba_podwladnych	liczba_os_do_rozmow	subiektywny_stan_zdrowia	miejsce_urodzenia_matki
CORR	wiek	silna	słaba	słaba	średnia	średnia	średnia	słaba	słaba	słaba	średnia	słaba
CORR	plec	słaba	silna	słaba	słaba	słaba	słaba	słaba	słaba	słaba	słaba	słaba
CORR	stan_cywilny	słaba	słaba	silna	średnia	słaba	średnia	słaba	słaba	słaba	słaba	słaba
CORR	dochod_gospodarstwa_net	średnia	słaba	średnia	silna	słaba	średnia	słaba	słaba	słaba	średnia	słaba
CORR	liczba_dzieci	średnia	słaba	słaba	słaba	silna	słaba	słaba	słaba	słaba	słaba	słaba
CORR	liczba_os_w_gosp_dom	średnia	słaba	średnia	średnia	słaba	silna	słaba	słaba	słaba	słaba	słaba
CORR	liczba_os_w_pracy	słaba	słaba	słaba	słaba	słaba	słaba	silna	słaba	słaba	słaba	słaba
CORR	liczba_podwladnych	słaba	słaba	słaba	słaba	słaba	słaba	słaba	silna	słaba	słaba	słaba
CORR	liczba_os_do_rozmow	słaba	słaba	słaba	słaba	słaba	słaba	słaba	słaba	silna	słaba	słaba
CORR	subiektywny_stan_zdrowia	średnia	słaba	słaba	średnia	słaba	słaba	słaba	słaba	słaba	silna	słaba
CORR	miejsce_urodzenia_matki	słaba	słaba	słaba	słaba	słaba	słaba	słaba	słaba	słaba	słaba	silna

Tabela 26. Korelacja zmiennych - ocena korelacji

Jak wynika z macierzy korelacji większość badanych zmiennych charakteryzuje się słabą lub umiarkowaną korelacją. Wartości bezwzględne korelacji wahają się między 0 a 0.58, co wskazuje na potencjalnie dobry dobór zmiennych objaśniających - są słabo skorelowane między sobą. Najsilniej skorelowana z innymi zmiennymi jest zmienna *wiek*, co jest w pełni



zrozumiałe. Niektóre zmienne takie, jak *Miejsce urodzenia matki*, *Liczba osób do rozmów* czy *Liczba podwładnych* nie wykazują żadnej korelacji (bliskie 0) z innymi zmiennymi objaśniającymi. Jednak przed podjęciem dalszych decyzji o eliminacji i przyjmowaniu zmiennych do modelu przeprowadzono analizę współliniowości.

### Badanie współliniowości - VIF

Badanie współliniowości przeprowadzono za pomocą współczynnika VIF (ang. Variance Inflation Factor), czyli tzw. wskaźnika inflacji wariancji lub też nazywanego czynnikiem rozdęcia wariancji. Współliniowość między zmiennymi może być ścisła lub przybliżona, współczynnik VIF bada tą drugą. Wzór na VIF wyraża się następującym wzorem:

$$VIF_k = \frac{1}{1 - R_k^2}$$

gdzie:

- $R_k^2$  - współczynnik determinacji pomiędzy zmienną  $k$  a pozostałymi zmiennymi niezależnymi.

Parametr VIF może osiągać minimalną wartość 1.0. W modelu regresji logistycznej przyjmuje się, że wartość  $VIF > 2,5$  wskazuje na występowanie współliniowości i należy ją usunąć z modelu.

Jeżeli jednak pomimo współliniowości nie chcemy usuwać zmiennej z modelu można spróbować usunąć inne zmienne z najniższą korelacją ze zmienną  $Y$ , zmienić zakres próby, rozszerzyć model o dodatkowe równanie, dokonać transformacji zmiennej odznaczającej się współliniowością lub użyć metody głównych składowych.

W celu dokonania pomiaru współczynnika VIF dla każdej ze zmiennych została wykorzystana procedura PROC REG. Jako zmienną objaśnianą ustalono badaną zmienną celu *happy* natomiast pozostałe zmienne zostały zadeklarowane jako zmienne objaśniające, co przedstawiają poniższe tabele:

Liczba obserwacji wczytanych	1269
Liczba obserwacji użytych	1269

Analiza wariancji					
Źródło	DF	Suma kwadratów	Średnia kwadratów	Wartość F	Pr. > F
Model	11	44.35693	4.03245	18.61	<.0001
Błąd	1257	272.42006	0.21672		
Razem skorygowane	1268	316.77699			

Pierw. z MSE	0.46553	R-kwadrat	0.1400
Średnia zależna	0.51931	Skor. R-kw.	0.1325
Wsp. zmienności	89.64542		

Tabela 27. Analiza wariancji

Oceny parametrów								
Zmienna	Etykieta	DF	Ocena parametru	Błąd standardowy	Wartość t	Pr. >  t	Tolerancja	Inflacja wariancji
Intercept	Intercept	1	0.46997	0.07963	5.90	<.0001	.	0
wiek	Wiek	1	-0.00113	0.00119	-0.96	0.3391	0.49412	2.02381
plec	Płeć	1	0.06307	0.02702	2.33	0.0197	0.95444	1.04773
stan_cywilny	Stan cywilny	1	0.06387	0.01605	3.98	<.0001	0.72315	1.38283
dochod_gospodarstwa_net	Łączny dochód netto gospodarstwa domowego	1	0.05483	0.01460	3.75	0.0002	0.63135	1.58391
liczba_dzieci	Liczba dzieci	1	0.00118	0.01697	0.07	0.9444	0.84344	1.18562
liczba_os_w_gosp_dom	Liczba osób w gospodarstwie domowym	1	-0.01016	0.02339	-0.43	0.6641	0.54620	1.83082
liczba_os_w_pracy	Liczba osób w miejscu pracy	1	0.03253	0.01880	1.73	0.0838	0.93096	1.07416
liczba_podwladnych	Jakiego typu stanowisko respondent obejmuje w pracy zwykłe/kierownicze	1	0.09860	0.04074	2.42	0.0157	0.96482	1.03647
liczba_os_do_rozmow	Liczba osób z którymi respondent może porozmawiać o sprawach osobistych	1	0.03396	0.01851	1.83	0.0668	0.93820	1.06587
subiektywny_stan_zdrowia	Subiektywny stan zdrowia	1	-0.16120	0.02440	-6.61	<.0001	0.63516	1.57440
miejsce_urodzenia_matki	Czy matka urodzona jest w kraju zamieszkania respondenta	1	0.05341	0.11103	0.48	0.6306	0.99069	1.00940

Tabela 28. Oceny parametrów zmiennych

Z powyższych tabel wynika, że współczynnik VIF dla żadnej zmiennej nie przekroczył 2.5 i na tej podstawie zdecydowano się przyjąć wszystkie zmienne objaśniające do dalszych analiz.

## Współczynnik V-Cramera

Jedną z miar zależności jest współczynnik V-Cramera. Jest to współczynnik określający poziom zależności pomiędzy dwiema zmiennymi nominalnymi, spośród których co najmniej jedna przyjmuje więcej niż dwie wartości. Twórcą tego współczynnika jest szwedzki statystyk Harald Cramér. Współczynnik V-Cramera jest rozwinięciem czy udoskonaleniem współczynnika  $\chi^2$ , gdyż współczynnik  $\chi^2$  może być używany tylko wtedy, gdy badamy związek

między dwiema zmiennymi, które obie są nominalne i dychotomiczne. Tabela kontyngencji ma wtedy wymiary  $2 \times 2$ . W przypadku bardziej złożonych tabel kontyngencji (np.  $2 \times 3$ ,  $3 \times 3$  etc.) należy użyć współczynnika V-Cramera. Wyraża się on następującym wzorem:

$$V = \sqrt{\frac{\chi^2}{n \cdot \min(r-1, k-1)}}$$

gdzie:

- **V** - Współczynnik V-Cramera pomiędzy dwoma zmiennymi
- $\chi^2$  - wynik testu Chi-kwadrat dla pary zmiennych
- **n** - liczba obserwacji
- **r** - liczba poziomów jednej zmiennej
- **k** - liczba drugiej jednej zmiennej
- $\min(r-1, k-1)$  - wartość mniejsza z dwóch: (k-1) lub (r-1).

Wartość współczynnika V-Cramera przyjmuje wartości z przedziału  $< 0; 1 >$ . Im jego wartość jest bliższa 0, tym siła związku między badanymi cechami jest słabsza, czyli dla 0 nie ma żadnego związku. I analogicznie - im wartość V-Cramera jest bliższa 1, tym siła związku między badanymi cechami jest mocniejsza (dla 1 jest najmocniejsza). Ponadto należy mieć na uwadze, że V-Cramer zależy też od wielkości tabel kontyngencji, więc powinno się porównywać tylko tabele o tej samej wielkości. Dokładniejsza interpretacja wartości współczynnika przedstawia się:

wartość	interpretacja
0-0.2	bardzo słaby związek
0.2-0.4	słaby związek
0.4-0.6	umiarkowany związek
0.6-0.8	silny związek
0.8-1	bardzo silny związek

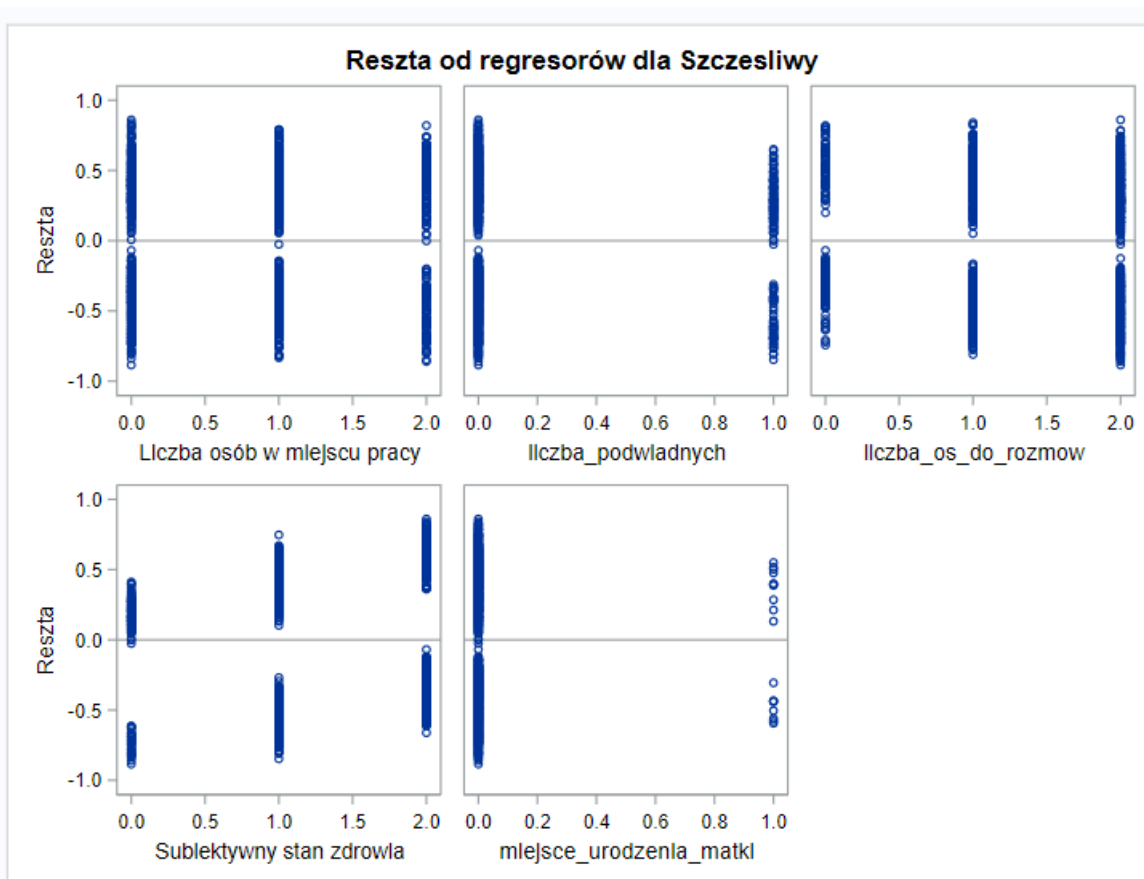
Tabela 28. Interpretacja wartości współczynnika V-Cramera

Współczynnik V Cramera jest istotny statystycznie, gdy wartość p, wyznaczana na podstawie statystyki testu Chi-kwadrat i rozkładu Chi-kwadrat jest mniejsza lub równa alfa, czyli poziom istotności (tutaj przyjęto  $\alpha = 0.05$ ). Poniższa tabela przedstawia wartość czynnika V-Cramera oraz prawdopodobieństwo oszacowane na podstawie przyjętego alfa oraz statystyki Chi-Kwadrat dla wszystkich zmiennych objaśniających:

Zmienna	P dla Chi Kwadrat	V-cramer
wiek	<.0001	0.3324
plec	0.773	0.0081
stan_cywilny	<.0001	0.2016
dochod_gospodarstwa_net	<.0001	0.2658
liczba_dzieci	0.0261	0.0854
liczba_os_w_gosp_dom	<.0001	0.1760
liczba_os_w_pracy	0.5426	0.0310
liczba_podwladnych	0.0002	0.1064
liczba_os_do_rozmow	<.0001	0.1357
subiektywny_stan_zdrowia	<.0001	0.2980
miejsce_urodzenia_matki	0.7565	0.0087

Tabela 29. Współczynniki V-Cramera dla zmiennych

Na pomarańczowo zaznaczono zmienne, które zdecydowano się odrzucić do dalszych badań ze względu na niski współczynnik V-Cramera, są to: *pleć* oraz *miejsce urodzenia matki*. Niektóre zmienne takie, jak liczba osób w pracy również posiadają niską wartość tego współczynnika jednak na podstawie informacji dostarczonych w analizie korelacji oraz współliniowości zdecydowano się je zachować do dalszych analiz i przyjąć do budowy modelu.



Wykres 30. Wykresy reszt regresorów

## Estymacja binarnego modelu regresji logistycznej

Do modelowania zmiennych zależnych posłużono się binarnym modelem regresji logistycznej, który modeluje prawdopodobieństwo zajścia jednego, z dwóch alternatywnych, zdarzenia. Analiza takiego modelu polega na oszacowaniu parametrów modelu logistycznego, inaczej nazywanych współczynnikami kombinacji liniowej. W binarnym modelu regresji występuje pojedyncza binarna zmienna zależna, oznaczona jako “0” i “1” (w naszym modelu “0” oznacza nieszczęśliwy, “1” szczęśliwy). Zmienne zależne mogą przyjmować zarówno postać binarną jak i ciągłą. Do weryfikacji takiego modelu wykorzystuje się test oceny proporcjonalności szans oraz test Walda badający istotność statystyczną parametrów.

Test punktowy służyć będzie do przetestowania założenia o proporcjonalności szans. Przeprowadza się go przy użyciu statystyki chi-kwadrat. Wynik tego testu pozwala na porównanie modelu uporządkowanego z modelem wielowymiarowym. Testowane są w nim następujące hipotezy:

- $H_0$ : Szanse są proporcjonalne
- $H_1$ : Szanse nie są proporcjonalne.

Jeżeli podejmujemy decyzję o odrzuceniu hipotezy zerowej, wtedy model jest słaby i niezbędne jest przejście na model wielowymiarowy, który posiada mniejszą moc wykrywania zależności.

Do oceny istotności statystycznej współczynników przeważnie wykorzystuje się test Walda, również opartego o statystykę chi-kwadrat. Sprawdza się przy nim następujące hipotezy:

- $H_0$ :  $\beta$  nie jest istotny statystycznie ( $\beta = 0$ )
- $H_1$ :  $\beta$  jest istotny statystycznie ( $\beta \neq 0$ ).

#### Schemat budowy modelu regresji wielomianowej

1. Pierwszy etap obejmuje model posiadający wszystkie efekty główne. Następnie badano, które zmienne są nieistotne statystycznie oraz które są zmiennymi zakłócającymi. Usuając po kolei zmienne najbardziej nieistotne, sprawdzano czy nastąpiła większa niż 15% zmiana oceny parametrów, jeśli tak zmienna była przywracana do modelu jako zmienna zakłócająca (oraz jeśli występuje zależność przyczynowo skutkowa między badanymi zmiennymi). Ze względu na powyższe kryteria w modelu zostały uwzględnione wszystkie zmienne jako efekty główne.
2. W kolejnym etapie uwzględniono interakcje pomiędzy wszystkimi zmiennymi oraz ich efekty główne, (nawet jeśli okazały się nieistotne statystycznie, ponieważ interakcja takiej zmiennej z inną zmienną może okazać się istotna statystycznie. Następnie dokonano selekcji interakcji za pomocą metody selekcji wstecznej oraz krokowej.
3. W ostatnim etapie porównano i wybrano najlepszy model selekcji, który zredukowano o zmienne nieistotne statystycznie (oraz nie będące w żadnej istotnie statystycznej interakcji) i zmienne nie będące zmiennymi zakłócającymi.

## Estymacja modelu

### Etap I

W pierwszej kolejności wyestymowano model wykorzystując wszystkie zmienne niezależne, ponieważ w analizie badanych zmiennych wykazano brak wyraźnych podstaw do ich odrzucenia. Uwzględniono w nim jedynie efekty główne (bez interakcji). Tabela X ukazuje podstawowe informacje dotyczące zbudowanego modelu.

Informacje o modelach		
Zbiór	WORK.SORTTEMPTABLESORTED	
Zmienna objaśniana	Szczesliwy	Jak bardzo jesteś szczeniwy
Liczba poziomów odpowiedzi	2	
Model	logit binarny	
Metoda optymalizacji	Ocena Fishera	

Liczba obserwacji wczytanych	1269
Liczba obserwacji użytych	1269

Profil odpowiedzi		
Wartość uporządkowana	Szczeniwy	Całkowita liczebność
1	0	610
2	1	659

Wymodelowane prawdopodobieństwo wynosi Szczeniwy=1.

Tabela 30. Informacje o modelu ze wszystkimi zmiennymi

Liczba obserwacji wczytanych równa 1269 okazała się taka sama, jak liczba obserwacji użytych, a więc do stworzenia modelu użyto wszystkich obserwacji. Ten model oraz wszystkie poniższe był estymowany dla wartości “1”, co oznacza osobę szczęśliwą. Model spełnia kryterium zbieżności, dzięki czemu możliwa jest interpretacja uzyskanych wyników.

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

Statystyki dopasowania		
Kryterium	Tylko wyraz wolny	Wyraz wolny i współzmiennie
AIC	1759.315	1606.251
SC	1764.461	1719.462
-2 log L	1757.315	1562.251

Tabela 31. Status zbieżności i statystyki dopasowania - model ze wszystkimi zmiennymi

Statystyki dopasowania świadczą o tym, że model nieuporządkowany z modelami głównymi jest lepszy, niż model tylko z wyrazem wolnym oraz ilość zmiennych w uwzględnianym modelu nie jest za duża.

Poniższa Tabela testuje globalną hipotezę o zerowym współczynniku  $\beta$  (nieistotność statystyczna współczynników), gdzie hipotezą alternatywną jest istotność statystyczna, przynajmniej jednego współczynnika. Wartość p w teście Walda powinna być mniejsza niż poziom istotności równego 5%, aby mieć podstawy do odrzucenia hipotezy zerowej.

Testowanie globalnej hipotezy zerowej: BETA=0			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kw.
Il. wiarygodn.	195.0642	21	<.0001
Mn. Lagrange'a	184.1547	21	<.0001
Walda	164.1837	21	<.0001

Tabela 32. Test Walda - model ze wszystkimi zmiennymi

Weryfikacja hipotezy o łącznej nieistotności statystycznej zmiennych wykazuje p-value mniejsze niż 0.001, a więc mamy podstawy do odrzucenia hipotezy zerowej, na poczet hipotezy alternatywnej. Oznacza to, że przynajmniej jedna zmienna w tym modelu jest istotna statystycznie.

Analiza efektów typu 3			
Efekt	DF	Chi-kwadrat Walda	Pr. > chi-kw.
wiek	1	1.5137	0.2186
plec	1	4.0995	0.0429
stan_cywilny	3	12.6414	0.0055
dochod_gospodarstwa_	3	14.7197	0.0021
liczba_dzieci	3	1.8178	0.6111
liczba_os_w_gosp_dom	2	0.0987	0.9518
liczba_os_w_pracy	2	2.9105	0.2333
liczba_podwladnych	1	6.1881	0.0129
liczba_os_do_rozmow	2	3.5615	0.1685
subiektywny_stan_zdro	2	41.0350	<.0001
miejsce_urodzenia_ma	1	0.2374	0.6261

Tabela 33. Analiza efektów typu 3 - model ze wszystkimi zmiennymi

Istotność statystyczna zmiennych była ustalana na podstawie p-value, możliwe do odczytania z tabeli “Analiza efektów typu 3”. Wartość mieszcząca się w kolumnie Pr.>chi-kw. powinna być mniejsza niż 0.05, aby parametr był istotny statystycznie. Model posiadający



jedynie efekty główne, jako zmienne istotnie statystycznie traktuje płeć, stan, cywilny, dochód\_gospodarstwa, liczba\_podwładnych oraz subiektywny\_stan\_zdrowia. Zmienne najbardziej nieistotne poddano badaniu czy są one zmiennymi zakłócającymi. Ich wpływ oraz przyczynowość skutkową wykazały, że są one zakłócające, więc są uwzględnione w dalszym badaniu.

Tabela "Informacje o poziomach klasyfikacji" dobrze obrazuje w jaki sposób oznaczono daną zmienną oraz który poziom zmiennych wielopoziomowych został oznaczony jako bazowy.

Informacje o poziomach klasyfikacji				
Klasa	Wartość	Zmienne projektowe		
płeć	K	1		
	M	0		
stan_cywilny	Nie w związku lub w związku nieformalnym	1	0	0
	Nigdy w formalnym związku	0	1	0
	Rozwiedziony/owdowiony	0	0	1
	W związku małżeńskim	0	0	0
dochod_gospodarstwa_net	1-2 decyl	1	0	0
	3-4 decyl	0	1	0
	5-6 decyl	0	0	1
	7-10 decyl	0	0	0
liczba_dzieci	1 dziecko	1	0	0
	2 dzieci	0	1	0
	3 i więcej dzieci	0	0	1
	Brak dzieci	0	0	0
liczba_os_w_gosp_dom	1 osoba	1	0	
	2 osoby	0	1	
	3 i więcej osób	0	0	
liczba_os_w_pracy	Miedzy 25 a 99	1	0	
	Ponizej 25	0	1	
	Powyzej 99	0	0	
liczba_podwladnych	Stanowisko kierownicze	1		
	Zwykle stanowisko	0		
liczba_os_do_rozmow	Z 1 osoba	1	0	
	Z 2 osobami i więcej	0	1	
	Z nikim	0	0	
subiektywny_stan_zdrowia	Bardzo dobre	1	0	
	Dobre	0	1	
	Srednie i gorsze	0	0	
miejsce_urodzenia_matki	Nie	1		
	Tak	0		

Tabela 34. Informacje o poziomach klasyfikacji - model ze wszystkimi zmiennymi

## Modele z interakcjami

### Etap II

W następnym kroku zbudowano modele ze wszystkimi efektami głównymi oraz ich interakcjami. Wykorzystano wszystkie zmienne (bez względu na ich statystyczną istotność), gdyż interakcje z tymi zmiennymi mogą okazać się istotne statystycznie. Najpierw został oszacowany model ze wszystkimi interakcjami, następnie dokonano selekcji interakcji metodą eliminacji wstecznej oraz selekcji krokowej.

### Model ze wszystkimi interakcjami

Statystyki dopasowania			
Kryterium	Tylko wyraz wolny	Wyraz wolny i współzmiennie	
AIC	1759.315	1731.038	
SC	1764.461	2832.279	
-2 log L	1757.315	1303.038	

R-kwadrat	0.3009	Maksymalnie przeskalowane R-kwadrat	0.4014
-----------	--------	-------------------------------------	--------

Testowanie globalnej hipotezy zerowej: BETA=0			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kw.
Il. wiarygodn.	454.2767	213	<.0001
Mn. Lagrange'a	367.2451	213	<.0001
Walda	233.6538	213	0.1581

Tabela 35. Statystyki dopasowania – model ze wszystkimi interakcjami

W modelu zostało uwzględnionych ponad 50 interakcji, dlatego też nie umieszczono tutaj tabeli ze statystykami dla poszczególnych zmiennych. Jednak większość tych zmiennych okazała się być nieistotna statystycznie.

### Model metodą selekcji wstecznej

W wyniku zastosowanie metody wstecznej selekcji interakcji do modelu zostały wybrane tylko trzy interakcje: zmiennych wiek i liczba\_os\_w\_pracy, subiektywny\_stan\_zdrowia i stan\_cywilny oraz liczba\_os\_w\_gosp\_dom i liczba\_os\_do\_rozmowy. Zastosowanie odpowiedniej selekcji interakcji pozwoliło na osiągnięcie lepszego dopasowania modelu do danych. Statystyki AIC i SC znacznie się polepszyły (odpowiednio z 1731 na 1586 i z 2382 na 1761). Wartość funkcji wiarygodności również wzrosła.

Statystyki dopasowania			
Kryterium	Tylko wyraz wolny	Wyraz wolny i współzmiennie	
AIC	1759.315	1586.792	
SC	1764.461	1761.755	
-2 log L	1757.315	1518.792	

R-kwadrat	0.1714	Maksymalnie przeskalowane R-kwadrat	0.2286
-----------	--------	-------------------------------------	--------

Testowanie globalnej hipotezy zerowej: BETA=0			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kw.
Il. wiarygodn.	238.5233	33	<.0001
Mn. Lagrange'a	216.9120	33	<.0001
Walda	176.3937	33	<.0001

Testy łączne			
Efekt	DF	Chi-kwadrat Walda	Pr. > chi-kw.
liczba_dzieci	3	1.1160	0.7732
liczba_os_w_gosp_dom	2	4.1980	0.1226
liczba_os_w_pracy	2	10.8498	0.0044
subiektywny_stan_zdro	2	4.9106	0.0858
stan_cywilny	3	17.1170	0.0007
liczba_podwladnych	1	4.6738	0.0306
plec	1	3.4533	0.0631
miejsce_urodzenia_ma	1	0.1991	0.6554
dochod_gospodarstwa_	3	20.4145	0.0001
liczba_os_do_rozmow	2	1.6312	0.4424
wiek	1	5.8057	0.0160
wiek*liczba_os_w_pra	2	13.8374	0.0010
liczba_os*liczba_os_	4	11.8641	0.0184
subiektywny*stan_cywil	6	14.1821	0.0277

Tabela 36. Statystyki dopasowania, testy łączne – model metodą selekcji wstecznej

### Model metodą selekcji krokowej

Następnie dokonano selekcji interakcji metodą krokową. W wyniku selekcji tą metodą w modelu umieszczone zostały intrakcje zmiennych: wiek i liczba\_dzieci, wiek i liczba\_os\_w\_pracy, liczba\_os\_w\_gosp\_domowym i liczba\_os\_do\_rozmowy oraz subiektywny\_stan\_zdrowia i stan\_cywilny. Wartość statystyki AIC dla tej metody była niewiele lepsza niż dla metody wstecznej, jednak statystyka SC i funkcja wiarygodności znacząco pogorszyły się.

Statystyki dopasowania			
Kryterium	Tylko wyraz wolny	Wyraz wolny i współzmiennie	
AIC	1759.315	1583.790	
SC	1764.461	1774.191	
-2 log L	1757.315	1509.790	

R-kwadrat	0.1772	Maksymalnie przeskalowane R-kwadrat	0.2364
-----------	--------	-------------------------------------	--------

Testowanie globalnej hipotezy zerowej: BETA=0			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kw.
Il. wiarygodn.	247.5254	36	<.0001
Mn. Lagrange'a	224.5006	36	<.0001
Walda	181.5112	36	<.0001

Testy łączne			
Efekt	DF	Chi-kwadrat Walda	Pr. > chi-kw.
plec	1	3.5973	0.0579
liczba_dzieci	3	5.9894	0.1121
liczba_os_w_gosp_dom	2	3.1768	0.2043
liczba_os_w_pracy	2	9.6783	0.0079
subiektywny_stan_zdro	2	5.1081	0.0778
miejsce_urodzenia_ma	1	0.2298	0.6317
wiek	1	0.1591	0.6900
stan_cywilny	3	19.6251	0.0002
liczba_podwladnych	1	5.6297	0.0177
dochod_gospodarstwa_	3	20.8285	0.0001
liczba_os_do_rozmow	2	1.3772	0.5023
wiek*liczba_dzieci	3	8.4530	0.0375
wiek*liczba_os_w_pra	2	12.2213	0.0022
liczba_os*liczba_os_	4	11.1839	0.0246
subiektywny*stan_cywil	6	18.2386	0.0057

Tabela 37. Statystyki dopasowania, testy łączne – model metodą selekcji krokowej

## Model zredukowany

### Etap III

W celu zwiększenia liczby stopni swobody, zredukowano zmienne w modelu oszacowanym metodą selekcji wstecznej. Zmienne, które w tym modelu okazały się być nieistotne statystycznie oraz nie zawierały się w żadnej istotnej interakcji, to: liczba\_dzieci, miejsce\_urodzenia\_matki oraz płeć. Przy czym zmienna płeć jest bardzo blisko granicy istotności. W pierwszej kolejności zredukowano model o zmienną liczba\_dzieci, nie zaobserwowane zostały znaczne zakłócenia, dopasowanie modelu do zmiennych uległo poprawie. Następnie usunięto z modelu zmienną miejsce\_urodzenia\_matki, również nie

spowodowało to zakłóceń. Można na tej podstawie stwierdzić, że zmienne liczba\_dzieci oraz miejsce\_urodzenia\_matki nie są zmiennymi zakłócającymi. W trzecim kroku redukcji usunięto zmienną płeć, w tym przypadku wyraźne były zakłócenia. Dodatkowo po usunięciu pierwszych dwóch zmiennych, zmienna płeć jeszcze bardziej zbliżyła się do granicy istotności. Dlatego też zakończono redukcję zmiennych na zmiennej miejsce\_urodzenia\_matki. W wyniku otrzymano zdecydowanie bardziej dopasowany do zmiennych model. Zarówno statystyki AIC i S.C., jak i funkcja wiarygodności uległy polepszeniu. Maksymalnie przeskalowane R-kwadrat jest nieznacznie gorsze od modelu wyjściowego, jednak może to wynikać z mniejszej liczby zmiennych w modelu zredukowanym.

Statystyki dopasowania			
Kryterium	Tylko wyraz wolny	Wyraz wolny i współzmiennie	
AIC	1759.315	1580.101	
SC	1764.461	1734.480	
-2 log L	1757.315	1520.101	

R-kwadrat	0.1705	Maksymalnie przeskalowane R-kwadrat	0.2274
-----------	--------	-------------------------------------	--------

Testowanie globalnej hipotezy zerowej: BETA=0			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kw.
Il. wiarygodn.	237.2142	29	<.0001
Mn. Lagrange'a	215.8441	29	<.0001
Walda	175.6289	29	<.0001

Testy łączne			
Efekt	DF	Chi-kwadrat Walda	Pr. > chi-kw.
wiek	1	6.0864	0.0136
plec	1	3.7950	0.0514
stan_cywilny	3	17.3661	0.0006
dochod_gospodarstwa_	3	21.5142	<.0001
liczba_os_w_gosp_dom	2	4.3198	0.1153
liczba_os_w_pracy	2	10.9501	0.0042
liczba_podwladnych	1	4.4295	0.0353
liczba_os_do_rozmow	2	1.6394	0.4406
subiektywny_stan_zdro	2	4.9080	0.0860
wiek*liczba_os_w_pra	2	14.0002	0.0009
stan_cywi*subiektywny	6	14.3473	0.0260
liczba_os*liczba_os_	4	12.2437	0.0156

Tabela 38. Statystyki dopasowania, testy łączne – model zredukowany

## Wybór modelu

Poniższa tabela prezentuje statystyki dopasowania wyrazów wolnych i współczynników zmiennych dla dwóch wybranych modeli.

	Model zredukowany	Model selekcji krokowej
Akaike Information Criterion	1580	1584
Schwarz Criterion	1734	1774
-2 log L	1520	1510
maksymalne przeskalowanie R-kwadrat	0.2274	0.2364

*Tabela 39. Porównanie statystyk dopasowania – model zredukowany i model selekcji krokowej*

Kryteria informacyjne AIC, SC służą do porównania modeli między sobą, przy czym model o najmniejszych wartościach jest najlepszy. W tym przypadku jest to model zredukowany. Współczynnik R-kwadrat największą wartość przyjął w modelu selekcji krokowej jednak współczynnik R-kwadrat nie uwzględnia stopnia rozbudowania modelu.

W celu weryfikacji hipotez badawczych przyjrano się dokładniej ostatecznej zredukowanej formie modelu. Model szukał odpowiedzi na pytania “Czy interakcja pomiędzy wiekiem a ilością posiadanych dzieci jest istotna statystycznie?” oraz “Czy powyżej pewnego wieku brak dzieci czyni ludzi bardziej nieszczęśliwymi?”. Niestety zmienna liczba dzieci okazała się nieistotna statystycznie oraz nie znajdowała się w żadnej istotnie statystycznej interakcji, więc została usunięta z modelu w etapie III, w którym estymowany model był redukowany. Oznacza to, że powyższa praca odpowiada negatywnie na pytanie czy interakcja między wiekiem a ilością dzieci jest istotna statystycznie, natomiast nie jest w stanie odpowiedzieć na pytanie czy powyżej pewnego wieku ludzie nieposiadający dzieci są nieszczęśliwi.

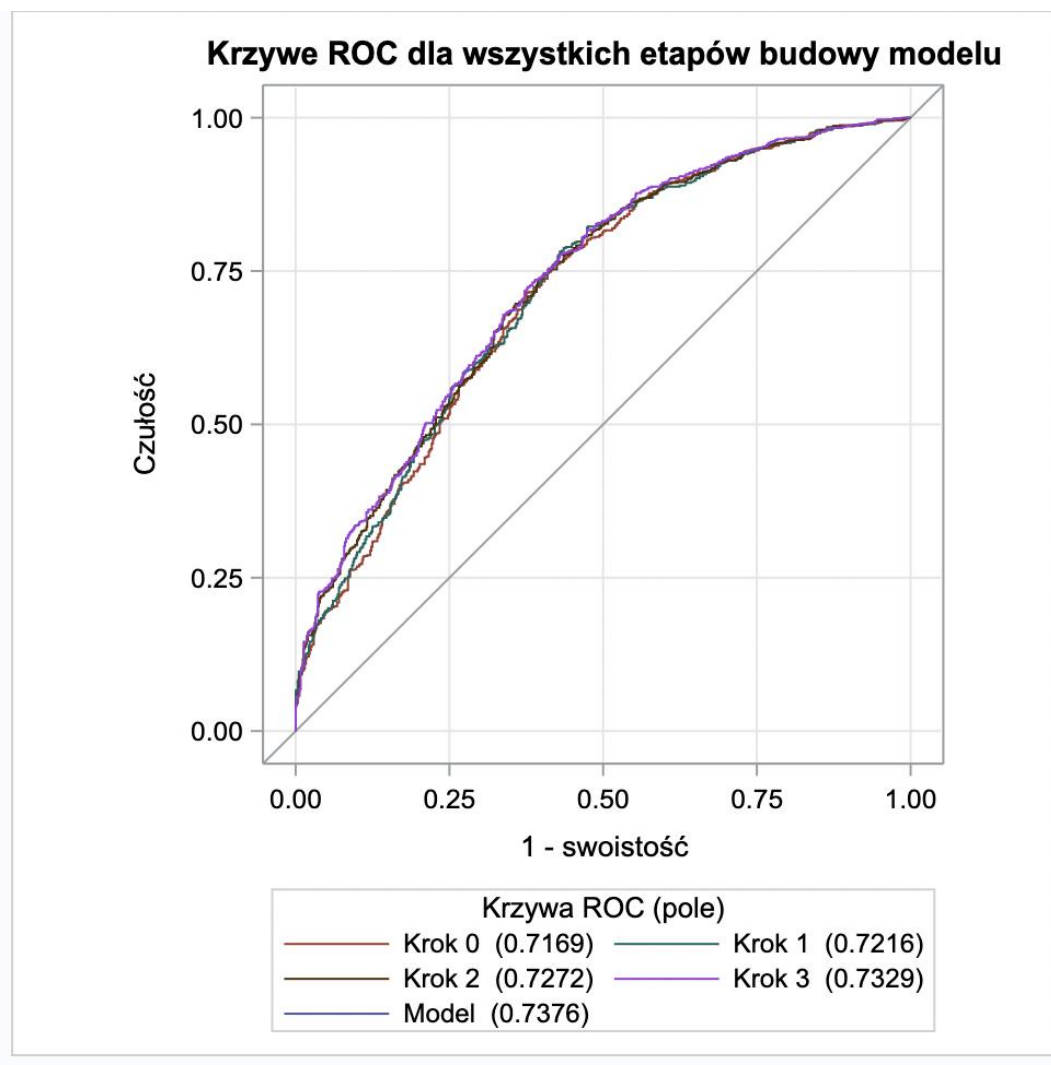
Analiza ocen maksymalnej wiarygodności							
Parametr			DF	Ocena	Błąd standardowy	Chi-kwadrat Walda	Pr. > chi-kw.
Intercept			1	-3.4532	1.0878	10.0778	0.0015
wiek			1	0.0291	0.0118	6.0864	0.0136
plec	K		1	0.2528	0.1298	3.7950	0.0514
stan_cywilny	Nie w związku lub w związku nieformalnym		1	2.0786	0.7755	7.1837	0.0074
stan_cywilny	Nigdy w formalnym związku		1	1.8361	0.8446	4.7255	0.0297
stan_cywilny	Rozwiedziony/owdowiony		1	1.1863	0.7720	2.3612	0.1244
dochod_gospodarstwa_	1-2 decyl		1	-0.7777	0.2183	12.6923	0.0004
dochod_gospodarstwa_	3-4 decyl		1	-0.8439	0.1901	19.7132	<.0001
dochod_gospodarstwa_	5-6 decyl		1	-0.3876	0.1966	3.8872	0.0487
liczba_os_w_gosp_dom	1 osoba		1	-0.7090	0.4500	2.4817	0.1152
liczba_os_w_gosp_dom	2 osoby		1	0.1711	0.4159	0.1691	0.6809
liczba_os_w_pracy	Miedzy 25 a 99		1	1.8710	0.8116	5.3148	0.0211
liczba_os_w_pracy	Ponizej 25		1	2.5568	0.7761	10.8520	0.0010
liczba_podwladnych	Stanowisko kierownicze		1	0.4215	0.2003	4.4295	0.0353
liczba_os_do_rozmow	Z 1 osoba		1	-0.2428	0.3428	0.5015	0.4788
liczba_os_do_rozmow	Z 2 osobami i wiecej		1	0.0391	0.3232	0.0146	0.9038
subiektywny_stan_zdro	Bardzo dobre		1	14.4553	256.3	0.0032	0.9550
subiektywny_stan_zdro	Dobre		1	1.9069	0.8610	4.9054	0.0268
wiek*liczba_os_w_pra	Miedzy 25 a 99		1	-0.0320	0.0129	6.1775	0.0129
wiek*liczba_os_w_pra	Ponizej 25		1	-0.0467	0.0125	13.8557	0.0002
stan_cywi*subiektywny	Nie w związku lub w związku nieformalnym	Bardzo dobre	1	-12.5795	256.3	0.0024	0.9609
stan_cywi*subiektywny	Nie w związku lub w związku nieformalnym	Dobre	1	-1.3366	0.8777	2.3188	0.1278
stan_cywi*subiektywny	Nigdy w formalnym związku	Bardzo dobre	1	-14.0542	256.3	0.0030	0.9563
stan_cywi*subiektywny	Nigdy w formalnym związku	Dobre	1	-1.2411	0.9585	1.6767	0.1954
stan_cywi*subiektywny	Rozwiedziony/owdowiony	Bardzo dobre	1	-12.5421	256.3	0.0024	0.9610
stan_cywi*subiektywny	Rozwiedziony/owdowiony	Dobre	1	-0.7874	0.8967	0.7710	0.3799
liczba_os*liczba_os_	1 osoba	Z 1 osoba	1	1.5694	0.5402	8.4404	0.0037
liczba_os*liczba_os_	1 osoba	Z 2 osobami i wiecej	1	0.9007	0.4898	3.3822	0.0659
liczba_os*liczba_os_	2 osoby	Z 1 osoba	1	-0.1949	0.4739	0.1691	0.6809
liczba_os*liczba_os_	2 osoby	Z 2 osobami i wiecej	1	-0.1286	0.4505	0.0814	0.7754

Tabela 40. Analiza ocen maksymalnej wiarygodności

Oszacowania finalnego modelu odbiegają nieco od modeli zaobserwowanych w literaturze. Parametr przy zmiennej wiek jest dodatni, w przeciwieństwie do obydwu cytowanych artykułów naukowych. Subiektywny stan zdrowia również okazał się znacznie gorszym predyktorem. Wypadnięcie z modelu zmiennych dotyczących dzieci również okazało się niespodzianką. Natomiast w przypadku stanu cywilnego to osoby zamężne okazały się najmniej szczęśliwe. Przyczyn tego typu zachowań należy szukać w odpowiedzi na dane pytanie. Dominującą odpowiedzią było oznajmienie ankietowanych, że nie są w związku, lub też są w związku nieformalnym (jedna odpowiedź oznaczona kodem '66'). Ze względu na połączenie tych dwóch skrajnie różnych grup w jedną całość trudno jest uzyskać miarodajne oszacowanie wpływu stanu cywilnego na szczęście. Z całą pewnością występują też różnice kulturowe pomiędzy Bułgarią a Hiszpanią, które tłumaczą inne oszacowania parametrów. Oszacowania pozostałych parametrów nie odbiegają od oczekiwań opartych na artykułach naukowych.

Krzywa ROC (Receiver Operating Characteristic) to wykres graficzny, który ilustruje zdolność diagnostyczną wybranego modelu, w tym przypadku regresji logistycznej. Krzywa ROC jest tworzona przez stosunek współczynnika TPR w stosunku do współczynnika FPR przy różnych ustawieniach progu. Innymi słowy jest to stosunek prawdopodobieństwa wykrycia do prawdopodobieństwa fałszywego wykrycia.

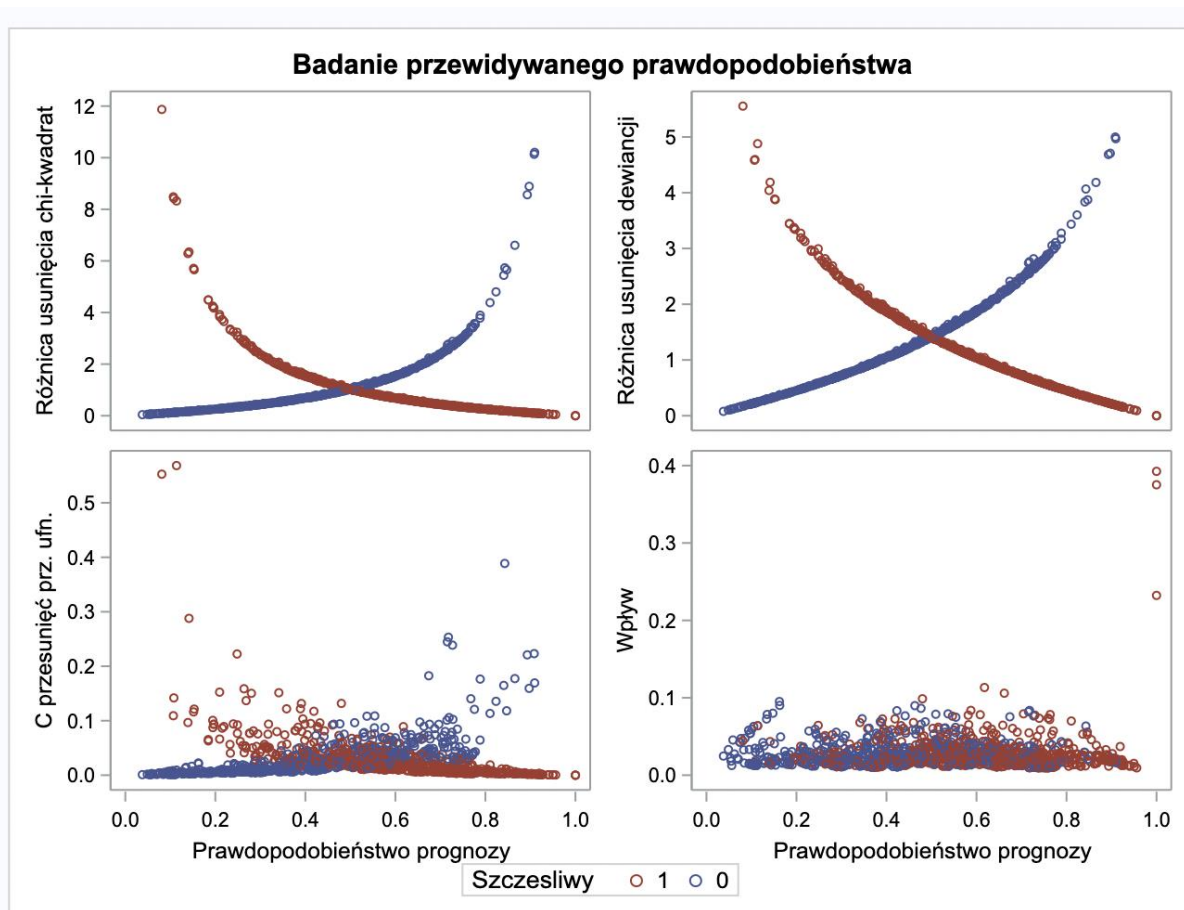
Jak widzimy z wykresu poniższej krzywej po każdym kroku następuje wzrost wartości pola pod wykresem co jest oznaką poprawy jakości modelu.



Wykres 31. Krzywa ROC -porównanie modeli

Badanie przewidywanego prawdopodobieństwa pozwala zwizualizować zależność prawdopodobieństwa prognozy w stosunku do takich parametrów jak wpływ, liczba przesunięć, różnica usunięcia chi-kwadrat czy różnica usunięcia dewiancji.





Wykres 32. Badanie przewidywanego prawdopodobieństwa

Z powyższego wykresu można zauważyć, że wartości najbardziej odstające zaliczają się do osób uznanych za szczęśliwe. Dodatkowo można zauważyć, że w przypadku przesunięcia najbardziej zróżnicowane są obserwacje ludzi szczęśliwych przy niskim prawdopodobieństwie prognozy natomiast dla obserwacji ludzi przeciwnych najbardziej zróżnicowane są obserwacje o wysokim prawdopodobieństwie prognozy.

## Podsumowanie

Na podstawie danych udało się oszacować model opisujący tłumaczący szczęścia czynnikami socjoekonomicznymi. Wyestymowany model uzyskał AUC na poziomie 0.7329 co potwierdza jego zdolności predykcyjne. Niestety wbrew oczekiwaniom dla Bułgarii w danej edycji *European Social Survey* zmienna związana z posiadaniem dzieci okazała się nieistotna statystycznie. W przypadku niektórych zmiennych oszacowania odbiegają od oczekiwanych na podstawie literatury, co można tłumaczyć odmienną kulturą w Bułgarii, lub też sposobem zbierania danych ankietowych.

## Bibliografia

United Nations. [1954]. International definition and measurement of standards and levels of living. New York: United Nations Publications

Rehdanz, K., & Maddison, D. [2005]. Climate and happiness. *Ecological Economics*, 52, 111–125.

Cunado, Gracia [2012] *Does Education Affect Happiness? Evidence for Spain*, *Social Indicators Research volume 108*, pages 185–196

Aassve, Goisis, Sironi [2012] Happiness and Childbearing Across Europe, *Social Indicators Research volume 108*, pages 65–86

## Kod

```
%macro enterpriseguide;
```

```
%global sasworklocation;
```

```
%local tempdsn unique_dsn path;
```

```
%if &sysscp=OS %then %do; /* MVS Server */
```

```
    %if %sysfunc(getoption(filesystem))=MVS %then %do;
```

```
    /* By default, physical file name will be considered a classic MVS data set. */
```

```
    /* Construct dsn that will be unique for each concurrent session under a particular account: */
```

```
        filename egtemp '&egtemp' disp=(new,delete); /* create a temporary data set */
```

```
        %let tempdsn=%sysfunc(pathname(egtemp)); /* get dsn */
```

```
        filename egtemp clear; /* get rid of data set - we only wanted its name */
```

```
        %let unique_dsn=".EGTEMP.%substr(&tempdsn, 1, 16).PDSE";
```

```
        filename egtmpdir &unique_dsn
```

```
            disp=(new,delete,delete) space=(cyl,(5,5,50))
```

```
            dsorg=po dsntype=library recfm=vb
```

```
            lrecl=8000 blksize=8004 ;
```

```
        options fileext=ignore ;
```

```
%end;
```

```
%else %do;
```

```
/*
```

```
    By default, physical file name will be considered an HFS
```

```

(hierarchical file system) file.
*/
%if "%sysfunc(getoption(filetempdir))"="" %then %do;
    filename egtmpdir '/tmp';
%end;
%else %do;
    filename egtmpdir "%sysfunc(getoption(filetempdir))";
%end;
%end;
%let path=%sysfunc(pathname(egtmpdir));
%let sasworklocation=%sysfunc(quote(&path));
%end; /* MVS Server */
%else %do;
    %let sasworklocation = "%sysfunc(getoption(work))"/";
%end;
%if &sysscp=VMS_AXP %then %do; /* Alpha VMS server */
    %let sasworklocation = "%sysfunc(getoption(work))";
%end;
%if &sysscp=CMS %then %do;
    %let path = %sysfunc(getoption(work));
    %let sasworklocation = "%substr(&path, %index(&path,%str( )))";
%end;
%mend enterpriseguide;

%enterpriseguide

/* Conditionally delete set of tables or views, if they exists */
/* If the member does not exist, then no action is performed */
%macro _eg_conditional_dropds /parmbuff;

    %local num;
    %local stepneeded;
    %local stepstarted;

```

```

%local dsname;
%local name;

%let num=1;
/* flags to determine whether a PROC SQL step is needed */
/* or even started yet */
%let stepneeded=0;
%let stepstarted=0;
%let dsname= %qscan(&syspbuff,&num,',()');
%do %while(&dsname ne);
    %let name = %sysfunc(left(&dsname));
    %if %qsysfunc(exist(&name)) %then %do;
        %let stepneeded=1;
        %if (&stepstarted eq 0) %then %do;
            proc sql;
                %let stepstarted=1;
        %end;
        drop table &name;
    %end;

    %if %sysfunc(exist(&name,view)) %then %do;
        %let stepneeded=1;
        %if (&stepstarted eq 0) %then %do;
            proc sql;
                %let stepstarted=1;
        %end;
        drop view &name;
    %end;

    %let num=%eval(&num+1);
%let dsname=%qscan(&syspbuff,&num,',()');
%end;
%if &stepstarted %then %do;
    quit;

```

```

        %end;
%mend _eg_conditional_dropds;

/* save the current settings of XPIXELS and YPIXELS */
/* so that they can be restored later          */
%macro _sas_pushchartsize(new_xsize, new_ysize);
    %global _savedxpixels _savedypixels;
    options nonotes;
    proc sql noprint;
        select setting into :_savedxpixels
        from sashelp.vgopt
        where optname eq "XPIXELS";
        select setting into :_savedypixels
        from sashelp.vgopt
        where optname eq "YPIXELS";
    quit;
    options notes;
    GOPTIONS XPIXELS=&new_xsize YPIXELS=&new_ysize;
%mend _sas_pushchartsize;

/* restore the previous values for XPIXELS and YPIXELS */
%macro _sas_popchartsize;
    %if %symexist(_savedxpixels) %then %do;
        GOPTIONS XPIXELS=&_savedxpixels YPIXELS=&_savedypixels;
        %symdel _savedxpixels / nowarn;
        %symdel _savedypixels / nowarn;
    %end;
%mend _sas_popchartsize;

%macro _SAS_VERCOMP(major, minor, maint);
    %_SAS_VERCOMP_FV(&major, &minor, &maint, &major, &minor, &maint)
%mend _SAS_VERCOMP;

```

```

%*-----*
* Tests the current version against either the required      *
* foundation or Viya required version depending on whether the *
* SYSVLONG version is a foundation or Viya one. A negative    *
* result means that the SAS server version is less than the  *
* version required. A positive result means that the SAS     *
* server version is greater than the version required. A     *
* result of zero indicates that the SAS server is exactly the *
* version required.                                         *
*                                                         *
* NOTE: The *maint parameters are optional.                *
*-----*;
%macro _SAS_VERCOMP_FV(fmajor, fminor, fmaint, vmajor, vminor, vmaint);
  %local major;
  %local minor;
  %local maint;
  %local CurMaj;
  %local CurMin;
  %local CurMnt;

  %* Pull the current version string apart.;
  %let CurMaj = %scan(&sysvlong, 1, %str(.));

  %* The Viya version number has a V on the front which means
  we need to adjust the Maint SCAN function index and also
  get the appropriate parameters for the major, minor, and
  maint values we need to check against (foundation or Viya);
  %if %eval(&CurMaj EQ V) %then
    %do;
      %* MM mm t      MM = Major version , mm = Minor version , t = Maint
version ;
      %* V.03.04M2P07112018 ;

```

```

%let major = &vmajor;
%let minor = &vminor;
%let maint = &vmaint;

%let CurMaj = %scan(&sysvlong, 2, %str(.));
%* Index is purposely 2 because V is now one of the scan delimiters ;
%let CurMin = %scan(&sysvlong, 2,
%str(.ABCDEFGHIJKLMNQRSTUWXYZ));
%let CurMnt = %scan(&sysvlong, 3,
%str(.ABCDEFGHIJKLMNQRSTUWXYZ));
%end;
%else
%do;

%* M mm t M = Major version , mm = Minor version , t = Maint
version ;

%* 9.01.02M0P11212005 ;

%let major = &fmajor;
%let minor = &fminor;
%let maint = &fmaint;

%let CurMin = %scan(&sysvlong, 2, %str(.));
%let CurMnt = %scan(&sysvlong, 4,
%str(.ABCDEFGHIJKLMNQRSTUWXYZ));
%end;

%* Now perform the version comparison.;
%if %eval(&major NE &CurMaj) %then
%eval(&CurMaj - &major);
%else
%if %eval(&minor NE &CurMin) %then
%eval(&CurMin - &minor);
%else
%if "&maint" = "" %then
%str(0);
%else

```

```

        %eval(&CurMnt - &maint);
%mend _SAS_VERCOMP_FV;

%*-----*
* This macro calls _SAS_VERCONDCODE_FV() with the passed *
* version. If the current server version matches or is newer, *
* then the true code (tcode) is executed, else the false code *
* (fcode) is executed. *
* Example: *
* %let isV92 = *
* %_SAS_VERCONDCODE(9,2,0, *
* tcode=%nrstr(Yes), *
* fcode=%nrstr(No)) *
*-----*;
%macro _SAS_VERCONDCODE( major, minor, maint, tcode=, fcode= );
    %_SAS_VERCONDCODE_FV( &major, &minor, &maint, &major, &minor, &maint,
&tcode, fcode )
%mend _SAS_VERCONDCODE;

%*-----*
* This macro calls _SAS_VERCOMP_FV() with the passed versions. *
* If the current server version matches or is newer, then the *
* true code (tcode) is executed, else the false code (fcode) *
* is executed. *
* Example: *
* %let isV92 = *
* %_SAS_VERCONDCODE_FV(9,2,0, 3,5,0 *
* tcode=%nrstr(Yes), *
* fcode=%nrstr(No)) *
*-----*;
%macro _SAS_VERCONDCODE_FV( fmajor, fminor, fmaint, vmajor, vminor, vmaint,
tcode=, fcode= );
    %if %_SAS_VERCOMP_FV(&fmajor, &fminor, &fmaint, &vmajor, &vminor, &vmaint)
>= 0 %then

```



```

        %do;
        &tcode
        %end;
    %else
        %do;
        &fcode
        %end;
%mend _SAS_VERCONDCODE_FV;

%*-----*
% Tests the current version to see if it is a Viya version *
% number. *
% A result of 1 indicates that the SAS server is a Viya *
% server. *
% A zero result indicates that the server version is not *
% that of a Viya server. *
%-----*;

%macro _SAS_ISVIYA;
    %local Major;

    %* Get the major component of the current version string.;
    %let Major = %scan(&sysvlong, 1, %str(.));

    %* Check if it is V for Viya.;
    %if %eval(&Major EQ V) %then
        %str(1);
    %else
        %str(0);
%mend _SAS_ISVIYA;

ODS PROCTITLE;
OPTIONS DEV=SVG;
GOPTIONS XPIXELS=0 YPIXELS=0;

```

```

%macro HTML5AccessibleGraphSupported;
    %if %_SAS_VERCOMP_FV(9,4,4, 0,0,0) >= 0 %then ACCESSIBLE_GRAPH;
%mend;

FILENAME EGHTMLX TEMP;
ODS HTML5(ID=EGHTMLX) FILE=EGHTMLX
    OPTIONS(BITMAP_MODE='INLINE')
    %HTML5AccessibleGraphSupported
    ENCODING='utf-8'
    STYLE=HtmlBlue
    NOGTITLE
    NOGFOOTNOTE
    GPATH=&sasworklocation
;

/* POCZATEK WEZLA: RL-Projekt.sas - Kopiuj */
%LET _CLIENTTASKLABEL='RL-Projekt.sas - Kopiuj';
%LET _CLIENTPROCESSFLOWNAME='Przebieg procesu';
%LET _CLIENTPROJECTPATH='Z:\RL_projekt\Projekt_RL.egp';
%LET _CLIENTPROJECTPATHHOST='DESKTOP-1MJKF6S';
%LET _CLIENTPROJECTNAME='Projekt_RL.egp';
%LET _SASPROGRAMFILE='';
%LET _SASPROGRAMFILEHOST='';

LIBNAME dane "Z:\RL_projekt";

data bulgaria;
    set dane.ess9bg;
    keep happy
        agea gndr marsts hinctnta nbthcld
        hhmbb estsz njbspv inprdsc health mocntr;
run;

%let varlist=agea gndr marsts hinctnta nbthcld

```

```
hhmmb estsz njbspv inprdsc health mocntr;
```

```
%macro analiza_zmiennych(var, data);
```

```
    proc freq data=&data  
        order=internal;  
        tables &var/ trend  
        plots=freqplot;  
    run;
```

```
%mend analiza_zmiennych;
```

```
/*
```

```
ods graphics on;
```

```
%analiza_zmiennych(&varlist, bulgaria)
```

```
*/
```

```
/* USUNIECIE PUSTYCH OBSERWACJI */
```

```
data bulgaria2;
```

```
set bulgaria;
```

```
if hinctnta in (77,88)
```

```
    or nbtheld in (77,88)
```

```
    or hhmmb in (77,88,99)
```

```
    or estsz in (6,7,8,9)
```

```
    or njbspv in (77777,88888,99999)
```

```
    or agea = 999
```

```
    or marsts in (77,88)
```

```
    or inprdsc in (77,88)
```

```
    or health in (7,8,9)
```

```
    or mocntr in (7,8,9)
```

```
    or happy in (77,88) then
```

```
        delete;
```

```
keep happy &varlist;
```

```
run;
```

```

/*
%analiza_zmiennych(&varlist, bulgaria2)
*/

```

## Kodowanie zmiennych

```

/* KATEGORYZACJA ZMIENNYCH */

```

```

/* ZMIENNA ZALEZNA */

```

```

data bulgaria2;
set bulgaria2;
if happy in (0,1,2,3,4,5) then happy =0;
if happy in (6,7,8,9,10) then happy=1;
run;

```

```

proc freq data=bulgaria2;
    table agea*marsts;
run;

```

```

/* ZMIENNE NIEZALEZNE */

```

```

data temp;
    set bulgaria2;

    if hhmb = 1 then
        hhmb_kat="1 osoba    ";
    if hhmb = 2 then
        hhmb_kat="2 osoby";
    if hhmb in (3,4,5,6,7,8,9,11) then
        hhmb_kat="3 i wiecej osob";

    if nbthcld = 66 then
        nbthcld_kat="Brak dzieci    ";
    if nbthcld = 1 then
        nbthcld_kat="1 dziecko";

```

```

if nbthcld = 2 then
    nbthcld_kat="2 dzieci";
if nbthcld in (3,4,5,6,7,8,9) then
    nbthcld_kat="3 i wiecej dzieci";

if health=1 then
    health_kat="Bardzo dobre    ";
if health=2 then
    health_kat="Dobre";
if health in(3,4,5) then
    health_kat="Srednie i gorsze";

if hinctnta in(1,2) then
    hinctnta_kat="1-2 decyl";
if hinctnta in(3,4) then
    hinctnta_kat="3-4 decyl";
if hinctnta in(5,6) then
    hinctnta_kat="5-6 decyl";
if hinctnta in(7,8,9,10) then
    hinctnta_kat="7-10 decyl";

if mocntr =1 then
    mocntr_kat="Tak";
if mocntr=2 then
    mocntr_kat="Nie";

if marsts=1 then
    marsts_kat="W zwiazku malzenskim    ";
if marsts in(4,5) then
    marsts_kat="Rozwiedziony/owdowiony";
if marsts in(6) then
    marsts_kat="Nigdy w formalnym zwiazku";
if marsts in(66) then
    marsts_kat="Nie w zwiazku lub w zwiazku nieformalnym";

```

```

if estsz in (1,2) then
    estsz_kat="Ponizej 25    ";
if estsz = 3 then
    estsz_kat="Miedzy 25 a 99";
if estsz in (4,5) then
    estsz_kat="Powyzej 99";

if njbspv = 66666 then
    njbspv_kat="Zwykle stanowisko    ";
if njbspv < 66666 then
    njbspv_kat="Stanowisko kierownicze";

if inprdsc = 0 then
    inprdsc_kat="Z nikim    ";
if inprdsc = 1 then
    inprdsc_kat="Z 1 osoba";
if inprdsc in(2,3,4,5,6) then
    inprdsc_kat="Z 2 osobami i wiecej";

if gndr=1 then gndr_kat='M';
if gndr=2 then gndr_kat='K';

run;

%let varlist_kat = gndr_kat marsts_kat hinctnta_kat nbtheld_kat
                    hhmmb_kat estsz_kat njbspv_kat inprdsc_kat health_kat
mocntr_kat;
/*
%analiza_zmiennych(&varlist_kat, temp);
*/

```

```

data temp2;
    set bulgaria2;
    if hhmmb = 1 then
        hhmmb=0;
    if hhmmb = 2 then
        hhmmb=1;
    if hhmmb in (3,4,5,6,7,8,9,11) then
        hhmmb=2;

    if nbthcld = 66 then
        nbthcld=0;
    if nbthcld = 1 then
        nbthcld=1;
    if nbthcld = 2 then
        nbthcld=2;
    if nbthcld in (3,4,5,6,7,8,9) then
        nbthcld=3;

    if health=1 then
        health=0;
    if health=2 then
        health=1;
    if health in(3,4,5) then
        health=2;

    if hinctnta in(1,2) then
        hinctnta=0;
    if hinctnta in(3,4) then
        hinctnta=1;
    if hinctnta in(5,6) then
        hinctnta=2;
    if hinctnta in(7,8,9,10) then
        hinctnta=3;

```

```
if mocntr =1 then
    mocntr=0;
if mocntr=2 then
    mocntr=1;
```

```
if marsts=1 then
    marsts=0;
if marsts in(4,5) then
    marsts=1;
if marsts in(6) then
    marsts=2;
if marsts in(66) then
    marsts=3;
```

```
if estsz in (1,2) then
    estsz=0;
if estsz = 3 then
    estsz=1;
if estsz in (4,5) then
    estsz=2;
```

```
if njbspv = 66666 then
    njbspv= 0;
if njbspv > 0 then
    njbspv= 1;
```

```
if inprdsc = 0 then
    inprdsc=0;
if inprdsc = 1 then
    inprdsc=1;
if inprdsc in(2,3,4,5,6) then
    inprdsc=2;
```



```

        if gndr=1 then gndr=0;
        if gndr=2 then gndr=1;
run;

/*
%analiza_zmiennych(&varlist, temp2)
*/
/* ZMIANA NAZW I DODANIE ETYKIET */

data dane.bulgaria_przed_kat(rename=(
    happy=Szczesliwy
    agea = wiek
    gndr = plec
    marsts = stan_cywilny
    hinctnta =dochod_gospodarstwa_net
    nbthcld =liczba_dzieci
    hhmmb = liczba_os_w_gosp_dom
    estsz =liczba_os_w_pracy
    njbspv =liczba_podwladnych
    inprdsc =liczba_os_do_rozmow
    health =subiektywny_stan_zdrowia
    mocntr = miejsce_urodzenia_matki

));
label
    happy= 'Jak bardzo jestes szczesliwy'
    agea = 'Wiek'
    gndr = 'Plec'
    marsts = 'Stan cywilny'
    hinctnta = 'Laczny dochod netto gospodarstwa domowego'
    nbthcld = 'Liczba dzieci'
    hhmmb = 'Liczba osob w gospodarstwie domowym'
    estsz = 'Liczba osob w miejscu pracy'

```

```

        njbspv = 'Jakiego typu stanowisko respondent obejmuje w pracy
zwykle/kierownicze'
        inprdsc = 'Liczba osob z ktorymi respondent moze porozmawiac o sprawach
osobistych'
        health = 'Subiektywny stan zdrowia'
        mocntr = 'Czy matka urodzona jest w kraju zamieszkania respondenta';
set bulgaria2;
keep happy &varlist;
run;
data dane.bulgaria_kat_opisowe (rename=(
    happy =Szczeniwy
    agea = wiek
    gndr_kat = plec
    marsts_kat = stan_cywilny
    hinctnta_kat =dochod_gospodarstwa_net
    nbthcld_kat =liczba_dzieci
    hhmmb_kat = liczba_os_w_gosp_dom
    estsz_kat =liczba_os_w_pracy
    njbspv_kat =liczba_podwladnych
    inprdsc_kat =liczba_os_do_rozmow
    health_kat =subiektywny_stan_zdrowia
    mocntr_kat = miejsce_urodzenia_matki
));
label
happy= 'Jak bardzo jestes szczeniwy'
    agea = 'Wiek'
    gndr_kat = 'Plec'
    marsts_kat = 'Stan cywilny'
    hinctnta_kat = 'Laczny dochod netto gospodarstwa domowego'
    nbthcld_kat = 'Liczba dzieci'
    hhmmb_kat = 'Liczba osob w gospodarstwie domowym'
    estsz_kat = 'Liczba osob w miejscu pracy'
    njbspv_kat = 'Jakiego typu stanowisko respondent obejmuje w pracy
zwykle/kierownicze'

```

```

        inprdsc_kat = 'Liczba osob z ktorymi respondent moze porozmawiac
o sprawach osobistych'
        health_kat = 'Subiektywny stan zdrowia'
        mocntr_kat = 'Czy matka urodzona jest w kraju zamieszkania respondenta';
set temp;
keep happy agea &varlist_kat;
run;

data dane.bulgaria_kat_num (rename=(
    happy=Szczesliwy
    agea = wiek
    gndr = plec
    marsts = stan_cywilny
    hinctnta =dochod_gospodarstwa_net
    nbtheld =liczba_dzieci
    hhmmb = liczba_os_w_gosp_dom
    estsz =liczba_os_w_pracy
    njbspv =liczba_podwladnych
    inprdsc =liczba_os_do_rozmow
    health =subiektywny_stan_zdrowia
    mocntr = miejsce_urodzenia_matki

));
label
    happy= 'Jak bardzo jestes szczesliwy'
    agea = 'Wiek'
    gndr = 'Plec'
    marsts = 'Stan cywilny'
    hinctnta = 'Laczny dochod netto gospodarstwa domowego'
    nbtheld = 'Liczba dzieci'
    hhmmb = 'Liczba osob w gospodarstwie domowym'
    estsz = 'Liczba osob w miejscu pracy'
    njbspv = 'Jakiego typu stanowisko respondent obejmuje w pracy
zwykle/kierownicze'

```

```

        inprdsc = 'Liczba osob z ktorymi respondent moze porozmawiac o sprawach
osobistych'
        health = 'Subiektywny stan zdrowia'
        mocntr = 'Czy matka urodzona jest w kraju zamieszkania respondenta';
        set temp2;
        keep happy &varlist;
run;

/*
%analiza_zmiennych(szczesliwy wiek plec stan_cywilny dochod_gospodarstwa_net
        liczba_dzieci liczba_os_w_gosp_dom liczba_os_w_pracy
liczba_podwladnych
        liczba_os_do_rozmow subiektywny_stan_zdrowia
miejsce_urodzenia_matki, dane.bulgaria_kat_opisowe)
*/

/* TABLICE KONTYNGENCJI */

%let variables = plec stan_cywilny dochod_gospodarstwa_net
        liczba_dzieci liczba_os_w_gosp_dom liczba_os_w_pracy
liczba_podwladnych
        liczba_os_do_rozmow subiektywny_stan_zdrowia
miejsce_urodzenia_matki;

%macro values_check(var,dataset);
%do i=1 %to %sysfunc(countw(&var));
%let zmienna=%sysfunc(scan(&var,&i.));
%put &=zmienna;
proc freq
data=dane.&dataset order=FORMATTED;
tables &zmienna.*szczesliwy /nopercent ;
tables &zmienna./ nocol norow nopercent nocum nofreq plots=freqplot;
title "Tablica kontyngencji";
run;

```

```

%end;
%mend;
/*przed kategoryzacja */
ods graphics on;
%values_check(&variables,bulgaria_przed_kat)

/*po kategoryzacji */
ods graphics on;
%values_check(&variables,bulgaria_kat_opisowe)

/*histogramy porównawcze*/
ods graphics on;
proc capability data=dane.bulgaria_kat_num noprint;
comphist &variables / class = ( szczeniwy )
intertile = 1.0
vscale=count
ncols = 2
nrows = 1
BARLABEL=percent
midpoints = 0 1
cfill = ligr
cframetop = blue
cframeside = blue;
inset cpk (4.2) / noframe pos = n;
run;

%macro wykresy;
%do i=1 %to %sysfunc(countw(&variables.));
%let zm=%scan(&variables.,&i.,' ');

proc sgplot data=dane.bulgaria_kat_opisowe;
vbar &zm. / stat=freq group=szczeniwy;
run;

```

```
%end;  
%mend;  
% wykresy  
% analiza_zmiennych(happy,bulgaria2)
```

## Analiza korelacji

```
/*TABLICE KORELACJI*/
```

```
/*libname dane "C:\Justyna\Zadania\Mgr\sem  
IV\Regresja_logistyczna_SAS\projekt\ESS9BG.sas";*/
```

```
/*ods rtf file= "C:\Justyna\Zadania\Mgr\sem  
IV\Regresja_logistyczna_SAS\projekt\korealcje.rtf" style=Sapphire;*/
```

```
proc freq data=dane.bulgaria_kat_num;  
tables wiek*Szczesliwy  
       plec*Szczesliwy  
       stan_cywilny*Szczesliwy  
       dochod_gospodarstwa_net*Szczesliwy  
       liczba_dzieci*Szczesliwy  
       liczba_os_w_gosp_dom*Szczesliwy  
       liczba_os_w_pracy*Szczesliwy  
       liczba_podwladnych*Szczesliwy  
       liczba_os_do_rozmow*Szczesliwy  
       subiektywny_stan_zdrowia*Szczesliwy  
       miejsce_urodzenia_matki*Szczesliwy;
```

```
run;
```

```
/*Badanie współliniowości
```

```
Macierz korelacji Pearsona*/
```

```
proc corr data=dane.bulgaria_kat_num out=bulgaria_pearson noprint pearson nosimple;  
var
```

```

        wiek
        plec
        stan_cywilny
        dochod_gospodarstwa_net
        liczba_dzieci
        liczba_os_w_gosp_dom
        liczba_os_w_pracy
        liczba_podwladnych
        liczba_os_do_rozmow
        subiektywny_stan_zdrowia
        miejsce_urodzenia_matki;

run;

/*Pearson*/
proc report
    data = bulgaria_pearson;
    title 'Macierz korelacji Pearsona';
    define _name_ / display " style=[font_weight=bold];
    define wiek    / format = 5.2;
    define plec    / format = 5.2;
    define stan_cywilny / format = 5.2;
    define dochod_gospodarstwa_net / format = 5.2;
    define liczba_dzieci / format = 5.2;
    define liczba_os_w_gosp_dom / format = 5.2;
    define liczba_os_w_pracy / format = 5.2;
    define liczba_podwladnych / format = 5.2;
    define liczba_os_do_rozmow / format = 5.2;
    define subiektywny_stan_zdrowia / format = 5.2;
    define miejsce_urodzenia_matki / format = 5.2;

run;

proc format;
value corr

```

```

0 - 0.3 = 'slaba'
0.3 - 0.6 = 'umiarkowana'
0.6-1 = 'silna'
-0.3 - 0 = 'slaba'
-0.6 - -0.3 = 'umiarkowana'
-1 - -0.6 = 'silna'
;
run;

```

```
/*Raport wynikow Pearsona po formatowaniu*/
```

```

proc report
    data = bulgaria_pearson;
    title 'Macierz korelacji Pearsona';
    define _name_ / display " style=[font_weight=bold];
    define wiek    / format = corr.;
    define plec    / format = corr.;
    define stan_cywilny / format = corr.;
    define dochod_gospodarstwa_net / format = corr.;
    define liczba_dzieci / format = corr.;
    define liczba_os_w_gosp_dom / format = corr.;
    define liczba_os_w_pracy / format = corr.;
    define liczba_podwladnych / format = corr.;
    define liczba_os_do_rozmow / format = corr.;
    define subiektywny_stan_zdrowia / format = corr.;
    define miejsce_urodzenia_matki / format = corr.;
run;

```

```

/*VIF - Variance Inflation Factor*/
proc reg data=dane.bulgaria_kat_num;
    model Szczesliwy=

```



```

        wiek
        plec
        stan_cywilny
        dochod_gospodarstwa_net
        liczba_dzieci
        liczba_os_w_gosp_dom
        liczba_os_w_pracy
        liczba_podwladnych
        liczba_os_do_rozmow
        subiektywny_stan_zdrowia
        miejsce_urodzenia_matki/
        vif tol collin;

run;

%let varlist= wiek
        plec
        stan_cywilny
        dochod_gospodarstwa_net
        liczba_dzieci
        liczba_os_w_gosp_dom
        liczba_os_w_pracy
        liczba_podwladnych
        liczba_os_do_rozmow
        subiektywny_stan_zdrowia
        miejsce_urodzenia_matki;

proc freq data=dane.bulgaria_kat_num;
tables Szczęśliwy*(&varlist)/chisq noprint;
run;

/*ods rtf close;*/

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

```

Wygenerowany dnia: sobota, 21 maja 2022 o godz. 21:49:15

Przez zadanie: RL wsteczna - redukcja

Dane wejsciowe: Local:RL\_LIB\_3.BULGARIA\_KAT\_OPISOWE

Serwer: Local

----- \*/

ODS GRAPHICS ON;

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

/\* -----

Sortowanie zbioru Local:RL\_LIB\_3.BULGARIA\_KAT\_OPISOWE

----- \*/

PROC SQL;

CREATE VIEW WORK.SORTTempTableSorted AS

SELECT T.Szczesliwy, T.wiek, T.plec, T.stan\_cywilny,

T.dochod\_gospodarstwa\_net, T.liczba\_os\_w\_gosp\_dom, T.liczba\_os\_w\_pracy,

T.liczba\_podwladnych, T.liczba\_os\_do\_rozmow, T.subiektyny\_stan\_zdrowia

FROM RL\_LIB\_3.BULGARIA\_KAT\_OPISOWE as T

;

QUIT;

TITLE;

TITLE1 "Rezultaty regresji logistycznej";

FOOTNOTE;

FOOTNOTE1 "Wygenerowane przez SAS-a (&\_SASSERVERNAME, &SYSSCPL) dnia

%TRIM(%QSYSFUNC(DATE(), NLDATE20.)) o godz. %TRIM(%QSYSFUNC(TIME(),

NLTIMAP25.))";

PROC LOGISTIC DATA=WORK.SORTTempTableSorted

PLOTS(ONLY)=ALL

;

CLASS plec (PARAM=REF) stan\_cywilny (PARAM=REF)

dochod\_gospodarstwa\_net (PARAM=REF) liczba\_os\_w\_gosp\_dom (PARAM=REF)

liczba\_os\_w\_pracy (PARAM=REF) liczba\_podwladnych (PARAM=REF)

liczba\_os\_do\_rozmow (PARAM=REF) subiektyny\_stan\_zdrowia (PARAM=REF);

```

MODEL Szczesliwy (Event = '1')=wiek plec stan_cywilny dochod_gospodarstwa_net
liczba_os_w_gosp_dom liczba_os_w_pracy liczba_podwladnych liczba_os_do_rozmow
subiektyny_stan_zdrowia wiek*plec wiek*stan_cywilny wiek*dochod_gospodarstwa_net
wiek*liczba_os_w_gosp_dom wiek*liczba_os_w_pracy wiek*liczba_podwladnych
wiek*liczba_os_do_rozmow wiek*subiektyny_stan_zdrowia plec*stan_cywilny
plec*dochod_gospodarstwa_net plec*liczba_os_w_gosp_dom plec*liczba_os_w_pracy
plec*liczba_podwladnych plec*liczba_os_do_rozmow plec*subiektyny_stan_zdrowia
stan_cywilny*dochod_gospodarstwa_net stan_cywilny*liczba_os_w_gosp_dom
stan_cywilny*liczba_os_w_pracy stan_cywilny*liczba_podwladnych
stan_cywilny*subiektyny_stan_zdrowia stan_cywilny*liczba_os_do_rozmow
dochod_gospodarstwa_net*liczba_os_w_gosp_dom
dochod_gospodarstwa_net*liczba_os_w_pracy
dochod_gospodarstwa_net*liczba_podwladnych
dochod_gospodarstwa_net*liczba_os_do_rozmow
dochod_gospodarstwa_net*subiektyny_stan_zdrowia
liczba_os_w_gosp_dom*liczba_os_w_pracy liczba_os_w_gosp_dom*liczba_podwladnych
liczba_os_w_gosp_dom*liczba_os_do_rozmow
liczba_os_w_gosp_dom*subiektyny_stan_zdrowia liczba_os_w_pracy*liczba_podwladnych
liczba_os_w_pracy*liczba_os_do_rozmow liczba_os_w_pracy*subiektyny_stan_zdrowia
liczba_podwladnych*liczba_os_do_rozmow liczba_podwladnych*subiektyny_stan_zdrowia
liczba_os_do_rozmow*subiektyny_stan_zdrowia /

```

```

SELECTION=BACKWARD

```

```

SLS=0.05

```

```

INCLUDE=9

```

```

INFLUENCE

```

```

LACKFIT

```

```

AGGREGATE SCALE=NONE

```

```

RSQUARE

```

```

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;

%LET \_CLIENTTASKLABEL=;

%LET \_CLIENTPROCESSFLOWNAME=;

%LET \_CLIENTPROJECTPATH=;

%LET \_CLIENTPROJECTPATHHOST=;

%LET \_CLIENTPROJECTNAME=;

%LET \_SASPROGRAMFILE=;

%LET \_SASPROGRAMFILEHOST=;

;\*,\*";\*/;quit;run;

ODS \_ALL\_ CLOSE;