

Analiza czynników wpływających na poziom wykształcenia.

Celem tej pracy jest potwierdzenie lub odrzucenie opinii, którą większość społeczeństwa uznaje za oczywistą, o tym że w Polsce lepiej wykształceni są mieszkańcy dużych miast. Jednocześnie będę chciał sprawdzić czy kobiety są lepiej wykształcone od mężczyzn.

1 Dane.

Dane do modelu pochodzą z badania gospodarstw domowych GUS z roku 1998, dostosowanego na WNE UW do potrzeb projektu badawczego CHER. W badaniu uczestniczyło 31,819 gospodarstw domowych, do których należało 100,909 osób. Z tak otrzymanej próby usunąłem dane o 164 gospodarstwach domowych do których należały 672 osoby, ponieważ zadeklarowali oni dochody niższe od zera.

2 Metoda badawcza.

W celu wyjaśnienia wykształcenia za pomocą charakterystyk społeczno - gospodarczych zdecydowałem się użyć jednego z modeli nieliniowych. Do prawidłowej analizy zjawiska potrzeby jest model z wielowartościową dyskretną zmienną objaśnianą. W związku z tym zdecydowałem że będzie to model probitowy lub logitowy. Ponieważ model Multinomial Logit ma lepsze własności obliczeniowe niż Multinomial Probit ¹ zdecydowałem się na zastosowanie pierwszego modelu. Poziom wykształcenia (W) jest to zmienna dyskretna która przyjmuje skończoną liczbę wartości. Dla potrzeb tego modelu ograniczyłem ją do 4 wartości (wykształcenie wyższe, średnie lub policealne, zawodowe, podstawowe). Równanie ma postać:

$$W_j = X'_{i,k} \beta_{k,j} + \epsilon_{i,j} \quad (1)$$

gdzie: i - numer obserwacji, $i = 1..N$;

j - ilość kategorii zmiennej objaśnianej, $j = 1..4$;

k - ilość zmiennych objaśniających.

Oraz:

gdzie $\epsilon_{i,j}$ są niezależne o rozkładzie o funkcji gęstości ².

Macierz X_{ij} zawiera zmienne objaśniające: wiek, płeć, miejsce zamieszkania oraz dochód na głowę w rodzinie ankietowanego.

¹J.Long "Regression Models for Categorical and Limited Dependent Variables, str.185

²J.Long "Regression Models..", str. 156

3 Opis zmiennych modelu.

Aby stworzyć model spośród zmiennych dostępnych w badaniu panelowym, wybrałem te charakterystyki społeczno-demograficzne ludności, które mają uzasadniony ekonomicznie wpływ na poziom wykształcenia badanej jednostki. Za takie cechy uznałem wiek, płeć, miejsce zamieszkania oraz poziom dochodu na głowę w rodzinie. Badając wpływ wieku na poziom wykształcenia podzieliłem całą populację na grupy wiekowe, a następnie na podgrupy osób uczących się i nie uczących się. Jednakże dla wszystkich grup osób uczących się i zarazem powyżej 30 roku życia wyestymowane współczynniki okazały się nieistotne statystycznie. Z tego powodu pominąłem je prezentując wyniki.

Badanie przeprowadziłem w dwóch etapach. W pierwszym model estymowałem na całej dostępnej próbie. Jednak wyliczone współczynniki dla ludzi młodych i uczących się były bardzo duże i zaburzały pozostałe parametry modelu. Dlatego chcąc uniknąć błędu wyjaśniania poziomu osiągniętego wykształcenia na podstawie danych o osobach, które się jeszcze uczą przed drugą estymacją z pliku danych wyeliminowałem 25739 obserwacji - ludzi uczących się. W tym przypadku ze zbioru danych zostały wyeliminowane osoby, które uczą się. Przeważająca ich większość (99%) należała do grupy wiekowej do 30 lat. W ten sposób z modelu został wyeliminowany problem, zbyt dużych współczynników dla wykształcenia podstawowego.

4 Zmienna objaśniana.

gredu -zmienna określająca poziom wykształcenia badanych osób.

1. - wykształcenie wyższe, w tym licencjackie,
2. - wykształcenie policealne, średnie ogólne lub średnie zawodowe,
3. - wykształcenie zasadnicze zawodowe,
4. - wykształcenie podstawowe, lub niepełne podstawowe. Wszystkim dzieciom w wieku do 14 lat zostało przypisane wykształcenie podstawowe.

Zmienna objaśniana ma w badanej próbie następujący rozkład:

Wykształcenie	Liczba obserwacji	Częstość
Wyższe	5435	0,0723028
Średnie	19774	0,2630571
Zawodowe	21672	0,2883065
Podstawowe	28289	0,3763336
SUMA	75170	1,0000000

Zmienne objaśniające.

Płeć - płeć osoby badanej.

0 - kobieta,

1 - mężczyzna.

Donagl - dochód gospodarstwa domowego na jednostkę ekonomiczną

Zmienne obliczone na podstawie zmiennej KLM z badania GUS.

m500 - zmienna zero-jedynkowa określająca czy badane gospodarstwo domowe mieszka w miejscowości o liczbie mieszkańców równej bądź większej od 500.000 tysięcy.

m100 - zmienna zero-jedynkowa określająca czy badane gospodarstwo domowe mieszka w miejscowości o liczbie mieszkańców zawartej w przedziale od 100.000 do 500.000.

m20 - zmienna zero-jedynkowa określająca czy badane gospodarstwo domowe mieszka w miejscowości o liczbie mieszkańców zawartej w przedziale od 20.000 do 100.000.

m0wies - zmienna zero-jedynkowa określająca czy badane gospodarstwo domowe mieszka w miejscowości o liczbie mieszkańców nie większej niż 20.000.

Zmienne obliczone na podstawie zmiennej określającej wiek osoby badanej, w latach.

Mlodz_u - zmienna zero-jedynkowa określająca czy osoba nie jest starsza niż 30 lat i jednocześnie się uczy, czyli młodzież ucząca się,

Mlodz_n - zmienna zero-jedynkowa określająca czy osoba nie jest starsza niż 30 lat i jednocześnie się nie uczy, czyli młodzież nie ucząca się,

Do40 - zmienna zero-jedynkowa określająca czy osoba ma więcej niż 30 lat, a nie więcej niż 40 lat,

Do50 - zmienna zero-jedynkowa określająca czy osoba ma więcej niż 40 lat, a nie więcej niż 50 lat,

Do60 - zmienna zero-jedynkowa określająca czy osoba ma więcej niż 50 lat, a nie więcej niż 60 lat,

Pow60 - zmienna zero-jedynkowa określająca czy osoba ma więcej niż 60 lat.

5 Model.

Pierwszym krokiem w konstrukcji modelu ekonometrycznego jest sprawdzenie dopasowania założonego modelu do danych empirycznych. W tym celu przeprowadziłem test na łączną nieistotność zmiennych. Większość testów, które zawarłem w mojej pracy bazuje na teście ilorazu wiarygodności³. Test ten polega na porównaniu wartości funkcji wiarygodności modelu pełnego z wartością funkcji wiarygodności modelu ze stałą jako jedyną zmienną objaśniającą.

Otrzymana wartość funkcji wiarygodności dla modelu ze stałą ($-95283,442$) jest znacznie mniejsza od wartości funkcji wiarygodności założonego modelu, która wynosi $L0 = -92005,495$. Na tej podstawie odrzucam hipotezę o łącznej nieistotności zmiennych objaśniających w modelu. Kolejnym krokiem było sprawdzenie istotności każdej zmiennej objaśniającej osobno. Wszystkie zmienne okazały się istotne. Wyniki estymacji stanowią załącznik do niniejszej pracy.

Iteration 0: log likelihood = -95283.442

Multinomial regression	Number of obs	=	75170
	LR chi2(0)	=	0.00
	Prob > chi2	=	.
Log likelihood = -95283.442	Pseudo R2	=	0.0000

	gredu	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
1						
	_cons	-1.649614	.0148102	-111.38	0.000	-1.678641 -1.620586
2						
	_cons	-.3581051	.0092694	-38.63	0.000	-.3762727 -.3399375
3						
	_cons	-.2664519	.0090273	-29.52	0.000	-.2841451 -.2487588

(Outcome gredu==4 is the comparison group)

³William H. Greene "Econometric Analysis", str. 826

Podstawowym pytaniem na które należy odpowiedzieć konstruując model, jest to czy wykształcenie jest cechą, która się daje sensownie uhierarchizować. To znaczy czy można podzielić je na wykluczające się kategorie, wśród których kolejna będzie oznaczała wyższy poziom od poprzedniej. Ponieważ wydaje się że, wykształcenie wyższe to coś więcej niż wykształcenie średnie, a to z kolei jest wyższe od zawodowego, to na początku wyestymowałem wielowartościowy model logitowy ze zmienną objaśnianą dającą się uhierarchizować. Ponieważ model ten okazał się słabo dopasowany do danych - jedna z wartości okazała się zdeterminowana przez zmienne objaśniające, przeprowadziłem test dopasowania modelu do danych. Statystyka testowa $-2L_1 - L_0$ ma rozkład $\chi^2(p(c - 2))$ stopniami swobody, gdzie:

L_1 - logarytm wiarygodności modelu z uhierarchizowaną zmienną objaśnianą,

L_0 - logarytm wiarygodności modelu bez hierarchii,

p - liczba zmiennych niezależnych,

c - liczba kategorii zmiennej objaśnianej.

Obliczone logarytmy wiarygodności wynoszą odpowiednio:

$$L_1 = -84985.12$$

$$L_0 = -81145.621$$

Wobec tego statystyka testowa ma wartość większą od 7000. Należy więc odrzucić hipotezę zerową, o tym że model z uhierarchizowaną zmienną objaśnianą lepiej opisuje badane zjawisko. Hierarchia wykształcenia, pomimo tego że intuicyjnie wydaje się dobrym założeniem, nie znajduje potwierdzenia w danych empirycznych. Model z hierarchią wykazuje, że jedna z wartości zmiennej objaśnianej jest zdeterminowana przez zmienne objaśniające. Oznacza to, że odchylenia standardowe, oraz jednocześnie statystyki istotności zmiennych mogą nie być prawdziwe.

```
. ologit gredu mlodz_u mlodz_n do40 do50 do60 pow60 plec donagl
m500 m100 m20 m0wies
```

```
note: mlodz_u dropped due to collinearity
note: do40 dropped due to collinearity
note: m500 dropped due to collinearity
```

```
Iteration 0: log likelihood = -95144.287
Iteration 1: log likelihood = -85211.485
Iteration 2: log likelihood = -84986.656
Iteration 3: log likelihood = -84985.121
Iteration 4: log likelihood = -84985.12
```

Ordered logit estimates	Number of obs	=	75067
	LR chi2(9)	=	20318.33
	Prob > chi2	=	0.0000
Log likelihood = -84985.12	Pseudo R2	=	0.1068

gredu	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
mlodz_n	1.274097	.0206598	61.67	0.000	1.233604	1.314589
do50	.2663757	.0204495	13.03	0.000	.2262955	.3064559
do60	.9324897	.0244593	38.12	0.000	.8845503	.980429
pow60	1.799464	.0240559	74.80	0.000	1.752316	1.846613
plec	.0548645	.0138912	3.95	0.000	.0276382	.0820907
donagl	-.00149	.0000205	-72.68	0.000	-.0015302	-.0014498
m100	.3354071	.0261653	12.82	0.000	.2841241	.3866902
m20	.5129949	.0260448	19.70	0.000	.4619479	.5640419
m0wies	1.088917	.0236576	46.03	0.000	1.042549	1.135285
-----+-----						
_cut1	-2.43268	.0334912	(Ancillary parameters)			
_cut2	-.2367079	.0307241				
_cut3	1.220811	.0311282				

```
note: 1 observation completely determined. Standard errors questionable.
```

Rezultaty estymacji modelu multinomial logit zawierają dwie części: pierwsza to współczynniki β przy kolejnych zmiennych, w drugiej części obliczone są ilorazy względnych szans (relative risk ratio). Dodatni współczynnik β wskazuje na to że dany czynnik zwiększa szansę wystąpienia wybranej kategorii zmiennej objaśnianej w stosunku od kategorii bazowej, ujemny współczynnik ją zmniejsza.

Wyniki estymacji ilorazów względnych szans otrzymane na podstawie całej próby.

```
. mlogit, rrr
Multinomial regression
```

Number of obs	=	100237
LR chi2(30)	=	51096.38
Prob > chi2	=	0.0000
Pseudo R2	=	0.2173

Log likelihood = -92005.495

gredu	RRR	Std. Err.	z	P> z	[95% Conf. Interval]	
2						
mlodz_u	19.59259	2.898202	20.11	0.000	14.66153	26.1821
mlodz_n	1.316388	.0685035	5.28	0.000	1.188744	1.457738
do50	.9823264	.0436867	-0.40	0.688	.9003276	1.071793
do60	.8014832	.0409061	-4.34	0.000	.7251883	.8858049
pow60	.8456349	.0434387	-3.26	0.001	.7646422	.9352065
plec	.6965565	.0218854	-11.51	0.000	.654956	.7407993
donagl	.9992413	.0000256	-29.59	0.000	.999191	.9992915
m100	1.18614	.0507829	3.99	0.000	1.090669	1.289968
m20	1.679887	.0779426	11.18	0.000	1.533862	1.839813
m0wies	2.212931	.0973294	18.06	0.000	2.030159	2.412156
3						
mlodz_u	4.357989	.6685751	9.60	0.000	3.226267	5.886703
mlodz_n	1.606045	.0841519	9.04	0.000	1.449297	1.779747
do50	.9735652	.0442056	-0.59	0.555	.8906672	1.064179
do60	.7472809	.0395029	-5.51	0.000	.6737324	.8288583
pow60	.5142573	.0283565	-12.06	0.000	.4615774	.5729495
plec	1.728275	.0558334	16.94	0.000	1.622236	1.841245
donagl	.9980789	.0000355	-54.13	0.000	.9980094	.9981484
m100	1.717458	.0820891	11.32	0.000	1.563872	1.886126
m20	2.500712	.1270233	18.04	0.000	2.263742	2.762488
m0wies	4.921487	.2352772	33.33	0.000	4.481297	5.404916
4						
mlodz_u	655.6744	98.06203	43.37	0.000	489.0838	879.0085
mlodz_n	10.86745	.6079842	42.64	0.000	9.738829	12.12686
do50	1.899059	.0979757	12.43	0.000	1.71642	2.101132
do60	4.518279	.2528059	26.95	0.000	4.04899	5.041959
pow60	10.32066	.5672193	42.47	0.000	9.266715	11.49448
plec	.8873437	.0284055	-3.73	0.000	.8333805	.944801
donagl	.9977258	.0000342	-66.46	0.000	.9976588	.9977928
m100	1.578642	.0734458	9.81	0.000	1.441059	1.729361
m20	2.545354	.1264653	18.80	0.000	2.309173	2.805692
m0wies	7.324298	.3408443	42.79	0.000	6.685816	8.023754

(Outcome gredu==1 is the comparison group)

Bardzo wysokie współczynniki przy zmiennej mlodz_u są zgodnie z intuicją jednak zaburzają pozostałe wielkości, ponieważ wszystkie równania są maksymalizowane równocześnie⁴.

W dalszej części pracy wszystkie otrzymane wyniki estymacji są uzyskane na podstawie próby bez ludzi uczących się.

⁴J.Long "regression Models .., str 151

```
. mlogit gredu mlodz_u mlodz_n do40 do50 do60 pow60 plec donagl m500 m100 m20 m0wies, basecat(1)
```

```
note: mlodz_u dropped due to collinearity
note: do40 dropped due to collinearity
note: m500 dropped due to collinearity
Iteration 0: log likelihood = -95144.287
Iteration 5: log likelihood = -81145.621
```

```
Multinomial regression                                Number of obs =      75067
LR chi2(27) = 27997.33
Prob > chi2 = 0.0000
Pseudo R2 = 0.1471
Log likelihood = -81145.621
```

	gredu	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
2							
	mlodz_n	.2923667	.0523184	5.59	0.000	.1898246	.3949088
	do50	-.0132477	.0448235	-0.30	0.768	-.1011002	.0746048
	do60	-.208641	.0513328	-4.06	0.000	-.3092515	-.1080304
	pow60	-.150829	.0516556	-2.92	0.004	-.2520722	-.0495858
	plec	-.355444	.0319374	-11.13	0.000	-.4180402	-.2928478
	donagl	-.0007951	.0000274	-29.02	0.000	-.0008488	-.0007414
	m100	.1798811	.0437132	4.12	0.000	.0942048	.2655575
	m20	.5326299	.0472485	11.27	0.000	.4400245	.6252353
	m0wies	.7786722	.0447159	17.41	0.000	.6910306	.8663137
	_cons	1.812797	.0541805	33.46	0.000	1.706605	1.918989
3							
	mlodz_n	.4905399	.0527933	9.29	0.000	.3870669	.594013
	do50	-.0203448	.0458136	-0.44	0.657	-.1101379	.0694483
	do60	-.2701229	.0532827	-5.07	0.000	-.374555	-.1656908
	pow60	-.6366965	.055564	-11.46	0.000	-.7455999	-.527793
	plec	.5456788	.0327234	16.68	0.000	.4815421	.6098155
	donagl	-.0020779	.0000371	-55.93	0.000	-.0021507	-.0020051
	m100	.5385324	.0485799	11.09	0.000	.4433175	.6337474
	m20	.904955	.0515634	17.55	0.000	.8038927	1.006017
	m0wies	1.572926	.0484245	32.48	0.000	1.478015	1.667836
	_cons	1.85425	.0601936	30.80	0.000	1.736272	1.972227
4							
	mlodz_n	2.412951	.0564607	42.74	0.000	2.30229	2.523612
	do50	.6601065	.0520641	12.68	0.000	.5580626	.7621503
	do60	1.561102	.0565876	27.59	0.000	1.450193	1.672012
	pow60	2.403481	.0556924	43.16	0.000	2.294326	2.512636
	plec	-.1384315	.0328943	-4.21	0.000	-.2029032	-.0739599
	donagl	-.0026293	.0000394	-66.81	0.000	-.0027064	-.0025521
	m100	.4217368	.0492354	8.57	0.000	.3252372	.5182364
	m20	.8819526	.0521299	16.92	0.000	.7797798	.9841254
	m0wies	1.953041	.0484724	40.29	0.000	1.858037	2.048045
	_cons	1.036218	.0639389	16.21	0.000	.9108996	1.161536

(Outcome gredu==1 is the comparison group)

. mlogit, rrr

Multinomial regression

Number of obs = 75067

LR chi2(27) = 27997.33

Prob > chi2 = 0.0000

Log likelihood = -81145.621

Pseudo R2 = 0.1471

	gredu	RRR	Std. Err.	z	P> z	[95% Conf. Interval]	
2							
	mlodz_n	1.339594	.0700854	5.59	0.000	1.209038	1.484249
	do50	.9868396	.0442336	-0.30	0.768	.9038425	1.077458
	do60	.8116866	.0416662	-4.06	0.000	.7339962	.8976003
	pow60	.8599947	.0444236	-2.92	0.004	.7771886	.9516235
	plec	.7008622	.0223837	-11.13	0.000	.6583357	.7461357
	donagl	.9992052	.0000274	-29.02	0.000	.9991516	.9992589
	m100	1.197075	.052328	4.12	0.000	1.098785	1.304158
	m20	1.703406	.0804834	11.27	0.000	1.552745	1.868686
	m0wies	2.178578	.0974171	17.41	0.000	1.995771	2.378128
3							
	mlodz_n	1.633198	.0862219	9.29	0.000	1.472655	1.811242
	do50	.9798607	.044891	-0.44	0.657	.8957106	1.071917
	do60	.7632857	.0406699	-5.07	0.000	.6875952	.8473082
	pow60	.5290372	.0293954	-11.46	0.000	.4744496	.5899054
	plec	1.725779	.0564734	16.68	0.000	1.618568	1.840092
	donagl	.9979242	.0000371	-55.93	0.000	.9978516	.9979969
	m100	1.71349	.0832413	11.09	0.000	1.557867	1.88466
	m20	2.471821	.1274554	17.55	0.000	2.234221	2.734688
	m0wies	4.820732	.2334417	32.48	0.000	4.384236	5.300685
4							
	mlodz_n	11.16687	.6304894	42.74	0.000	9.997052	12.47357
	do50	1.934998	.100744	12.68	0.000	1.747284	2.142879
	do60	4.76407	.2695875	27.59	0.000	4.263936	5.322868
	pow60	11.06161	.6160474	43.16	0.000	9.917747	12.33741
	plec	.8707229	.0286418	-4.21	0.000	.8163573	.9287089
	donagl	.9973742	.0000393	-66.81	0.000	.9972973	.9974511
	m100	1.524607	.0750646	8.57	0.000	1.384359	1.679064
	m20	2.415612	.1259257	16.92	0.000	2.180992	2.675471
	m0wies	7.050093	.3417349	40.29	0.000	6.411138	7.75273

(Outcome gredu==1 is the comparison group)

```
. mlogit gredu mlodz_u mlodz_n do40 do50 do60 pow60 plec donagl m500 m100 m20 m0wies, basecat(2)
```

```
note: mlodz_u dropped due to collinearity
note: do40 dropped due to collinearity
note: m500 dropped due to collinearity
Iteration 0: log likelihood = -95144.287
Iteration 1: log likelihood = -82635.267
Iteration 2: log likelihood = -81238.723
Iteration 3: log likelihood = -81146.505
Iteration 4: log likelihood = -81145.621
Iteration 5: log likelihood = -81145.621
```

```
Multinomial regression                                Number of obs   =      75067
LR chi2(27)                                           =    27997.33
Prob > chi2                                           =      0.0000
Pseudo R2                                             =      0.1471

Log likelihood = -81145.621
```

	gredu	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1							
	mlodz_n	-.2923667	.0523184	-5.59	0.000	-.3949088	-.1898246
	do50	.0132477	.0448235	0.30	0.768	-.0746048	.1011002
	do60	.208641	.0513328	4.06	0.000	.1080304	.3092515
	pow60	.150829	.0516556	2.92	0.004	.0495858	.2520722
	plec	.355444	.0319374	11.13	0.000	.2928478	.4180402
	donagl	.0007951	.0000274	29.02	0.000	.0007414	.0008488
	m100	-.1798811	.0437132	-4.12	0.000	-.2655575	-.0942048
	m20	-.5326299	.0472485	-11.27	0.000	-.6252353	-.4400245
	m0wies	-.7786722	.0447159	-17.41	0.000	-.8663137	-.6910306
	_cons	-1.812797	.0541805	-33.46	0.000	-1.918989	-1.706605
3							
	mlodz_n	.1981732	.0299235	6.62	0.000	.1395243	.2568222
	do50	-.0070971	.0281231	-0.25	0.801	-.0622173	.0480231
	do60	-.0614819	.0357682	-1.72	0.086	-.1315863	.0086224
	pow60	-.4858675	.0386192	-12.58	0.000	-.5615596	-.4101753
	plec	.9011228	.0208057	43.31	0.000	.8603444	.9419012
	donagl	-.0012828	.0000311	-41.19	0.000	-.0013439	-.0012218
	m100	.3586513	.0378482	9.48	0.000	.2844701	.4328325
	m20	.3723251	.0375779	9.91	0.000	.2986737	.4459765
	m0wies	.7942536	.0344981	23.02	0.000	.7266385	.8618687
	_cons	.0414527	.0439067	0.94	0.345	-.0446028	.1275082
4							
	mlodz_n	2.120584	.0353945	59.91	0.000	2.051213	2.189956
	do50	.6733542	.0369266	18.23	0.000	.6009793	.745729
	do60	1.769743	.0397932	44.47	0.000	1.69175	1.847737
	pow60	2.55431	.0379568	67.30	0.000	2.479916	2.628704
	plec	.2170125	.0210094	10.33	0.000	.1758348	.2581902
	donagl	-.0018342	.0000334	-54.88	0.000	-.0018997	-.0017687
	m100	.2418557	.0384377	6.29	0.000	.1665192	.3171921
	m20	.3493228	.0380523	9.18	0.000	.2747415	.423904
	m0wies	1.174369	.0342793	34.26	0.000	1.107182	1.241555
	_cons	-.7765794	.048465	-16.02	0.000	-.871569	-.6815897

(Outcome gredu==2 is the comparison group)

. mlogit, rrr

Multinomial regression

Number of obs = 75067

LR chi2(27) = 27997.33

Prob > chi2 = 0.0000

Pseudo R2 = 0.1471

Log likelihood = -81145.621

	gredu	RRR	Std. Err.	z	P> z	[95% Conf. Interval]	
1							
	mlodz_n	.7464947	.0390554	-5.59	0.000	.6737415	.8271042
	do50	1.013336	.0454213	0.30	0.768	.9281102	1.106387
	do60	1.232003	.0632422	4.06	0.000	1.114082	1.362405
	pow60	1.162798	.0600651	2.92	0.004	1.050836	1.286689
	plec	1.426814	.0455688	11.13	0.000	1.340239	1.518982
	donagl	1.000795	.0000274	29.02	0.000	1.000742	1.000849
	m100	.8353695	.0365167	-4.12	0.000	.7667783	.9100964
	m20	.587059	.0277377	-11.27	0.000	.5351355	.6440206
	m0wies	.4590151	.0205253	-17.41	0.000	.4204988	.5010594
3							
	mlodz_n	1.219174	.0364819	6.62	0.000	1.149727	1.292815
	do50	.992928	.0279242	-0.25	0.801	.9396787	1.049195
	do60	.9403699	.0336353	-1.72	0.086	.8767036	1.00866
	pow60	.6151633	.0237571	-12.58	0.000	.5703189	.6635339
	plec	2.462366	.0512312	43.31	0.000	2.363975	2.564853
	donagl	.998718	.0000311	-41.19	0.000	.998657	.998779
	m100	1.431398	.0541759	9.48	0.000	1.329058	1.541618
	m20	1.451105	.0545295	9.91	0.000	1.34807	1.562015
	m0wies	2.212789	.0763371	23.02	0.000	2.068117	2.367581
4							
	mlodz_n	8.336008	.2950488	59.91	0.000	7.777326	8.934824
	do50	1.960803	.0724058	18.23	0.000	1.823904	2.107978
	do60	5.869347	.2335598	44.47	0.000	5.428974	6.345441
	pow60	12.86242	.4882162	67.30	0.000	11.94026	13.8558
	plec	1.24236	.0261013	10.33	0.000	1.192241	1.294585
	donagl	.9981675	.0000334	-54.88	0.000	.9981021	.9982329
	m100	1.27361	.0489546	6.29	0.000	1.181186	1.373266
	m20	1.418107	.0539623	9.18	0.000	1.31619	1.527915
	m0wies	3.236099	.1109313	34.26	0.000	3.025821	3.460991

(Outcome gredu==2 is the comparison group)

```
. mlogit gredu mlodz_u mlodz_n do40 do50 do60 pow60 plec donagl m500 m100 m20 m0wies, basecat(3)
```

```
note: mlodz_u dropped due to collinearity
note: do40 dropped due to collinearity
note: m500 dropped due to collinearity
Iteration 0: log likelihood = -95144.287
Iteration 1: log likelihood = -82635.267
Iteration 2: log likelihood = -81238.723
Iteration 3: log likelihood = -81146.505
Iteration 4: log likelihood = -81145.621
Iteration 5: log likelihood = -81145.621
```

```
Multinomial regression                                Number of obs   =      75067
LR chi2(27)                                           =    27997.33
Prob > chi2                                           =      0.0000
Pseudo R2                                             =      0.1471
Log likelihood = -81145.621
```

	gredu	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1							
	mlodz_n	-.4905399	.0527933	-9.29	0.000	-.594013	-.3870669
	do50	.0203448	.0458136	0.44	0.657	-.0694483	.1101379
	do60	.2701229	.0532827	5.07	0.000	.1656908	.374555
	pow60	.6366965	.055564	11.46	0.000	.527793	.7455999
	plec	-.5456788	.0327234	-16.68	0.000	-.6098155	-.4815421
	donagl	.0020779	.0000371	55.93	0.000	.0020051	.0021507
	m100	-.5385324	.0485799	-11.09	0.000	-.6337474	-.4433175
	m20	-.904955	.0515634	-17.55	0.000	-1.006017	-.8038927
	m0wies	-1.572926	.0484245	-32.48	0.000	-1.667836	-1.478015
	_cons	-1.85425	.0601936	-30.80	0.000	-1.972227	-1.736272
2							
	mlodz_n	-.1981732	.0299235	-6.62	0.000	-.2568222	-.1395243
	do50	.0070971	.0281231	0.25	0.801	-.0480231	.0622173
	do60	.0614819	.0357682	1.72	0.086	-.0086224	.1315863
	pow60	.4858675	.0386192	12.58	0.000	.4101753	.5615596
	plec	-.9011228	.0208057	-43.31	0.000	-.9419012	-.8603444
	donagl	.0012828	.0000311	41.19	0.000	.0012218	.0013439
	m100	-.3586513	.0378482	-9.48	0.000	-.4328325	-.2844701
	m20	-.3723251	.0375779	-9.91	0.000	-.4459765	-.2986737
	m0wies	-.7942536	.0344981	-23.02	0.000	-.8618687	-.7266385
	_cons	-.0414527	.0439067	-0.94	0.345	-.1275082	.0446028
4							
	mlodz_n	1.922411	.0326307	58.91	0.000	1.858456	1.986366
	do50	.6804513	.0352682	19.29	0.000	.6113268	.7495758
	do60	1.831225	.0383991	47.69	0.000	1.755965	1.906486
	pow60	3.040177	.0397598	76.46	0.000	2.96225	3.118105
	plec	-.6841104	.019863	-34.44	0.000	-.7230411	-.6451796
	donagl	-.0005513	.0000341	-16.16	0.000	-.0006182	-.0004845
	m100	-.1167956	.0423231	-2.76	0.006	-.1997474	-.0338439
	m20	-.0230024	.0417128	-0.55	0.581	-.104758	.0587533
	m0wies	.3801151	.0375253	10.13	0.000	.3065669	.4536634
	_cons	-.8180321	.0505625	-16.18	0.000	-.9171327	-.7189315

(Outcome gredu==3 is the comparison group)

. mlogit, rrr

Multinomial regression

Number of obs = 75067

LR chi2(27) = 27997.33

Prob > chi2 = 0.0000

Pseudo R2 = 0.1471

Log likelihood = -81145.621

	gredu	RRR	Std. Err.	z	P> z	[95% Conf. Interval]	
1							
	mlodz_n	.6122957	.0323251	-9.29	0.000	.5521072	.6790456
	do50	1.020553	.0467553	0.44	0.657	.9329084	1.116432
	do60	1.310125	.069807	5.07	0.000	1.180208	1.454344
	pow60	1.890226	.1050285	11.46	0.000	1.695187	2.107706
	plec	.5794483	.0189615	-16.68	0.000	.5434511	.6178299
	donagl	1.00208	.0000372	55.93	0.000	1.002007	1.002153
	m100	.5836041	.0283515	-11.09	0.000	.5305997	.6419034
	m20	.4045601	.0208605	-17.55	0.000	.3656724	.4475833
	m0wies	.2074374	.0100451	-32.48	0.000	.1886549	.2280899
2							
	mlodz_n	.8202277	.0245441	-6.62	0.000	.7735058	.8697719
	do50	1.007122	.0283234	0.25	0.801	.9531117	1.064194
	do60	1.063411	.0380363	1.72	0.086	.9914147	1.140636
	pow60	1.625585	.0627787	12.58	0.000	1.507082	1.753405
	plec	.4061134	.0084495	-43.31	0.000	.3898859	.4230164
	donagl	1.001284	.0000312	41.19	0.000	1.001223	1.001345
	m100	.6986179	.0264415	-9.48	0.000	.6486691	.7524129
	m20	.6891302	.0258961	-9.91	0.000	.6401988	.7418014
	m0wies	.4519184	.0155903	-23.02	0.000	.4223721	.4835317
4							
	mlodz_n	6.837425	.2231097	58.91	0.000	6.413828	7.288999
	do50	1.974769	.0696466	19.29	0.000	1.842875	2.116102
	do60	6.24153	.2396689	47.69	0.000	5.789029	6.729401
	pow60	20.90895	.8313353	76.46	0.000	19.34143	22.60351
	plec	.5045389	.0100217	-34.44	0.000	.4852742	.5245683
	donagl	.9994488	.0000341	-16.16	0.000	.999382	.9995157
	m100	.889767	.0376577	-2.76	0.006	.8189376	.9667224
	m20	.9772602	.0407643	-0.55	0.581	.9005424	1.060514
	m0wies	1.462453	.054879	10.13	0.000	1.358752	1.574068

(Outcome gredu==3 is the comparison group)

```
. mlogit gredu mlodz_u mlodz_n do40 do50 do60 pow60 plec donagl m500 m100 m20 m0wies, basecat(4)
```

```
note: mlodz_u dropped due to collinearity
note: do40 dropped due to collinearity
note: m500 dropped due to collinearity
Iteration 0: log likelihood = -95144.287
Iteration 1: log likelihood = -82635.267
Iteration 2: log likelihood = -81238.723
Iteration 3: log likelihood = -81146.505
Iteration 4: log likelihood = -81145.621
Iteration 5: log likelihood = -81145.621
```

```
Multinomial regression                                Number of obs   =      75067
LR chi2(27)                                           =    27997.33
Prob > chi2                                           =      0.0000
Pseudo R2                                             =      0.1471

Log likelihood = -81145.621
```

	gredu	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
1						
	mlodz_n	-2.412951	.0564607	-42.74	0.000	-2.523612 -2.30229
	do50	-.6601065	.0520641	-12.68	0.000	-.7621503 -.5580626
	do60	-1.561102	.0565876	-27.59	0.000	-1.672012 -1.450193
	pow60	-2.403481	.0556924	-43.16	0.000	-2.512636 -2.294326
	plec	.1384315	.0328943	4.21	0.000	.0739599 .2029032
	donagl	.0026293	.0000394	66.81	0.000	.0025521 .0027064
	m100	-.4217368	.0492354	-8.57	0.000	-.5182364 -.3252372
	m20	-.8819526	.0521299	-16.92	0.000	-.9841254 -.7797798
	m0wies	-1.953041	.0484724	-40.29	0.000	-2.048045 -1.858037
	_cons	-1.036218	.0639389	-16.21	0.000	-1.161536 -.9108996
2						
	mlodz_n	-2.120584	.0353945	-59.91	0.000	-2.189956 -2.051213
	do50	-.6733542	.0369266	-18.23	0.000	-.745729 -.6009793
	do60	-1.769743	.0397932	-44.47	0.000	-1.847737 -1.69175
	pow60	-2.55431	.0379568	-67.30	0.000	-2.628704 -2.479916
	plec	-.2170125	.0210094	-10.33	0.000	-.2581902 -.1758348
	donagl	.0018342	.0000334	54.88	0.000	.0017687 .0018997
	m100	-.2418557	.0384377	-6.29	0.000	-.3171921 -.1665192
	m20	-.3493228	.0380523	-9.18	0.000	-.423904 -.2747415
	m0wies	-1.174369	.0342793	-34.26	0.000	-1.241555 -1.107182
	_cons	.7765794	.048465	16.02	0.000	.6815897 .871569
3						
	mlodz_n	-1.922411	.0326307	-58.91	0.000	-1.986366 -1.858456
	do50	-.6804513	.0352682	-19.29	0.000	-.7495758 -.6113268
	do60	-1.831225	.0383991	-47.69	0.000	-1.906486 -1.755965
	pow60	-3.040177	.0397598	-76.46	0.000	-3.118105 -2.96225
	plec	.6841104	.019863	34.44	0.000	.6451796 .7230411
	donagl	.0005513	.0000341	16.16	0.000	.0004845 .0006182
	m100	.1167956	.0423231	2.76	0.006	.0338439 .1997474
	m20	.0230024	.0417128	0.55	0.581	-.0587533 .104758
	m0wies	-.3801151	.0375253	-10.13	0.000	-.4536634 -.3065669
	_cons	.8180321	.0505625	16.18	0.000	.7189315 .9171327

(Outcome gredu==4 is the comparison group)

```
. mlogit, rrr
```

```

Multinomial regression              Number of obs   =      75067
                                   LR chi2(27)        =    27997.33
                                   Prob > chi2         =      0.0000
Log likelihood = -81145.621          Pseudo R2       =      0.1471

```

	gredu	RRR	Std. Err.	z	P> z	[95% Conf. Interval]	
1							
	mlodz_n	.0895506	.0050561	-42.74	0.000	.0801695	.1000295
	do50	.5167963	.0269066	-12.68	0.000	.4666619	.5723168
	do60	.2099045	.011878	-27.59	0.000	.1878687	.2345251
	pow60	.0904027	.0050347	-43.16	0.000	.0810543	.1008294
	plec	1.148471	.0377781	4.21	0.000	1.076764	1.224954
	donagl	1.002633	.0000395	66.81	0.000	1.002555	1.00271
	m100	.6559067	.0322938	-8.57	0.000	.59557	.722356
	m20	.4139738	.0215804	-16.92	0.000	.373766	.4585069
	m0wies	.1418421	.0068754	-40.29	0.000	.1289868	.1559786
2							
	mlodz_n	.1199615	.004246	-59.91	0.000	.1119216	.1285789
	do50	.5099951	.0188324	-18.23	0.000	.4743883	.5482744
	do60	.1703767	.0067798	-44.47	0.000	.1575935	.1841969
	pow60	.0777459	.002951	-67.30	0.000	.072172	.0837503
	plec	.8049199	.0169109	-10.33	0.000	.7724483	.8387566
	donagl	1.001836	.0000335	54.88	0.000	1.00177	1.001901
	m100	.7851695	.0301801	-6.29	0.000	.7281909	.8466065
	m20	.7051655	.0268332	-9.18	0.000	.6544867	.7597685
	m0wies	.309014	.0105928	-34.26	0.000	.2889346	.3304888
3							
	mlodz_n	.1462539	.0047724	-58.91	0.000	.1371931	.1559131
	do50	.5063884	.0178594	-19.29	0.000	.472567	.5426304
	do60	.1602171	.0061522	-47.69	0.000	.1486016	.1727405
	pow60	.0478264	.0019016	-76.46	0.000	.0442409	.0517025
	plec	1.982008	.0393686	34.44	0.000	1.906329	2.060691
	donagl	1.000551	.0000341	16.16	0.000	1.000485	1.000618
	m100	1.12389	.0475665	2.76	0.006	1.034423	1.221094
	m20	1.023269	.0426835	0.55	0.581	.9429394	1.110442
	m0wies	.6837827	.0256592	-10.13	0.000	.6352965	.7359693

(Outcome gredu==4 is the comparison group)

Otrzymane za pomocą tego modelu wyniki nieznacznie różnią się od wyników estymacji na pełnym zbiorze danych (str. 6). Różnice w otrzymanych współczynnikach nie są większe niż 0.02.

Rezultaty estymacji interpretuję według metody opisanej przez J.S.Longa. Wskazują one na następujące tendencje w społeczeństwie:

1. Ze wszystkich porównań widać, że wraz ze wzrostem dochodu na członka gospodarstwa domowego, przy wszystkich pozostałych czynnikach stałych, rośnie poziom uzyskiwanego wykształcenia, chociaż zróżnicowanie dochodowe nie jest silne. Współczynniki przy ilorazach względnych szans są bliskie jedności i różnią się trzecim miejscem po przecinku (str. 9,11,13 i 15 zmienna donagl),

2. W grupie wiekowej 40-50 lat nie jest istotne rozróżnienie pomiędzy wykształceniem zawodowym, średnim i wyższym. (współczynniki przy ilorazach względnych szans zmiennej do50 są nieistotne, str. 9,11,13). Statystyczny przedstawiciel tej grupy ma 2 razy częściej wykształcenie podstawowe niż wyższe, dwa razy częściej podstawowe niż średnie, oraz dwa razy częściej podstawowe niż zawodowe. A zarazem mają oni częściej wykształcenie ponadpodstawowe niż podstawowe.
3. Szansa, że przedstawiciel grupy powyżej 60 lat ma wykształcenie podstawowe jest ponad 2 razy większa od szansy ze ma jakiegokolwiek inne wykształcenie. Świadczą o tym współczynniki relatywnych szans przy zmiennej pow60 mniejsze od jedności.
4. Młodzi ludzie, którzy przestali się uczyć mają w zdecydowanej większości wykształcenie podstawowe.
5. Fakt, że badany jest mężczyzną zmniejsza szansą osiągnięcia wykształcenia średniego w stosunku do wyższego (str. 9, kategoria średnie, płeć=0,7) oraz zwiększa 1,7 razy szansę osiągnięcia wykształcenia zawodowego w stosunku do wyższego (str. 9, kategoria zawodowe, zmienna płeć). W stosunku do wykształcenia podstawowego te współczynniki wynoszą odpowiednio 0.8 i 1.98. (str. 15, kategoria średnie i zawodowe). Wynika z tego, że mężczyźni albo legitymują się wykształceniem zawodowym, albo kończą studia. Dla kobiet te współczynniki należy interpretować odwrotnie. Kobiety częściej niż mężczyźni osiągają wykształcenie średnie. Płeć ma małe znaczenie przy wyjaśnieniu czy badany należy do grupy osób o najniższym wykształceniu.
6. Ludzie mieszkający na wsi lub w małych miastach mają ponad 2 razy większą szansę na posiadanie wykształcenia podstawowego niż wyższego (str. 9, kategoria podstawowe, zmienna m0wies), oraz 1,5 razy większą szansę na legitymowanie się wykształceniem zawodowym niż wyższym lub średnim w stosunku do wyższego (str. 12, kategoria wyższe i średnie, zmienna m0wies),
7. Najczęściej ludzi z wyższym wykształceniem spotyka się w dużych miastach - str. 9, wszystkie kategorie, im niższe wykształcenie w stosunku do wyższego i im mniejsze miasto tym wyższy współczynnik szans.

Otrzymane wyniki potwierdzają hipotezę o tym, że w dużych miastach mieszkają ludzie o wyższym od przeciętnego wykształceniu, natomiast na wsi

i w małych miasteczkach ludzie są słabiej wykształceni. Poziom wykształcenia rośnie z wiekiem. Im ktoś jest starszy tym większa jest szansa że ma wyższe lub średnie wykształcenie. Małe odstępstwo od tej reguły obserwowane jest w grupie osób najstarszych. Odstępstwo to może mieć dwie przyczyny: po pierwsze jest to grupa ludzi, którzy urodzili się przed wojną i ich lata szkolne przypadają na okres II wojny światowej, a drugą przyczyną może być to że grupa ta nie musi być reprezentatywna - więcej osób lepiej wykształconych niż słabo wykształconych może nie dożywać 60 lat. Na drugie postawione pytanie nie jestem w stanie udzielić jednoznacznej odpowiedzi. Na pewno przeciętna kobieta jest lepiej wykształcona niż przeciętny mężczyzna, i na pewno spośród osób , które zdają maturę więcej mężczyzn niż kobiet osiąga wyższe wykształcenie. Dlatego też nie da się sformułować jednoznacznego stwierdzenia oddającego zależność między płcią a wykształceniem.

Literatura

- [1] William H. Greene *Econometric Analysis*, 4th edition, 1999.
- [2] Scott J, Long *Regression Models for Categorical and Limited Dependent Variables*, Thousand Oak, 1997.

Załącznik

Test na istotność kolejnych zmiennych objaśniających

```
. test mlodz_n
```

```
( 1)  [1]mlodz_n = 0.0  
( 2)  [2]mlodz_n = 0.0  
( 3)  [3]mlodz_n = 0.0
```

```
      chi2( 3) = 4565.97  
Prob > chi2 =    0.0000
```

```
. test do50
```

```
( 1)  [1]do50 = 0.0  
( 2)  [2]do50 = 0.0  
( 3)  [3]do50 = 0.0
```

```
      chi2( 3) = 418.88  
Prob > chi2 =    0.0000
```

```
. test do60
```

```
( 1)  [1]do60 = 0.0  
( 2)  [2]do60 = 0.0  
( 3)  [3]do60 = 0.0
```

```
      chi2( 3) = 2700.28  
Prob > chi2 =    0.0000
```

```
. test pow60
```

```
( 1)  [1]pow60 = 0.0  
( 2)  [2]pow60 = 0.0  
( 3)  [3]pow60 = 0.0
```

```
      chi2( 3) = 6982.70  
Prob > chi2 =    0.0000
```

```

. test plec

( 1)  [1]plec = 0.0
( 2)  [2]plec = 0.0
( 3)  [3]plec = 0.0

      chi2( 3) = 2135.56
Prob > chi2 =    0.0000

. test donagl

( 1)  [1]donagl = 0.0
( 2)  [2]donagl = 0.0
( 3)  [3]donagl = 0.0

      chi2( 3) = 5145.56
Prob > chi2 =    0.0000

. test m100

( 1)  [1]m100 = 0.0
( 2)  [2]m100 = 0.0
( 3)  [3]m100 = 0.0

      chi2( 3) = 161.64
Prob > chi2 =    0.0000

. test m20

( 1)  [1]m20 = 0.0
( 2)  [2]m20 = 0.0
( 3)  [3]m20 = 0.0

      chi2( 3) = 383.92
Prob > chi2 =    0.0000

. test m0wies

```

```
( 1) [1]m0wies = 0.0
( 2) [2]m0wies = 0.0
( 3) [3]m0wies = 0.0
```

```
      chi2( 3) = 2177.64
Prob > chi2 =    0.0000
```

Test na łączną istotność miejsca zamieszkania.

```
. test m100 m20 m0wies
```

```
( 1) [1]m100 = 0.0
( 2) [2]m100 = 0.0
( 3) [3]m100 = 0.0
( 4) [1]m20 = 0.0
( 5) [2]m20 = 0.0
( 6) [3]m20 = 0.0
( 7) [1]m0wies = 0.0
( 8) [2]m0wies = 0.0
( 9) [3]m0wies = 0.0
```

```
      chi2( 9) = 3205.71
Prob > chi2 =    0.0000
```

Test na łączną istotność wieku.

```
. test mlodz_n do50 do60 pow60
```

```
( 1) [1]mlodz_n = 0.0
( 2) [2]mlodz_n = 0.0
( 3) [3]mlodz_n = 0.0
( 4) [1]do50 = 0.0
( 5) [2]do50 = 0.0
( 6) [3]do50 = 0.0
( 7) [1]do60 = 0.0
( 8) [2]do60 = 0.0
( 9) [3]do60 = 0.0
(10) [1]pow60 = 0.0
(11) [2]pow60 = 0.0
(12) [3]pow60 = 0.0
```

```
      chi2(12) =10272.23
Prob > chi2 =    0.0000
```