

1. Wprowadzenie

Tematyka zdrowia jest niezwykle istotna w XXI wieku dla każdego państwa. Autorzy raportu wydanego przez ONZ podkreślają, że w latach 2000-2019 oczekiwana liczba lat życia wzrosła z 67 do 73¹. Jednym z czynników ryzykownych dla zdrowia jest alkohol. Globalnie liczba spożytego alkoholu na mieszkańca od 2015 spada². Warto zaznaczyć, że mężczyźni spożywali średnio 4 razy więcej alkoholu niż kobiety. Kolejnym ryzykownym czynnikiem jest palenie papierosów. Globalnie procent populacji czynnie palących w latach 2000-2020 zmalał z poziomu 32,7% do 22,3%³. ONZ podkreśla, że m. in. dzięki poprawie czynników ryzyka wymienionych wcześniej polepszyło się zdrowie ludzi i wzrosła oczekiwana liczba lat życia.

Aktualność problemu także z teoretycznego punktu widzenia potwierdza ilość zapytań frazy „zdrowie” w Google Scholar w dniu 21.04.2024 r., która wyniosła 165 000 wyników, natomiast w języku angielskim „health” osiągnęła 8 100 000.

Mateusz Kolenda w swojej publikacji podjął tematykę analizy zdrowia poprzez wybrane determinanty na przykładzie krajów Unii Europejskiej⁴. W pracy wykorzystano model klasycznej metody najmniejszych kwadratów. Wybrane zostały zmienne z liczbą lekarzy, finansami oraz powiązane ze stylem życia takie jak palenie papierosów oraz picie alkoholu. W wynikach autor otrzymał, że wzrost odsetek osób palących oraz spożywających alkohol ma istotny wpływ na pogorszenie się stanu zdrowia i były to zmienne istotne statystycznie. Wzrost zmiennych o 1% powodował odpowiednio pogorszenie oceny stanu zdrowia o około 0,69% oraz 0,85%. W pracy sprawdzono testy takie jak normalność rozkładu reszt, heteroskedantyczność, nieliniowość. Model spełnił wszystkie założenia i uznano go za poprawny. Warto podkreślić jest to, że wyniki otrzymane przez autora prowadzą do konkluzji, że styl życia ma większy wpływ na ocenę zdrowia niż zasoby państw. Rezultaty sugerują, że ważne jest skupienie się na zapobieganiu chorob, a nie tylko przekazywanie środków na ich leczenie.

¹ ONZ, *World health statistics 2023*

<https://www.who.int/publications/i/item/9789240074323> (dostęp: 28.04.2024).

² Ibidem, s. 12.

³ Ibidem, s. 13.

⁴ Mateusz Kolenda, *Analiza ekonometryczna wybranych determinant zdrowia na przykładzie krajów UE*, Uniwersytet Ekonomiczny, Wrocław 2022.

Tetyana Demydas badała wpływ czynników stylu życia na zdrowie na przykładzie mikrodanych ze Stanów Zjednoczonych zebranych w latach 2005-2006⁵. Ważnym punktem jest to, że zmienna objaśniana zdrowia nie jest obiektywną oceną respondenta, lecz pomiarem stanu zdrowia przez wykwalifikowanych pracowników⁶. Autorka wykorzystowała metodę modelowania równań strukturalnych. Model podobnie jak w poprzednim artykule wykorzystuje zmienne dotyczące palenia papierosów oraz spożywania alkoholu. Dodatkowo w tej pracy dodano inne zmienne m.in. dotyczące diety, wagi, edukacji oraz aktywność fizyczną, która jest mierzona nie wprost poprzez liczbę godzin spędzonych na oglądaniu telewizji w ciągu miesiąca. W wynikach otrzymano, że waga, palenie papierosów, picie alkoholu oraz dieta są istotne. Są one silnie skorelowane ze stanem zdrowia i mają na niego negatywny wpływ⁷. Jednak największy wpływ na zdrowie w badanym modelu ekonometrycznym miała waga badanej osoby. Nie znaleziono korelacji między stanem zdrowia, a dochodami. Autorka zaznaczyła, że na zdrowie ma wpływ wiele zmiennych i nie da się wyjaśnić jego stanu za pomocą pojedynczej⁸.

Paul Contoyannis and Andrew M Jones podjęli, także próbę wyjaśnienia oceny stanu zdrowia z pomocą wybranych zmiennych stylu życia⁹. Ich model korzystał z danych panelowych z lat 1984 i 1991 z Wielkiej Brytanii. Wykorzystano metodę największej wiarygodności. Brak palenia papierosów i umiarkowane picie alkoholu w 1984 roku miało istotne znaczenie na dobry oraz bardzo dobry stan zdrowia w 1991 roku. Ponownie styl życia okazał się kluczowym czynnikiem wpływającym na ocenę zdrowia.

Przytoczone artykuły potwierdzają istotę tematu jakim jest ocena stanu zdrowia. W pracy zostaną wykorzystane dane pochodzące z ankiet z prawie 20 różnych krajów z Europy oraz Azji. W pracy skonstruowano model uporządkowanej regresji logistycznej, a zmienną objaśnianą jest ocena stanu zdrowia przez respondentów, natomiast zmienne objaśniające dotyczą stylu życia tej osoby. Sprawdzono hipotezę badawczą według której styl życia w postaci palenia papierosów, picia alkoholu, aktywności fizycznej, ma istotny wpływ na stan zdrowia.

⁵ Tetyana Demydas, *Lifestyle factors, dietary quality and health: Econometrics evidence from US micro data*, Review of Agricultural and Environmental Studies 94-2 (2013).

⁶ Ibidem, s. 144.

⁷ Tetyana Demydas, *Lifestyle factors...*, op. cit., s. 154.

⁸ Ibidem, s. 157.

⁹ Paul Contoyannis and Andrew M Jones, *Socio-Economic Status, Health and Lifestyle*, Department of Economics and Related Studies, University of York, 2001/19.

2. Analiza danych

Wykorzystane dane pochodzą z International Social Survey Programme i były zbierane w latach 2021-2023¹⁰. Do ankiet mogły przystąpić tylko osoby powyżej 18 roku życia. Tak jak wspomniano wcześniej dane pochodzą z 19 krajów z Azji, Europy oraz Oceanii. Dokładne kraje zawarte w danych to: Szwajcaria, Czechy, Niemcy, Dania, Finlandia, Chorwacja, Izrael, Islandia, Włochy, Japonia, Norwegia, Nowa Zelandia, Filipiny, Polska, Rosja, Słowacja, Słowenia, Tajlandia oraz Tajwan. Zbiór danych zawiera 27 276 obserwacji oraz 279 kolumn. Jednak do analizy na potrzeby niniejszej pracy wykorzystaną zostaną wykorzystane tylko 8 zmiennych, które mogą mieć potencjalny wpływ na stan zdrowia. W danych wystąpiło prawie 12 000 brakujących wartości, zatem opuszczono je i do analizy zostały wykorzystane 15 291 obserwacje. Poniższa tabela 1. zawiera nazwę zmiennych oraz jej definicje:

Tabela 1. Nazwa i opis zmiennych w zbiorze danych

Nazwa	Wyjaśnienie
V47	Palenie papierosów
V48	Częstotliwość picia alkoholu 4 drinków lub więcej
V49	Częstotliwość aktywności fizycznej trwającej minimum 20 minut
V50	Częstotliwość jedzenie owoców lub warzyw
V51	Stan zdrowia
SEX	Płeć
AGE	Wiek w latach
WRKHRS	Liczba godzin przepracowanych w tygodniu

Źródło: Opracowanie własne na podstawie danych ISSP.

Zmienna V47 przyjmuję następujące uporządkowanie wartości:

- 1, gdy osoba nie pali i nigdy nie paliła,
- 2, gdy osoba nie pali, ale kiedyś paliła,
- 3, gdy osoba pali od 1 do 5 papierosów dziennie,
- 4, gdy osoba pali od 6 do 10 papierosów dziennie,
- 5, gdy osoba pali od 11 do 20 papierosów dziennie,

¹⁰ ISSP, *Health and Health Care II – ISSP 2021*, https://search.gesis.org/research_data/ZA8000 (dostęp: 28.04.2024).

- 6, gdy osoba pali od 21 do 40 papierosów dziennie,
- 7, gdy osoba pali ponad 40 papierosów dziennie.

AGE jest zmienną odpowiadającą za wiek respondenta wyrażoną w latach. WRKHR jest wyrażona w liczbie godzin przepracowanych w tygodniu. W celu uproszczenia pogrupowano WRKHR w pięć grup w odpowiednie przedziały oznaczające liczbę przepracowanych godzin w tygodniu, które przyjmują wartość od 1 do 5. Podobnie jak v47 zmienne v48, v49, v50, v51 przyjmują uporządkowane wartości. Poniższa tabela 2. przedstawia wyjaśnienia dla konkretnych wartości wymienionych zmiennych.

Tabela 2. Znaczenia wartości zmiennych v48, v49, v50, v51, WRKHR

Wyszczególnienie	V48	V49	V50	V51	WRKHS_ grouped
1	Nigdy	Nigdy	Nigdy	Idealny	0-20h
2	Raz w miesiącu	Raz w miesiącu	Raz w miesiącu	Bardzo dobry	21-40h
3	Kilka razy w miesiącu	Kilka razy w miesiącu	Kilka razy w miesiącu	Dobry	41-60h
4	Kilka razy tygodniowo	Kilka razy tygodniowo	Kilka razy tygodniowo	Dostateczny	61-80h
5	Codziennie	Codziennie	Codziennie	Słaby	81-100h

Źródło: Opracowanie własne na podstawie danych ISSP

Zatem zmienna v51 dotycząca oceny stanu zdrowia ma wartości od 1 do 5, które oznaczają odpowiednio idealny i słaby stan zdrowia. Natomiast wartości zmiennych v47, v48, v49, v50 oznaczają częstotliwość w skali od 1 do 5. Poniższa tabela 3. przedstawia wybrane

statystyki opisowe dla wszystkich opisanych zmiennych zaokrąglone do drugiego miejsca po przecinku.

Tabela 3. Wybrane statystyki opisowe dla wykorzystywanych zmiennych

Wyszczególnienie	Średnia	Odchylenie standardowe	Skośność	Percentyl 5%	Mediana	Percentyl 95%
V47	1,93	1,38	1,50	1	1	5
V48	1,79	0,95	0,91	1	2	4
V49	3,18	1,26	1,60	1	3	5
V50	4,33	0,83	-1,33	2	5	5
V51	2,77	0,92	0,04	1	3	4
AGE	44,93	13,37	0,04	24	45	66
WRKHRS_grouped	2,35	0,76	0,45	1	2	3

Źródło: Opracowanie własne na podstawie danych ISSP.

Dla zmiennej v47 średnia wyniosła 1,93, co oznacza, że średnio osoba z próby nie pali, ale paliła w przeszłości, a mediana równa 1 wskazuje, że minimum połowa osób nigdy nie paliła papierosów. Odchylenie standardowe 1,38 wskazuje na dużą zmienność. Zmienna v48 dotycząca spożywania alkoholu ma podobną średnią, natomiast mediana wynosi 2, niższa wartość skośności wynosząca 0,91 wskazuje na mniejszą zmienność. W zmiennych dotyczących aktywności fizycznej, zdrowia, wieku, liczby przepracowanych godzin, czyli odpowiednio v49, v51, AGE, WRKHRS_grouped średnia jest bliska medianie. Sugeruje to, że rozkłady zmiennych są bliskie rozkładowi normalnemu. Szczególnie potwierdza to niska skośność wynosząca 0,04 przy zmiennych v51 oraz AGE. Przeciętna osoba w zbiorze danych ma 45 lat pracuje między 20, a 40 godzin w tygodniu, nie pali papierosów oraz ma dobry stan zdrowia. Zmienna v50 ma rozkład lewoskośny, co potwierdza, skośność na poziomie -1,33

oraz mediana wynosząca 5, a jest to maksymalna wartość. Oznacza to, że minimum połowa osób spożywa owoce i warzywa codziennie.

3. Opis zastosowanego modelu

W pracy wykorzystano wielomianowy uporządkowany model logitowy. Zmienna zależna obejmuje dyskretne wybory, które można łatwo uporządkować na skali Likerta.¹¹ W przypadku badań przedstawionych w tym artykule zmienna zależną jest stan zdrowia. Podczas definiowania modelu uporządkowanego zakłada się, że zmienna porządkowa y jest ograniczonym zapisem pewnej nieobserwowalnej zmiennej ciągłej y^* [Gruszczyński 2010]. Ta zmienna ciągła jest zależna liniowo od zbioru zmiennych objaśniających, zapisanych w wektorze x oraz nieznanymi parametrami, które znajdują się w wektorze β .¹²

$$y_i^* = x_i' \beta + u_i$$

gdzie $i = 1, 2, \dots, n$ oznaczają kolejne obiekty (respondenci), u_i to składniki losowe o rozkładzie logistycznym, opisane pewną funkcją gęstości $f(u)$ o zerowej wartości oczekiwanej i stałej wariancji. Przekształcenie zmiennej nieobserwowalnej na zmienną obserwowalną przyjmującą 5 wartości może zostać wykonane poprzez podział całego zakresu zmienności tej zmiennej (oznaczonej jako y^*) na 5 równych przedziałów. W tym celu trzeba wyznaczyć 4 punkty odcięcia, które będą określały granice tych przedziałów. Każdy z tych punktów odcięcia będzie reprezentował wartość, która oddziela sąsiadujące ze sobą przedziały. W ten sposób otrzymujemy 5 zakresów, które można przyporządkować do odpowiednich wartości obserwowalnych. Model logitowy opiera się na założeniu regresji równoległych (proporcjonalnych szans). Stosunek szans jest stały, nie zależy od wartości zmiennych objaśniających. Aby zweryfikować, czy przyjęte założenie nie jest zbyt restrykcyjne, konieczne jest przeprowadzenie testu statystycznego. W naszym badaniu zastosowaliśmy test Branta. Wyniki testu wskazały, że założenie proporcjonalnych szans zostało naruszone, więc oszacowano uogólniony model logitowy. Główna różnica między tym modelem a modelem standardowym polega na sposobie traktowania progów. W standardowym modelu progi są stałymi wartościami. W tym modelu natomiast zakłada się, że progi również są funkcjami liniowymi zmiennych objaśniających. Oznacza to, że wartości progowe mogą się zmieniać w

¹¹ Clark, A.E., Lelkes, O.: Deliver us from evil: religion as insurance. PSE working paper 2005-43 (2005).

¹² Adam Waszkowski, Wielomianowy uporządkowany model logitowy w prognozowaniu zagrożenia finansowego przedsiębiorstw.

zależności od wartości tych zmiennych. Dodatkowo, każdy próg ma swój własny zestaw parametrów, co pozwala na bardziej szczegółowe modelowanie zależności między zmiennymi objaśniającymi a progami.

Do oceny jakości modelu logitowego używa się mierników opartych na wartości funkcji wiarygodności, pseudo- R^2 (McKelveya i Zavoina, Coxa-Snella, Cragga i Uhlera) oraz kryteria informacyjne. Warto zaznaczyć, że tylko pseudo- R^2 McKelveya i Zavoina można by interpretować jak klasyczny R^2 dla modelu liniowego. Pozostałe wymienione, które są oparte na wartości funkcji wiarygodności, można wykorzystać jedynie do porównania modeli konkurencyjnych. Zdolność predykcyjna modelu została przedstawiona w tabeli klasyfikacyjnej. Interpretacji parametrów modelu dokonano za pomocą efektów krańcowych.

4. Weryfikacja oraz interpretacja modelu

Współliniowość ma ważne znaczenie w modelu, gdyż pozwala wykryć nadmierne zależności pomiędzy niezależnymi zmiennymi. Ich zbyt duża wartość może prowadzić do błędnych estymacji. Poniższa tabela 4. przedstawia macierz współliniowości dla zmiennych objaśniających w modelu.

Tabela 4. Macierz współliniowości dla zmiennych objaśniających

Wyszczególnienie	V47	V48	V49	V50	AGE	WRKHRS_grouped
V47	1					
V48	0,2345	1				
V49	-0,0496	0,0544	1			
V50	-0,1151	-0,0896	0,2094	1		
AGE	0,0083	-0,0468	-0,0284	0,1066	1	
WRKHRS_grouped	0,0493	0,0546	-0,0043	-0,0422	-0,0485	1

Źródło: Opracowanie własne na podstawie danych ISSP.

Najwyższą wartość otrzymano dla zmiennych V48 i V47 oraz v50 i v49 wyniosła odpowiednio 0,2345 i 0,2094. Jednak wartości poniżej 0,3 oznaczają nadal małą korelację tych zmiennych. Pozostałe wartości są mniejsze, zatem w danych nie występuje problem współliniowości.

W tabeli 6 przedstawiono wyniki uporządkowanego modelu logitowego. Jako zmienne niezależne przyjęto v47, v48, v49, v50, AGE, WRKHRS_grouped. Czynniki istotne statystycznie (przy poziomie istotności równym 0.05) to: v47 (wszystkie kategorie), v48 (Codziennie), v49 (Kilka razy w miesiącu, Kilka razy tygodniowo, Codziennie), v50 (Raz

w miesiącu lub rzadziej), AGE oraz WRKHRS_grouped (2, 3). Dodatni współczynnik oznacza, że dany czynnik zwiększa szansę wystąpienia w wyższej kategorii zmiennej objaśnianej w porównaniu do kategorii bazowej. Z kolei ujemny współczynnik ją zmniejsza.

Tabela 5. Wyniki uporządkowanego modelu logitowego

Wyszczególnienie	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
v47						
Do not smoke now but smoked in the past	0.1699646	0.037903	4.48	0	0.095676	0.244253
Smoke 1-5 cigarettes per day	0.1576826	0.068325	2.31	0.021	0.023769	0.291597
Smoke 6-10 cigarettes per day	0.1560704	0.064565	2.42	0.016	0.029525	0.282616
Smoke 11-20 cigarettes per day	0.1922358	0.05822	3.3	0.001	0.078126	0.306345
Smoke 21-40 cigarettes per day	0.677147	0.114047	5.94	0	0.453619	0.900675
Smoke more than 40 cigarettes per day	0.3407014	0.338898	1.01	0.315	-0.32353	1.00493
v48						
Once a month or less often	-0.0459177	0.034906	-1.32	0.188	-0.11433	0.022496
Several times a month	-0.0245254	0.04592	-0.53	0.593	-0.11453	0.065476
Several times a week	0.1178902	0.071309	1.65	0.098	-0.02187	0.257653
Daily	0.3818513	0.13393	2.85	0.004	0.119354	0.644349
v49						
Once a month or less often	0.0397222	0.055901	0.71	0.477	-0.06984	0.149286
Several times a month	-0.2739131	0.05097	-5.37	0	-0.37381	-0.17401
Several times a week	-0.6967637	0.048822	-14.27	0	-0.79245	-0.60108
Daily	-0.6849654	0.057816	-11.85	0	-0.79828	-0.57165
v50						
Once a month or less often	0.4699359	0.195055	2.41	0.016	0.087635	0.852237
Several times a month	0.2649297	0.178689	1.48	0.138	-0.08529	0.615153
Several times a week	0.0777301	0.174177	0.45	0.655	-0.26365	0.419111
Daily	-0.2564681	0.174072	-1.47	0.141	-0.59764	0.084706
AGE	0.0240235	0.001164	20.63	0	0.021742	0.026306
hours_grouped						
2	-0.1458992	0.051229	-2.85	0.004	-0.24631	-0.04549
3	-0.2230963	0.053069	-4.2	0	-0.32711	-0.11908
4	-0.0349235	0.096354	-0.36	0.717	-0.22377	0.153926
5	0.0802455	0.144589	0.55	0.579	-0.20314	0.363635
/cut1	-2.020509	0.185316			-2.38372	-1.6573
/cut2	0.0518687	0.184065			-0.30889	0.41263
/cut3	1.960251	0.18472			1.598207	2.322295
/cut4	4.420214	0.191272			4.045328	4.7951

Źródło: Opracowanie własne na podstawie danych ISSP.

Po oszacowaniu pełnego modelu, oszacowano model z ograniczeniami. W nowym modelu nie uwzględniono zmiennej V50. Wynik test ilorazu wiarygodności przedstawiony w

poniższej tabeli 6., wskazuje na odrzucenie hipotezy o tym, że model o mniejszej ilości zmiennych (model_2) zawiera się w modelu o większej ilości zmiennych. To wskazuje, że zastosowane ograniczenia względem modelu są nieprawidłowe.

Tabela 6. Test Ilorazu wiarygodności

Model	N	log-likelihood (null)	log-likelihood (model)	df	AIC	BIC
model2	15,320	-20389.2	-19837.26	23	39720.51	39896.16
model1	15,320	-20389.2	-19745.22	27	39544.43	39750.63
Założenie: model2 zagnieżdżony w model1						
LR chi2(4)	184.08					
Prob > chi2	0.0000					

Źródło: Opracowanie własne na podstawie danych ISSP.

Uporządkowany model zakłada równoległość regresji, co oznacza jednakowy wpływ zmiennych objaśnianych na zmienną zależną. Do weryfikacji testu wykorzystano test Branta. Wyniki testu przedstawiono w poniższej tabeli 7.

Tabela 7. Test Branta

Wyszczególnienie	Chi2	p>ch2
Wszystkie zmienne	324,59	0,000

Źródło: Opracowanie własne na podstawie danych ISSP.

Na poziomie istotności 0,05 odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej, że występują istotne różnice we współczynnikach regresji w zależności od kategorii. W teście Branta na poziomie istotności 0,05 istotnie statycznie były zmienne v47 w drugiej, trzeciej i szóstej kategorii, v48 w trzeciej, czwartej i piątej kategorii oraz v49 w piątej kategorii. Zatem założenia modelu zostały naruszone i dalej zostanie wykorzystana uogólniona wersja uporządkowanej regresji liniowej.

Uogólniony uporządkowany model logitowy oszacowano z użyciem opcji „autofit”. Opcja ta testuje, które zmienne lub poziomy zmienne mogą naruszać założenie

proporcjonalnych szans oraz automatycznie dostosowuje model. To pozwoliło zaoszczędzić czas, eliminując konieczność ręcznego testowania i dostosowywania modelu. Jako bazowe poziomy zmiennych niezależnych przyjęto poziom 1. Kategoria bazowa zmiennej zależnej to „Słaby”. Zdolność predykcyjną modelu oceniono na podstawie tabeli 8.

Tabela 8. Tabela klasyfikacji

Przynależność rzeczywista	Prognoza przynależności do grupy			
Stan zdrowia	2	3	4	Razem
1 -Idealny	458	751	2	1 211
2- Bardzo dobry	1 330	3 321	15	4 666
3- Dobry	1 041	5 109	60	6 210
4- Dostateczny	319	2 498	55	2 872
5- Słaby	35	314	12	361
Razem	3 185	11 996	148	15 320

Źródło: Opracowanie własne na podstawie danych ISSP.

Model nie klasyfikuje przypadków do kategorii 1 oraz 5. Najlepiej prognozowaną kategorią jest 3 - Dobry, gdzie 5109 przypadków zostało sklasyfikowanych poprawnie. Zliczeniowy R², czyli stosunek liczby poprawnie zaklasyfikowanych grup, równy 6494, do liczby wszystkich 15320 obserwacji wynosi 42,4%. Jeśli model przypisałby wszystkie obserwacje do największej grupy, wartość zliczeniowego R² wyniosłaby 40,5%, czyli niewiele mniej niż wynik uzyskany przez model, co nie przemawia na jego korzyść. W tabeli 9. przedstawiono pozostałe miary oceny jakości modelu.

Tabela 9. Miary oceny jakości modelu

Log-likelihood	
Model	-19623.876
Intercept-only	-20389.197
Chi-square	
Deviance (df=15257)	39247.752
LR (df=59)	1530.643
p-value	0
R ²	
McFadden	0.038
McFadden (adjusted)	0.034
Cox-Snell/ML	0.095
Cragg-Uhler/Nagelkerke	0.102
Count	0.429

Count (adjusted)	0.039
IC	
AIC	39373.752
AIC divided by N	2.57
BIC (df=63)	39854.878

Źródło: Opracowanie własne na podstawie danych ISSP.

Niska wartość p-value przy teście chi-kwadrat sugeruje, że model ma istotne statystycznie dopasowanie do danych. Wartości AIC oraz BIC mogą posłużyć jako kryterium porównawcze do innych modeli. W celu zrozumienia parametrów oszacowanego modelu użyjemy podejścia opierającego się na efektach krańcowych. Efekty te określają wpływ zmian wartości zmiennych objaśniających na prawdopodobieństwo przynależności do poszczególnych kategorii.

Palacze należący do grupy: „Smoke 21-40 cigarettes per day” mają *ceteris paribus* o 4,3% mniejsze prawdopodobieństwo wyboru bardzo dobrej oceny zdrowia, o 11,7% większe prawdopodobieństwo wyboru kategorii dostatecznej oceny zdrowia oraz o 2,3% większe prawdopodobieństwo słabej oceny zdrowia niż osoby niepalące („Do not smoke and never did”).

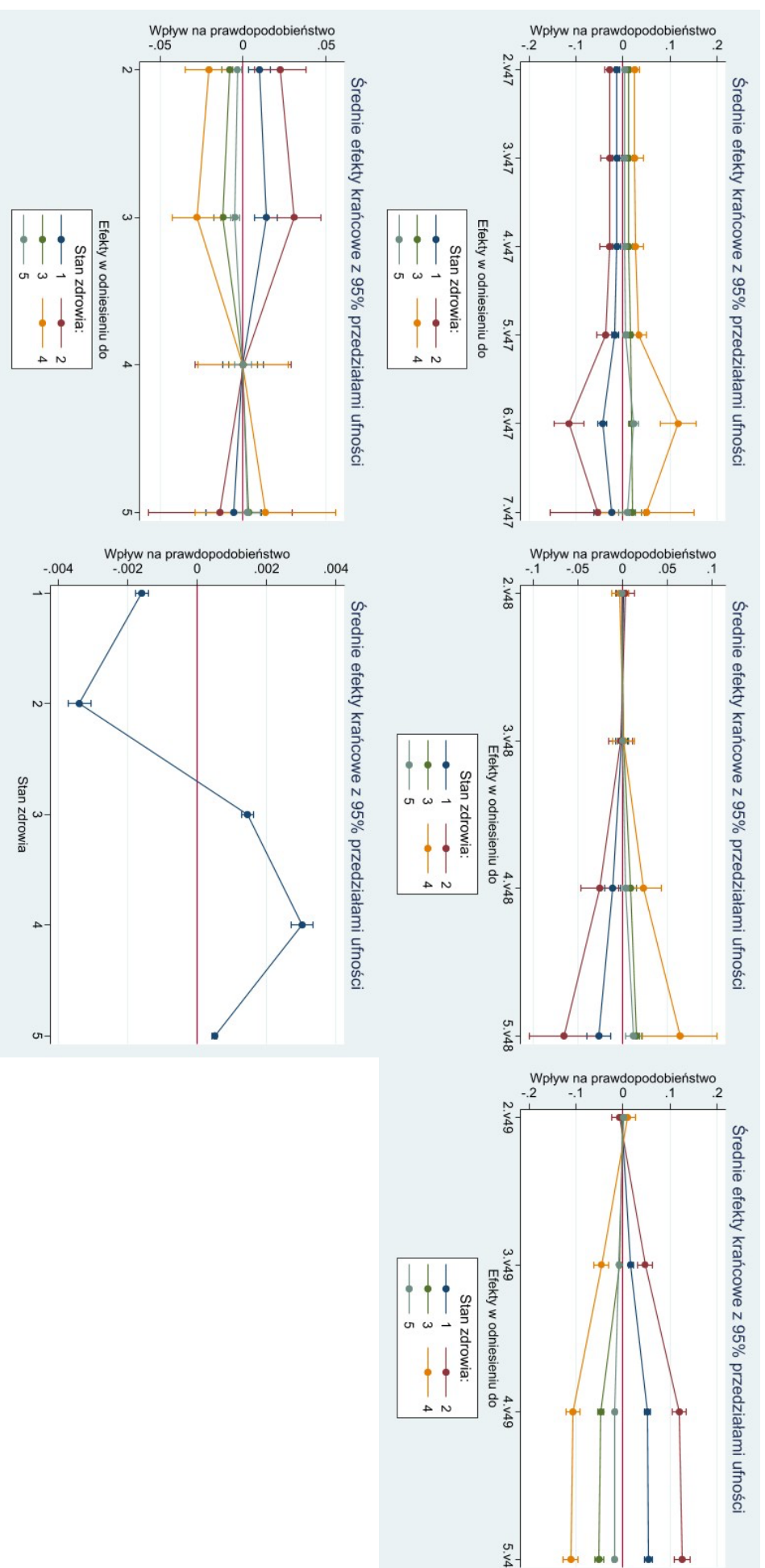
Osoby, które codziennie spożywają alkohol w ilości 4 lub więcej drinków mają *ceteris paribus* o 2,6% oraz 6,6% mniejsze prawdopodobieństwo wyboru kategorii odpowiednio idealnego oraz bardzo dobrego zdrowia niż abstynenci alkoholowi. Mają oni także o 6,4% i 1,2% większe prawdopodobieństwo wyboru odpowiednio dostatecznej i słabej oceny zdrowia niż osoby niespożywające alkoholu, *ceteris paribus*.

Osoby aktywne fizycznie przez przynajmniej 20 minut dziennie mają *ceteris paribus* o 5,5% i 12,5% większe prawdopodobieństwo wyboru kategorii odpowiednio idealnej i bardzo dobrej oceny zdrowia niż osoby nie wykazujące aktywności fizycznej. Dodatkowo te osoby mają 11,1% oraz 1,8% mniejsze prawdopodobieństwo wyboru odpowiednio dostatecznej oraz słabej oceny zdrowia w porównaniu do osób niećwiczących, *ceteris paribus*.

Osoby pracujące 41-60 godzin tygodniowo mają *ceteris paribus* o 1,3% i 3,1% większe prawdopodobieństwo wyboru odpowiednio kategorii oceny idealnej oraz bardzo dobrej zdrowia, niż osoby, które pracują tygodniowo do 20 godzin. Te osoby mają o 2,8% i 0,4% mniejsze prawdopodobieństwo wyboru odpowiednio dostatecznej oraz słabej oceny zdrowia, *ceteris paribus*. W ankietach odpowiadały osoby powyżej 16 roku życia, zatem przypuszczalnie

osoby pracujące mniejszą liczbę godzin niż 20, mogą mieć większe problemy zdrowotne w danym momencie, być niezdolne do pracy lub na emeryturze. Przy wzroście wieku o jeden rok, prawdopodobieństwo wyboru kategorii perfekcyjnej oraz bardzo dobrej oceny zdrowia spada odpowiednio o około 0,2% oraz 0,3%, *ceteris paribus*. Natomiast wtedy wzrasta prawdopodobieństwo wyboru oceny dobrej, dostatecznej i słabej o odpowiednio 0,1%, 0,3% oraz 0,05%. Zmianę efektów krańcowych przedstawiono na poniższym wykresie panelowym (wykres 10.).

Wykres 10. Średnie efekty krańcowe



Źródło: Opracowanie własne na podstawie danych ISSP.

5. Zakończenie

Podsumowując, model w istotny sposób pokazuje, że styl życia ma wpływ na ocenę zdrowia. Przyjmujemy zatem hipotezę badawczą, wedle której palenie papierosów, spożywanie alkoholu oraz uprawianie sportu ma istotny wpływ na stan zdrowia. Według badań przeprowadzonych w pracy palenie oraz alkohol mają negatywne oddziaływanie na zdrowie. Codzienne palenie oraz picie alkoholu zmniejszają prawdopodobieństwo, że respondent odpowie, że ma perfekcyjny lub bardzo dobry stan zdrowia. Inaczej jest w przypadku aktywności fizycznej. Codzienny sport przez minimum 20 minut zwiększa istotnie prawdopodobieństwo pozytywnej oceny zdrowia oraz zmniejsza prawdopodobieństwo oceny go w kategorii dostatecznej lub słabej.

Warto podkreślić, że prowadzony przez ludzi styl życia może poprawić stan zdrowia. Zdrowe społeczeństwo to klucz dla rozwoju każdej gospodarki. Z pewnością zmiana nawyków jest niełatwym zadaniem, lecz ważna jest edukacja na ten temat i świadomość, że codzienne czynności mają kluczowe znaczenie dla zdrowia w długoterminowym okresie.

6. Kod

Poniżej przedstawiono kod w programie STATA do replikacji obliczeń.

```
import spss country v47 v48 v49 v50 v51 AGE WRKHRS using "../ZA8000_v1-0-0.sav"

* Data Cleaning & Feature Engineering
*-----

gen to_use = !missing(v47, v48, v49, v50, v51, AGE, WRKHRS)

drop if to_use == 0

gen WRKHRS_grouped = WRKHRS

recode WRKHRS_grouped (0/20 = 1) (21/40 = 2) (41/60 = 3) (61/80 = 4) (81/100 = 5)

tab WRKHRS_grouped

* Data analysis
*-----

summarize, detail

* Models #1 Ordered logistic regression
*-----

ologit v51 i.v47 i.v48 i.v49 i.v50 c.AGE i.WRKHRS_grouped

estimates store model1

label variable _est_model1 "all variables"

ologit v51 i.v47 i.v48 i.v49 c.AGE i.WRKHRS_grouped

estimates store model2

label variable _est_model2 "without v50 - fruit & vegetables"

lrtest model1 model2
```

```

* Testing model assumptions
*-----
estimates restore modell
brant

* Models #2 Generalized Ordered Logit
*-----
gologit2 v51 i.v47 i.v48 i.v49 i.v50 c.AGE i.WRKHRS_grouped, autofit

* Model evaluation
*-----
fitstat

estimates restore modell

fitstat

* Classification Table
estimates restore model2

predict true1, outcome(1)
predict true2, outcome(2)
predict true3, outcome(3)
predict true4, outcome(4)
predict true5, outcome(5)

generate float max_prob =max(true1, true2, true3, true4, true5)
generate float prediction = .

replace prediction = 1 if max_prob==true1
replace prediction = 2 if max_prob==true2
replace prediction = 3 if max_prob==true3
replace prediction = 4 if max_prob==true4
replace prediction = 5 if max_prob==true5
table (v51) (prediction), stat(frequency)

* Marginal effects
margins, dydx(v47)
marginsplot , yline(0)
margins, dydx(v47) predict(outcome(1))
margins, dydx(v47) predict(outcome(2))
margins, dydx(v47) predict(outcome(3))
margins, dydx(v47) predict(outcome(4))
margins, dydx(v47) predict(outcome(5))

margins, dydx(v48)
marginsplot , yline(0)
margins, dydx(v48) predict(outcome(1))
margins, dydx(v48) predict(outcome(2))
margins, dydx(v48) predict(outcome(3))
margins, dydx(v48) predict(outcome(4))
margins, dydx(v48) predict(outcome(5))

margins, dydx(v49)
marginsplot , yline(0)
margins, dydx(v49) predict(outcome(1))
margins, dydx(v49) predict(outcome(2))
margins, dydx(v49) predict(outcome(3))
margins, dydx(v49) predict(outcome(4))
margins, dydx(v49) predict(outcome(5))

margins, dydx(WRKHRS_grouped)
marginsplot , yline(0)
margins, dydx(WRKHRS_grouped) predict(outcome(1))
margins, dydx(WRKHRS_grouped) predict(outcome(2))
margins, dydx(WRKHRS_grouped) predict(outcome(3))
margins, dydx(WRKHRS_grouped) predict(outcome(4))
margins, dydx(WRKHRS_grouped) predict(outcome(5))

margins, dydx(AGE)
marginsplot , yline(0)
margins, dydx(AGE) predict(outcome(1))
margins, dydx(AGE) predict(outcome(2))
margins, dydx(AGE) predict(outcome(3))
margins, dydx(AGE) predict(outcome(4))
margins, dydx(AGE) predict(outcome(5))

```