

Uniwersytet Warszawski  
Wydział Nauk Ekonomicznych

Mateusz Tomczak  
Nr albumu: 432561

**Edukacja rodziców a oceny dziecka – analiza  
wpływu edukacji rodziców na osiągnięcia dziecka w  
szkole**

Praca zaliczeniowa  
Ekonometria  
ćwiczenia, gr. 304  
WNE UW

Warszawa, styczeń 2023

## **Streszczenie**

*Praca przedstawia analizę czynników wpływających na osiągnięcia dzieci w szkole. Opisane zostały badania naukowe i ich wyniki, które wykorzystano jako uzasadnienie teoretyczne sformułowanych hipotez badawczych. Następnie zbudowany został model, który następnie zinterpretowano oraz zbadano pod kątem istotności oszacowań. Główna hipoteza badawcza, stanowiąca o pozytywnej zależności edukacji rodziców i osiągnięć w szkole dziecka została potwierdzona jedynie w przypadku edukacji matki. Dane otrzymane w modelu okazały się również sprzeczne z pierwszą dodatkową hipotezą badawczą, mówiącą o istotności wpływu zaangażowania rodziców w edukację dziecka a jego osiągnięciami w szkole. Jako powód tego wskazane zostały ograniczenia wykorzystanych danych. Ostatnia z hipotez, która wskazywała na zależność wedle której dzieci uczęszczające na zajęcia dodatkowe otrzymywały wyższe oceny w porównaniu do swoich rówieśników, również nie wykazała zgodności z otrzymanymi oszacowaniami modelu, a jako potencjalne powody takich wniosków wskazane zostały ponownie ograniczenia związane z danymi, na których zbudowany został model.*

## **Słowa kluczowe**

ekonometria, edukacja, ekonomia

## **Dziedzina pracy**

Edukacja

## SPIS TREŚCI

WSTĘP.....	3
ROZDZIAŁ I. Hipotezy badawcze oraz podstawy teoretyczne .....	4
1.1. Wyższy poziom edukacji rodziców przekłada się na wyższe oceny dziecka w szkole .....	4
1.2. Aktywne wsparcie rodziny w nauce dziecka przekłada się na wyższe osiągnięcia w szkole .....	5
1.3. Uczniowie uczęszczający na zajęcia dodatkowe osiągają lepsze wyniki w szkole w porównaniu do ich rówieśników.....	6
ROZDZIAŁ II. Dane wykorzystane do budowy modelu .....	7
2.1. Zbiór danych oraz wykorzystane zmienne.....	7
2.2. Wstępna analiza danych.....	9
ROZDZIAŁ III. Analiza ekonometryczna.....	10
3.1. Model ekonometryczny.....	10
3.2. Interpretacja oszacowanych wartości.....	12
3.3. Testy diagnostyczne modelu, obserwacje odstające oraz współliniowość .....	13
WNIOSKI .....	17
OGRANICZENIA MODELU .....	18
BIBLIOGRAFIA .....	19
ZESTAWIENIE SPISÓW .....	20

## WSTĘP

Edukacja jest znaczącym wyznacznikiem przyszłych osiągnięć. Dlatego istotne jest wskazanie czynników determinujących osiągnięcia dzieci w szkole. Otwiera to możliwość do lepszego dostosowania edukacji w celu zwiększenia jej jakości oraz efektywności.

Celem przedstawionej w tym tekście analizy jest budowa modelu ekonometrycznego, który pozwoli na weryfikację sformułowanych hipotez badawczych. Główna hipoteza stwierdza, że wyższy poziom edukacji rodziców przekłada się na wyższe oceny dziecka w szkole. Pierwsza hipoteza poboczna zakłada, iż aktywne wsparcie rodziny przekłada się na wyższe osiągnięcia dziecka, natomiast druga mówi o osiąganiu lepszych wyników, w porównaniu do swoich rówieśników, przez uczniów uczęszczających na zajęcia dodatkowe.

W pierwszym rozdziale analizowana jest literatura naukowa i przytoczone w niej wyniki badań. Następnie wysunięte z wymienionej literatury wnioski są porównywane do sformułowanych wcześniej hipotez badawczych.

Drugi rozdział skupia się na analizie wykorzystanej do zbudowania przedstawionego w pracy modelu bazy danych oraz na wstępnej weryfikacji wyselekcjonowanych do modelu zmiennych w celu eliminacji potencjalnych zależności mogących negatywnie wpłynąć na oszacowania modelu.

Trzeci rozdział przedstawia budowę modelu ekonometrycznego przy użyciu metody najmniejszych kwadratów. Przybliżony zostaje proces selekcji zmiennych oraz zostają uzasadnione podjęte decyzje wpływające na ostateczny wygląd modelu. Otrzymane wyniki są analizowane oraz interpretowane. Następnie model badany jest pod kątem spełnienia założeń klasycznego modelu regresji liniowej. Model zostaje również zweryfikowany pod kątem występowania obserwacji odstających oraz autokorelacji zmiennych.

W podsumowaniu, na bazie otrzymanych oszacowań modelu, zinterpretowane zostają wyniki analizy a wnioski z niej wyciągnięte zestawiane są z przyjętymi hipotezami badawczymi w celu ich weryfikacji. Następnie opisane zostają ograniczenia zbudowanego modelu oraz możliwości jego ulepszenia.

## **ROZDZIAŁ I**

### **Hipotezy badawcze oraz podstawy teoretyczne**

Przeprowadzona w tej pracy analiza ma na celu weryfikację hipotezy, wskazującej na pozytywną, istotną zależność edukacji rodziców i osiągnięć ich dzieci w szkole. Dodatkowo zaproponowane zostały dwie hipotezy poboczne – wskazujące na pozytywny wpływ zaangażowania rodziców oraz uczestnictwa w zajęciach dodatkowych na oceny dziecka w szkole. Poniżej przedstawione zostały badania oraz analizy naukowe wskazujące na teoretyczne i empiryczne uzasadnienia zaproponowanych wyżej hipotez badawczych.

#### **1.1. Wyższy poziom edukacji rodziców przekłada się na wyższe oceny dziecka w szkole**

Literatura naukowa wskazuje na znaczącą rolę edukacji rodziców w analizowaniu i przewidywaniu wyników osiąganych przez ich dzieci w szkole. P. Raudenska oraz D. Hamplova (2022) na podstawie danych dotyczących czeskich rodzin wskazały, że wykształcenie rodziców oraz ich dochody miały statystycznie istotny, niebezpośredni wpływ na osiągnięcia dziecka w szkole. Zaproponowany w pracy model zakładał, że efekty edukacji oraz dochodu transmitowane były poprzez aspiracje rodziców, pozytywne podejście do dziecka oraz pozytywny związek emocjonalny rodziców z dzieckiem. W efekcie przedstawione aspekty środowiska domowego wpływały na aspiracje dzieci, ich samopoczucie oraz zainteresowanie literaturą, co bezpośrednio pozytywnie wpływało na oceny dzieci w szkole. Ważnym wnioskiem z pracy jest również różnica wpływu edukacji i dochodu rodziców na dziecko ze względu na jego płeć. W przypadku chłopców wpływ ten nie był znacząco zauważalny poprzez przedstawione niebezpośrednie drogi wpływu, w przeciwieństwie do dziewczyn, w przypadku których badane charakterystyki rodziców przekładały się niebezpośrednio na ich aspiracje i samopoczucie. Opierając się na wnioskach z przytoczonej pracy edukacja rodziców powinna wskazywać pozytywną zależność z osiągnięciami dziecka w szkole. Równie ciekawym wnioskiem jest również wykazany pozytywny wpływ dochodu rodziny na badane zjawisko. Niestety wykorzystana w tej pracy baza danych nie zawierała odpowiednich danych do przeprowadzenia takiej analizy.

Również E.F. Dubow, P. Boxer oraz L.R. Huesmann (2009) wskazują na istotny wpływ edukacji rodziców na kształtowanie przyszłości dziecka. Analiza autorów opierała się na 856 obserwacjach pozyskanych z badania Columbia County Longitudinal Study, które rozpoczęło

się w 1960, kiedy to poprzez ankietę pozyskano informację dotyczące trzecioklasistów i ich rodziców. Uczestnicy zostali następnie ponownie ankietowani w wieku 19, 30 oraz 48 lat. Analiza wykazała istotną korelację wyników uczniów w szkole średniej z edukacją ich rodziców. Autorzy wskazują na pozytywny wpływ wykształcenia rodziców w okresie młodego wieku dziecka zarówno na jego osiągnięcia w szkole jak i na edukację wyższą i osiągnięcia zawodowe w okresie dorosłości. Również w przypadku tego badania wskazano, iż wpływ edukacji rodziców na osiągnięcia dzieci jest niebezpośredni – ponownie nadmieniono tu rolę jaką odgrywa wykształcenie rodziców na kształtowane aspiracji dziecka. Długoterminowy charakter tej analizy i jej wyniki wskazują, jak ważną rolę w procesie edukacyjnym dziecka odgrywa poziom edukacji rodziców.

Jak zostało zauważone we wskazanych badaniach, edukacja rodziców odgrywa znaczącą rolę w edukacji dziecka. Zależność ta wykazana została zarówno w przypadku okresu krótkoterminowego jak i długoterminowego – odgrywając istotną rolę w kształtowaniu przyszłości zawodowej dziecka. Wnioskując z przytoczonych analiz, przeprowadzone w tej pracy badanie powinno wskazać na pozytywny, istotny efekt edukacji rodziców na osiągnięcia dziecka w szkole.

## **1.2. Aktywne wsparcie rodziny w nauce dziecka przekłada się na wyższe osiągnięcia w szkole.**

Także w przypadku zaangażowania rodziców w edukację dziecka literatura naukowa wskazuje na pozytywną zależność z jego osiągnięciami w szkole. William H. Jeynes (2007) w przeprowadzonej metaanalizie, opartej na tekstach naukowych pozyskanych z 25 baz danych, potwierdza tę hipotezę. Autor wskazuje także na wykazywaną w badaniach istotność poszczególnych komponentów, przez które rodzice przejawiają swoje zaangażowanie. Styl wychowania oraz oczekiwania względem dziecka przejawiały się jako bardziej wpływowe na edukację dziecka niż takie czynniki jak zasady panujące w domu, dla których przeanalizowane badania nie wykazały istotnej statystycznie zależności, czy obecność rodziców na zebraniach w szkole, w przypadku której wpływ na osiągnięcia dziecka był niejednoznaczny.

Również A.R. Gonzalez-DeHass, P.P. Willems oraz M.F. Doan Holbein (2005) w przeprowadzonej analizie literatury naukowej wskazują na istotność zaangażowania rodziców na osiągnięcia dziecka w szkole. Przeanalizowane badania pokazały, że dzieci rodziców, których cechowało zaangażowanie w edukację swojego dziecka, wykazywały

wyższe poziomy koncentracji oraz uwagi. Przejawy zaangażowania w formie zapewniania wsparcia oraz pochwał przekładały się na zwiększenie wewnętrznej motywacji dziecka. Autorki wskazały, że ogólnym efektem przedstawionych czynników na dzieci było bardziej prawdopodobne wytrwanie w trudnościach związanych z edukacją czy wyższa satysfakcja związana z nauką. Wnioski te wskazują na zaangażowanie rodziców jako kluczowy czynnik wpływający na osiągnięcia dziecka w szkole.

Przytoczona literatura potwierdza tezę o istotności aktywnego wsparcia rodziców na osiągnięcia dziecka w szkole. Jednak zależność ta może różnić się między aspektami, przez które takie zaangażowanie można rozumieć. Jak zostało wskazane sposób wychowania dziecka a zasady panujące w domu mogą cechować się niejednorodną istotnością na osiągnięcia w szkole. Jednak wnioskując z zaprezentowanych analiz można spodziewać się dodatniej zależności zaangażowania rodziców i efektów edukacji dziecka.

### **1.3. Uczniowie uczęszczający na zajęcia dodatkowe osiągają lepsze wyniki w szkole w porównaniu do ich rówieśników.**

M. Fruda oraz M. Shuleski (2019) w analizie przeprowadzonej na podstawie średniej ocen 148 uczniów wskazali, że uczniowie uczęszczający na przynajmniej jedno dodatkowe zajęcie osiągalni lepsze wyniki w szkole w porównaniu do reszty uczniów. Na podstawie przeprowadzonych ankiet autorzy pokazali również, że uczniowie uczęszczający na zajęcia poza programowe częściej wskazywali, że chodzenie do szkoły sprawia im przyjemność oraz posiadali lepszą opinię o swojej szkole. Wnioski z analizy wskazują, że uczęszczanie na zajęcia dodatkowe istotnie przekładało się na osiągnięcia w szkole jak i na opinię o edukacji. Autorzy jednak zauważają, że ze względu na ograniczoną próbę badawczą wymagane jest przeprowadzenie szerszych badań w celu uzyskania dokładniejszych wniosków.

Mimo swoich ograniczeń, wskazane badanie potwierdza, zaproponowaną hipotezę. Wyniki przedstawionej analizy wskazują, że uczęszczanie na zajęcia dodatkowe powinno przełożyć się na wyższe osiągnięcia w szkole dzieci na nie uczęszczających, w porównaniu do reszty uczniów.

## ROZDZIAŁ II

### Dane wykorzystane do budowy modelu

#### 2.1. Zbiór danych oraz wykorzystane zmienne

W celu zbudowania modelu wykorzystana została baza danych złożona z 395 obserwacji zebranych w latach 2005-2006. Jej źródłem jest praca pt. „Using Data Mining to Predict Secondary School Student Performance” (2008), której autorami są P. Cortez oraz A. Silva. Obserwacje dotyczą uczniów dwóch publicznych portugalskich szkół średnich, ich sytuacji rodzinnej oraz wyników z przedmiotów matematyka oraz język portugalski.

Jako zmienna objaśniana w przedstawionej w pracy analizie przyjęta została końcowa ocena z matematyki (*grade*). Przyjmuje ona wartości od 0 do 20, gdzie 0 jest najniższą, a 20 najwyższą możliwą do uzyskania oceną. Zmienne objaśniające, ich opis oraz przyjmowane wartości przedstawione zostały w tabeli 1.

Tabela 1. Zmienne objaśniające wykorzystane w modelu

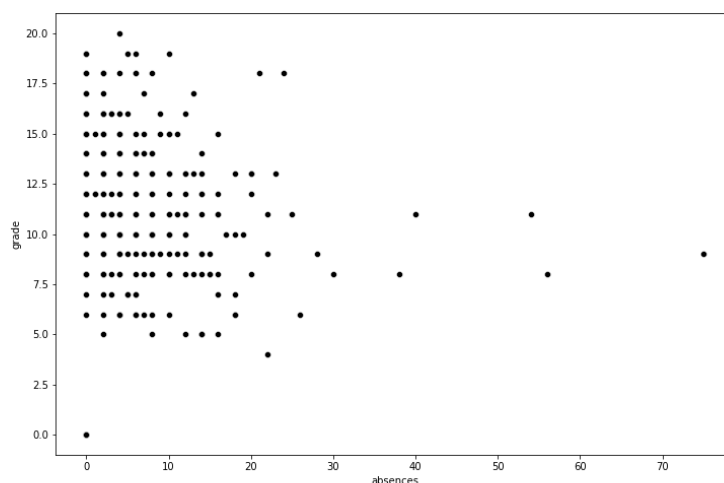
Lp.	Zmienna	Opis	Przyjmowane wartości
1.	<i>sex</i>	Płeć ucznia	1 – kobieta 0 – mężczyzna
2.	<i>age</i>	Wiek ucznia	od 15 do 22
3.	<i>studytime</i>	Ilość czasu w tygodniu przeznaczana na naukę	1 – <i>studytime</i> > 5h 0 – <i>studytime</i> ≤ 5h
4.	<i>had_failed</i>	Czy uczeń nie zaliczył 1 lub więcej przedmiotów	1 – tak 0 – nie
5.	<i>schoolsup</i>	Czy uczeń uzyskuje w szkole dodatkowe wsparcie edukacyjne	1 – tak 0 – nie
6.	<i>famsup</i>	Czy rodzina wspiera ucznia w edukacji	1 – tak 0 – nie
7.	<i>paid</i>	Czy uczeń uczęszcza na płatne zajęcia dodatkowe z matematyki	1 – tak 0 – nie
8.	<i>activities</i>	Czy uczeń uczęszcza na zajęcia dodatkowe	1 – tak 0 – nie
9.	<i>higher</i>	Czy uczeń planuje pójść na studia	1 – tak 0 – nie
10.	<i>freetime</i>	Ocena ilości czasu wolnego ucznia po szkole	od 1 (bardzo mało) do 5 (bardzo dużo)
11.	<i>absences</i>	Liczba nieobecności na zajęciach	od 0 do 93
12.	<i>absencessq</i>	Liczba nieobecności na zajęciach podniesiona do kwadratu	od 0 do 8649
13.	<i>Msec</i> oraz <i>Fsec</i>	Wykształcenie średnie jako najwyższy poziom wykształcenia odpowiednio matki oraz ojca	1 – tak 0 – nie
14.	<i>Mhigh</i> oraz <i>Fhigh</i>	Wykształcenie wyższe jako najwyższy poziom wykształcenia odpowiednio matki oraz ojca	1 – tak 0 – nie

Źródło: Opracowanie własne



Zmienne *Msec* i *Fsec* oraz *Mhigh* i *Fhigh* zostały opracowane na podstawie zmiennych *Medu* oraz *Fedu* z wykorzystanej bazy danych, gdzie dane dotyczące edukacji, odpowiednio matki oraz ojca, przedstawione zostały w skali 0-4. Przy przekształceniu zmiennych za podstawę edukacji rodzica w modelu przyjęto wykształcenie podstawowe lub brak wykształcenia. Zmieniona została również zmienna *studytime*. W wykorzystanej bazie danych przyjmowała ona wartości od 1 do 4 dla następujących, w kolejności, przedziałów czasu przeznaczanego na naukę w tygodniu: mniej niż 2h, od 2h do 5h, od 5h do 10h, powyżej 10h. Przekształcenie to spowodowane było faktem, że zmienna ta nie przedstawiała faktycznej ilości czasu wykorzystywanego na naukę przez ucznia, tylko wskazywała na numer odpowiedniego przedziału. W celu lepszego odzwierciedlenia informacji zawartej we wskazanej zmiennej, została ona przekształcona w taki sposób, by informować czy dany uczeń przeznacz średnio mniej lub więcej niż jedną godzinę na naukę dziennie, zakładając, że nie uczy się w weekend. Wprowadzenie zmiennej *absencesq* spowodowane było natomiast nieliniową zależnością między zmiennymi *absences* oraz *grade*. Wykres tej zależności przedstawiony został na rysunku 1.

Rysunek 1. Wykres zależności zmiennych *absences* oraz *grade*.



Źródło: Opracowanie własne przy użyciu biblioteki seaborn dla języka Python.

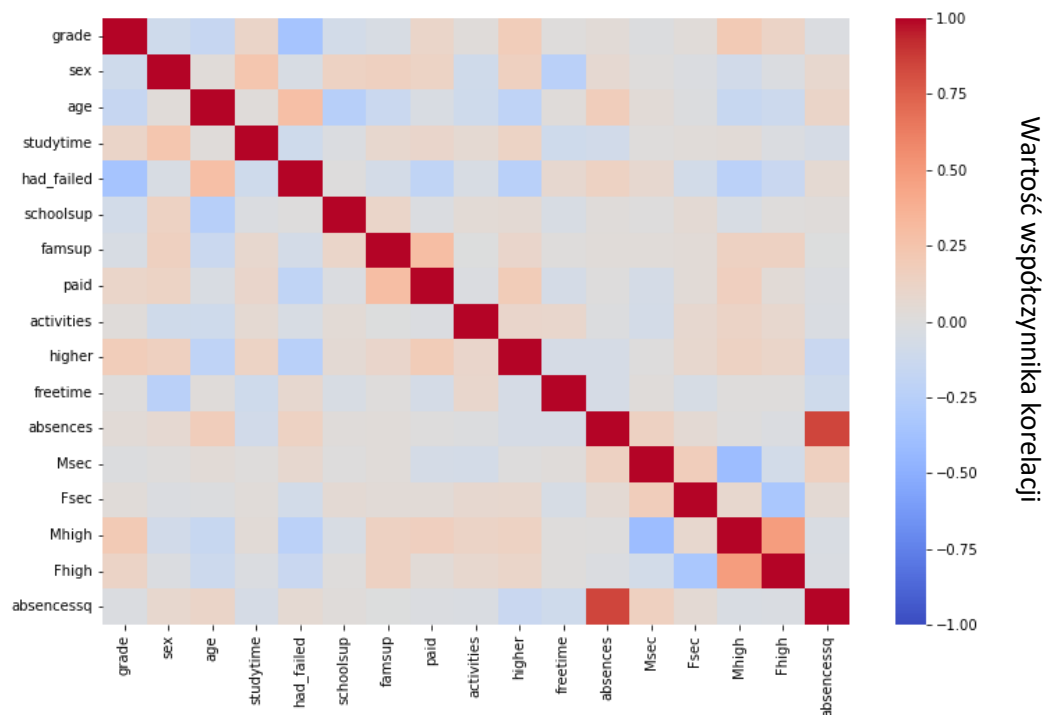
Dodatkowo do modelu zostały wprowadzone dwie interakcje: *absences\*had\_failed* oraz *higher\*studytime*. Dodanie pierwszej ma na celu wprowadzenie do modelu potencjalnej zależności, wedle której niezaliczenie jednego lub więcej przedmiotów przez ucznia spowodowane było nieobecnością na zajęciach. Natomiast druga interakcja wskazuje, że uczniowie, którzy są już zdecydowani na kontynuację swojej edukacji na studiach będą bardziej zdeterminowani do osiągnięcia lepszych wyników w szkole, przez co bardziej

prawdopodobne jest, że będą poświęcać na naukę więcej czasu niż uczniowie, którzy nie podjęli takiej decyzji.

## 2.2. Wstępna analiza danych

Przed określeniem ostatecznej formy modelu, który zostanie przeanalizowany, wykorzystane zmienne zostały zbadane pod kątem korelacji. Wyniki w postaci heatmapy zostały przedstawione na rysunku 2. Jak widać na wskazanej grafice, między zmiennymi wykorzystanymi w modelu nie występują wysokie korelacje, z wyjątkiem zależności między zmiennymi *absences* oraz *absencessq*, co jest spodziewane ze względu na to, że druga zmienna jest kwadratem pierwszej. Nie istnieje zatem potrzeba wykluczenia żadnej ze zmiennych z modelu.

Rysunek 2. Korelacja zmiennych wykorzystanych w modelu w formie heatmapy.



Źródło: Opracowanie własne przy użyciu biblioteki seaborn dla języka Python.

## ROZDZIAŁ III

### Analiza ekonometryczna

#### 3.1. Model ekonometryczny

Mając na uwadze wnioski z poprzedniego rozdziału, wyjściowa forma modelu, który zostanie przeanalizowany przyjmuje następującą postać:

$$\begin{aligned} grade_i = & \beta_0 + \beta_1 sex_i + \beta_2 age_i + \beta_3 studytime_i + \beta_4 had\_failed_i + \beta_5 schoolsup_i \\ & + \beta_6 famsup_i + \beta_7 absences_i + \beta_8 absencessq_i + \beta_9 paid_i \\ & + \beta_{10} activities_i + \beta_{11} higher_i + \beta_{12} freetime_i + \beta_{13} Msec_i + \beta_{14} Fsec_i \\ & + \beta_{15} Mhigh_i + \beta_{16} Fhigh_i + \beta_{17} absences_i * had\_failed_i + \beta_{18} higher_i \\ & * studytime_i + \varepsilon_i \end{aligned}$$

Model ten został następnie oszacowany przy pomocy metody najniższych kwadratów. W analizie przyjęty został poziom istotności na poziomie 10%. Zmienne nieistotne statystycznie były sukcesywnie usuwane, co zostało przedstawione w tabeli 2. Dodatkowo dla modelu 1 przetestowana została hipoteza o łącznej nieistotności zmiennych. Statystyka testowa F wyniosła 7.037, natomiast p-value = 7.044e-15. Zatem hipoteza o łącznej nieistotności zmiennych została odrzucona.

Ostatecznie do dalszej analizy wybrany został model 5. Decyzja ta wynika ze względu na najwyższą wartość skorygowanego współczynnika determinacji ( $adj. R^2 = 0.215$ ) w porównaniu do reszty przedstawionych modeli. Natomiast patrząc na wartość  $R^2$  możemy zauważyć, że wybrany model wyjaśnia 23.8% wariancji zmiennej objaśnianej *grade*. Zmienne istotne statystycznie oznaczone są poprzez symbol '\*', zgodnie z informacją na końcu tabeli, gdzie p oznacza p-value.

Tabela 2. Proces selekcji zmiennych objaśniających do modelu.

	<i>Dependent variable: grade</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	16.803*** (3.436)	16.855*** (3.427)	16.806*** (3.422)	17.127*** (3.350)	16.923*** (3.334)	15.580*** (3.231)	16.482*** (3.330)
sex	-1.107** (0.456)	-1.098** (0.455)	-1.086** (0.453)	-1.127** (0.444)	-1.092** (0.441)	-1.187*** (0.438)	-1.149*** (0.440)
age	-0.398** (0.183)	-0.404** (0.182)	-0.399** (0.182)	-0.398** (0.182)	-0.390** (0.181)	-0.313* (0.175)	-0.365** (0.181)
studytime	0.593** (0.257)	0.595** (0.256)	0.597** (0.256)	0.590** (0.255)	0.575** (0.254)	0.587** (0.254)	0.556** (0.254)
had_failed	-5.002*** (0.736)	-4.972*** (0.729)	-5.025*** (0.719)	-5.007*** (0.717)	-5.040*** (0.715)	-5.121*** (0.715)	-5.110*** (0.715)
schoolsup	-0.958 (0.652)	-0.987 (0.647)	-1.000 (0.646)	-1.006 (0.645)	-1.025 (0.644)		-0.990 (0.645)
famsup	-0.972** (0.459)	-0.974** (0.455)	-0.918** (0.439)	-0.907** (0.438)	-0.901** (0.437)	-0.944** (0.437)	-0.861** (0.438)
absences	0.082 (0.059)	0.083 (0.059)	0.083 (0.059)	0.085 (0.059)	0.082 (0.059)	0.077 (0.059)	0.001 (0.031)
absencessq	-0.002 (0.001)	-0.002 (0.001)	-0.002 (0.001)	-0.002* (0.001)	-0.002 (0.001)	-0.002 (0.001)	
paid	0.212 (0.451)	0.212 (0.449)					
activities	-0.269 (0.427)	-0.288 (0.423)	-0.294 (0.423)	-0.278 (0.421)			
higher	1.000 (1.047)	0.973 (1.043)	1.023 (1.037)	1.007 (1.035)	0.946 (1.030)	0.941 (1.032)	1.251 (1.015)
freetime	0.097 (0.216)	0.105 (0.215)	0.101 (0.215)				
Msec	0.636 (0.568)	0.558 (0.532)	0.550 (0.531)	0.559 (0.530)	0.574 (0.529)	0.602 (0.530)	0.550 (0.530)
Fsec	-0.293 (0.564)						
Mhigh	1.190* (0.622)	1.126** (0.511)	1.142** (0.509)	1.143** (0.509)	1.120** (0.507)	1.182** (0.507)	1.161** (0.508)
Fhigh	-0.041 (0.640)						
absences:had_failed	0.206*** (0.064)	0.205*** (0.063)	0.207*** (0.063)	0.206*** (0.063)	0.210*** (0.063)	0.213*** (0.063)	0.238*** (0.060)
higher:studytime	0.593** (0.257)	0.595** (0.256)	0.597** (0.256)	0.590** (0.255)	0.575** (0.254)	0.587** (0.254)	0.556** (0.254)
Observations	395	395	395	395	395	395	395
R <sup>2</sup>	0.241	0.240	0.240	0.239	0.238	0.233	0.233
Adjusted R <sup>2</sup>	0.207	0.210	0.212	0.213	0.215	0.211	0.211
Residual Std. Error	4.081 (df=377)	4.072 (df=379)	4.067 (df=380)	4.063 (df=381)	4.060 (df=382)	4.068 (df=383)	4.069 (df=383)
F Statistic	7.037*** (df=17;377)	7.990*** (df=15; 379)	8.562*** (df=14; 380)	9.222*** (df=13; 381)	9.969*** (df=12; 382)	10.603*** (df=11; 383)	10.585*** (df=11; 383)

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Źródło: Opracowanie własne przy pomocy biblioteki stargazer dla języka python

Niektóre ze zmiennych nieistotnych statystycznie w modelu 5 pozostały w nim ze względu na łączną istotność z interakcjami w których się zawierają. W przypadku zmiennej *higher*, mimo iż jest nieistotna statystycznie (p-value = 0.359), jej interakcja ze zmienną *studytime* jest istotna (p-value = 0.024). Podobnie w przypadku zmiennej *absences*, której interakcja ze zmienną *had\_failed* jest istotna statystycznie (p-value = 0.001). Dodatkowo istotne wnioski przynosi analiza modelu 7, w którym usunięto zmienną *absencessq* (p-value = 0.102). Spowodowało znaczący wzrost p-value dla zmiennej *absences* (z poziomu 0.160 do 0.971), co wskazuje na zasadność wprowadzenia kwadratu tej zmiennej do modelu.

### 3.2. Interpretacja oszacowanych wartości

Ostateczne oszacowane równanie liniowe dla zmiennej *grade* według danych obliczonych dla modelu 5 wygląda następująco:

$$\begin{aligned}\widehat{grade}_i = & 16.923 - 1.092 * sex_i - 0.390 * age_i + 0.575 * studytime_i - 5.040 \\ & * had\_failed_i - 1.025 * schoolsup_i - 0.901 * famsup_i \\ & + 0.082 * absences_i - 0.002 * absencessq_i + 0.946 * higher_i + 0.574 \\ & * Msec_i + 1.120 * Mhigh_i + 0.210 * absences_i * had\_failed_i \\ & + 0.575 * higher_i * studytime_i\end{aligned}$$

Na podstawie oszacowań zauważyć można, iż uczennice mają ocenę niższą o 1.092 od uczniów. Wzrost wieku również wiąże się ze spadkiem oceny, dokładnie o 0.390 przy zwiększeniu się wieku o rok. Natomiast wnioskując z wartości oszacowania przy zmiennej *studytime* uczniowie, którzy nie chcą iść na studia, uczący się więcej niż 5 godzin tygodniowo mogą liczyć na ocenę końcową wyższą o 0.575 niż uczniowie, którzy na naukę przeznaczają mniej czasu. Co ciekawe, zarówno wsparcie szkoły jak i rodziców wykazało w modelu negatywny wpływ na końcową ocenę z matematyki. Uczniowie otrzymujący wsparcie szkoły lub rodziców mogli spodziewać się oceny niższej o, odpowiednio, 1.025 lub 0.901. Natomiast wpływ dodatkowej nieobecności ucznia na ocenę końcową zależał od ich wcześniejszej ilości, zakładając, że uczeń zaliczył również wszystkie przedmioty. W takiej sytuacji optymalne dla ucznia było posiadanie 20.5 nieobecności, gdyż taka ich ilość miała maksymalny pozytywny wpływ na ocenę, wynoszący 0.841. Przekroczenie tej liczby powodowało stopniowe zmniejszenie się pozytywnego wpływu na ocenę końcową, aż do osiągnięcia 41 nieobecności,

kiedy to wpływ na ocenę zaczynał być negatywny. W przypadku uczniów deklarujących chęć pójścia na studia wpływ tej decyzji na ocenę zależny był od czasu poświęcanego przez tych uczniów na naukę. W tym przypadku uczniowie uczący się powyżej 5 godzin tygodniowo mogli spodziewać się oceny wyższej o, w sumie, 2.096, natomiast dla uczących się w tygodniu przez mniej czasu wzrost ten wyniósł 0.946. Najsilniejszy negatywny wpływ na ocenę ucznia miało wcześniejsze niezaliczenie dowolnego przedmiotu. W takim przypadku można było spodziewać się oceny niższej o 5.040, jeśli uczeń nie posiadał nieobecności w szkole. Natomiast, co ciekawe, jeśli nie zaliczył wcześniej przedmiotu, dodatkowe nieobecności mogły działać na jego korzyść. Nieobecności działały pozytywnie na ocenę takiego ucznia aż do wypracowania 73 nieobecności, osiągając w takim przypadku dodatni wpływ na nią na poziomie 5.618, następnie pozytywny efekt zmniejszał się. Natomiast negatywne efekty niezaliczenia przedmiotu niwelowane były już w przypadku posiadania 20 nieobecności. Istotny na ostateczną ocenę ucznia był również wpływ poziomu edukacji matki. W przypadku wykształcenia średniego ocena zwiększała się o 0.574, natomiast dla wykształcenia wyższego wzrost ten zwiększał się ponad dwukrotnie – do poziomu 1.120.

### 3.3. Testy diagnostyczne modelu, obserwacje odstające oraz współliniowość

W celu weryfikacji oszacowań modelu przeprowadzone zostały testy, których zadaniem było zweryfikowanie założeń Klasycznego Modelu Regresji Liniowej (KMRL). Spełnienie tych założeń wymagane jest, aby model był najlepszym nieobciążonym estymatorem parametrów modelu. Wyniki tej analizy przedstawione zostały w tabeli 3.

Tabela 3. Wyniki testów sprawdzających założenia KMRL.

Nazwa testu	Testowane założenie KMRL	Hipoteza zerowa	Statystyka testowa	p-value	Decyzja weryfikacyjna
test Breuscha-Pagana	Homoskedastyczność reszt	Homoskedastyczność jest obecna	24.650	0.026	Odrzucamy $H_0$
test White'a	Homoskedastyczność reszt	Homoskedastyczność jest obecna	99.787	0.029	Odrzucamy $H_0$
test RESET	Związek pomiędzy $y$ a $X$ jest liniowy	Liniowa forma funkcyjna	0.993	0.609	Brak podstaw do odrzucenia $H_0$
test Jarque-Bera	Składnik losowy ma rozkład normalny	Zmienne mają rozkład normalny	35.447	2.008e-08	Odrzucamy $H_0$
test Breuscha-Godfrey'a	$Cov(\varepsilon_i, \varepsilon_j) = 0$	Brak autokorelacji	14.922	0.135	Brak podstaw do odrzucenia $H_0$

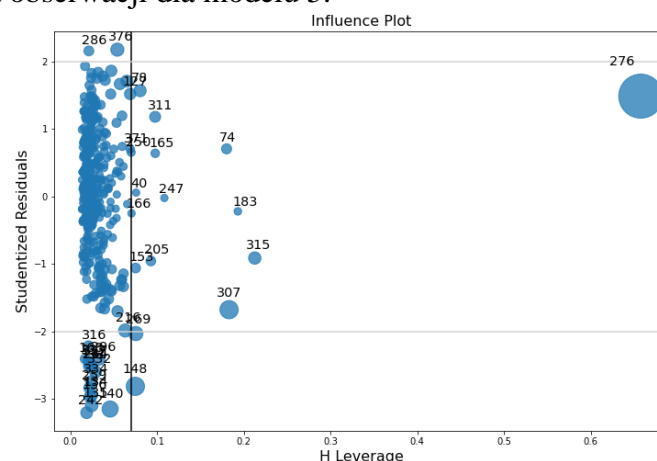
Źródło: Opracowanie własne.

Jak można zauważyć niespełnione zostały założenie o rozkładzie normalnym składnika losowego oraz o homoskedastyczności, zarówno w przypadku testu Breuscha-Pagana jak

i testu White'a. Jednak test sprawdzający założenie o rozkładzie normalnym składnika losowego, ze względu na ilość obserwacji, wynoszącą 395, nie jest wymagany. Dla dużej próby badawczej, wnoszącej ponad 100 obserwacji, przyjmuje się, iż założenie to nie ma istotnego wpływu na ocenę modelu, bez względu na wynik testu Jarque-Bera. Również nie jest wymagana weryfikacja założenia o braku autokorelacji, poprzez test Breuscha-Godfreya, ponieważ w przedstawionym modelu nie występuje element czasowy. Natomiast ze względu na odrzucenie hipotezy zerowej w testach dotyczących o homoskedastyczności wymagane jest zastosowanie macierzy odpornej w modelu w celu otrzymania poprawnych wartości błędów standardowych oraz wynikającym z nich wartości p-value. Spowodowane jest to faktem, że w przypadku niespełnienia założenia o homoskedastyczności wartości błędów standardowych w modelu będą obciążone. W tabeli 4 przedstawione zostało porównanie modelu 5 wyjściowego (1) oraz z zastosowaną macierzą odporną (2). Jak można zauważyć wartości błędów standardowych uległy zmianie. Dodatkowo zmienne *schoolsup* oraz *absencessq* po zastosowaniu macierzy odpornej stały się zmiennymi istotnymi statystycznie. Kolejne z założeń KMRL, mówiące o wartości oczekiwanej reszt równej 0, jest spełnione, ponieważ w modelu występuje stała. W przypadku założenia o nielosowości zmiennych objaśniających nie jest ono możliwe do przetestowania, więc uznaje się je za spełnione.

Natomiast do analizy obserwacji odstających wykorzystany został wykres widoczny na rysunku 3. Kolorem czarnym zaznaczona została wartość dźwigni H, natomiast na szaro zaznaczone zostały wartości graniczne reszt, odpowiednio 2 oraz -2. Na odległość Cooka danej obserwacji wskazuje rozmiar punktu na wykresie. Jak można zauważyć, w danych zastosowanych do budowy modelu nie występują obserwacje, które spełniałyby wszystkie z założeń kwalifikujących je jako obserwacje odstające.

Rysunek 3. Analiza obserwacji dla modelu 5.



Źródło: Opracowanie własne przy wykorzystaniu biblioteki statsmodels dla języka python

Tabela 4. Porównanie modelu 5 przed i po zastosowaniu macierzy odpornej.

	<i>Dependent variable: grade</i>	
	(1)	(2)
Intercept	16.923*** (3.334)	16.923*** (3.434)
sex	-1.092** (0.441)	-1.092*** (0.419)
age	-0.390** (0.181)	-0.390** (0.183)
studytime	0.575** (0.254)	0.575** (0.264)
had_failed	-5.040*** (0.715)	-5.040*** (0.823)
schoolsup	-1.025 (0.644)	-1.025* (0.585)
famsup	-0.901** (0.437)	-0.901** (0.429)
absences	0.082 (0.059)	0.082 (0.061)
absencessq	-0.002 (0.001)	-0.002* (0.001)
higher	0.946 (1.030)	0.946 (1.230)
Msec	0.574 (0.529)	0.574 (0.517)
Mhigh	1.120** (0.507)	1.120** (0.489)
absences:had_failed	0.210*** (0.063)	0.210*** (0.068)
higher:studytime	0.575** (0.254)	0.575** (0.264)
Observations	395	395
R <sup>2</sup>	0.238	0.238
Adjusted R <sup>2</sup>	0.215	0.215
Residual Std. Error	4.060 (df=382)	4.060 (df=382)
F Statistic	9.969*** (df=12; 382)	9.105*** (df=12; 382)

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Źródło: Opracowanie własne przy pomocy biblioteki stargazer dla języka python

W badaniu autokorelacji zmiennych przy pomocy współczynnika współliniowości VIF, przedstawionego w tabeli 5, nie zaobserwowano wysokich jego wartości, z wyjątkiem interakcji *studytime\*higher* i zmiennej *studytime*, dla których wartość VIF wyniosła nieskończoność, co wskazuje na idealną korelację. Z tego powodu zmienne te mogą odpowiadać za zwiększanie wariancji błędu statystycznego. Jednak wartość ta jest zaskakująca, ponieważ wskazuje, że zmienne te przyjmują jednakowe wartości. Źródłem tego



problemu okazał się fakt, iż w użytej do budowy modelu bazie danych nie zaistniał przypadek, w którym uczeń uczyłby się powyżej 5 godzin w tygodniu (*studytime* = 1) i nie planował jednocześnie kontynuować nauki na studiach wyższych (*higher* = 0). Spowodowało to zatem rozdzielenie wpływu wartości *studytime* na dwie zmienne. Można zatem usunąć tę interakcję z modelu a takie działanie spowoduje podwojenie się wartości oszacowania przy *studytime* oraz niewielką poprawę jego wartości p-value, i nie będzie miało innego wpływu na model. Natomiast w przypadku pozostałych zmiennych wartość wskaźnika utrzymywała się na niskim poziomie, nieprzekraczającym 10, co sugeruje brak istotnej autokorelacji między nimi.

Tabela 5. Wartości współczynnika współliniowości VIF.

	VIF	Tolerance
sex	1.160864	0.861427
age	1.279294	0.781681
studytime	inf	0.000000
had_failed	2.032220	0.492073
schoolsup	1.117825	0.894594
famsup	1.088083	0.919048
absences	5.244192	0.190687
absencessq	4.206742	0.237714
higher	1.222169	0.818217
Msec	1.260165	0.793547
Mhigh	1.366799	0.731637
absences:had_failed	2.504566	0.399271
higher:studytime	inf	0.000000

Źródło: Opracowanie własne

## WNIOSKI

Przeprowadzona analiza pozwala na lepsze zrozumienie czynników wpływających na oceny uczniów w szkole średniej. Zbudowany w pracy model wskazuje na istotność takich zmiennych jak płeć czy wiek, ale również obecność na zajęciach czy wcześniejsze niezaliczenia przedmiotów. Ciekawym wnioskiem z analizy jest negatywny wpływ wsparcia szkoły na oceny ucznia. Przybliżona stała przy zmiennej *schoolsup* przyjmuje wartość -1.025. Wartość ta wydaje się nieintuicyjna, jednak może być wiele wyjaśnień takiej zależności, i wymagałoby to głębszej analizy w celu weryfikacji jej przyczyn.

Niestety przedstawiona analiza nie pozwoliła na pełne potwierdzenie przyjętej w tej pracy głównej hipotezy badawczej. Wartości przybliżeń przedstawionego modelu wskazują na pozytywną zależność jedynie edukacji matki na oceny dziecka. W przypadku edukacji ojca zmienne ją określające zostały usunięte już w momencie przejścia do modelu 2, co wynikało z wysokich wartości p-value dla zmiennych *Fsec* oraz *Fhigh*. Taki wniosek, częściowo sprzeczny z przedstawioną literaturą naukową, może być spowodowany ograniczeniami wynikającymi z dostępnych danych.

W przypadku pierwszej dodatkowej hipotezy badawczej, mówiącej o wpływie wsparcia rodziców na wyniki dziecka w szkole, model wskazuje na istotną statystycznie negatywną zależność wsparcia rodziców na ostateczną ocenę dziecka. Wynik ten wydaje się być sprzeczny z literaturą naukową, jednak może ilustrować jeden z problemów wykorzystanej w tej pracy bazy danych. Zmienna *famsup* przekazuje bardzo ogólne informacje. Natomiast badania naukowe przytoczone w rozdziale 1 wskazują na istotność rozdzielenia poszczególnych aspektów wsparcia rodziców. Wynik analizy może zatem być spowodowany brakiem odpowiedniego rozdzielenia poszczególnych składowych ogólnie rozumianego wsparcia rodziców w nauce dziecka.

Ostatnia z postawionych hipotez także nie została potwierdzona. Zmienne *paid* oraz *activities* okazały się nieistotne statystycznie i zostały usunięte z ostatecznego modelu. Można zatem wyciągnąć wniosek mówiący, że dodatkowe zajęcia pozalekcyjne nie mają znaczącego wpływu na oceny dziecka w szkole. Jednak wniosek ten jest sprzeczny z literaturą przytoczoną wcześniej. Problem ten także może być spowodowany ograniczeniami modelu.

## OGRANICZENIA MODELU

Niestety przedstawiony w tej pracy model ekonometryczny posiada pewne ograniczenia. Głównym z nich jest zastosowana do jego zbudowania baza danych. Zawiera ona informacje dotyczące jedynie dwóch szkół średnich z jednego państwa europejskiego. W celu otrzymania dokładniejszych oszacowań pozwalających na wysunięcie bardziej pewnych wniosków wskazane byłoby wykorzystanie danych obejmujących szerszą grupę badawczą. Zmiana ta znacząco przyczyniłaby się do ulepszenia przedstawionego modelu, który w takim przypadku mógłby przedstawiać przybliżenia możliwe do zastosowania dla szerszej populacji.

Dodatkowym ograniczeniem jest również charakter zebranych w bazie danych informacji. Wiele z obecnych w niej zmiennych było mierzonych subiektywnie, uogólniając takie wartości jak czas przeznaczany na naukę czy ilość wolnego czasu do przedziałów czasowych czy deklaracji zebranych przy pomocy ankiet. Także ograniczenie danych dotyczących ocen do jedynie dwóch przedmiotów, języka portugalskiego oraz matematyki znacząco utrudnia przeprowadzenie analizy. Bardziej miarodajną zmienną byłaby w tym przypadku średnia ocen danego ucznia.

Również przypadek interakcji zmiennych *higher* oraz *studytime* wskazuje na ograniczenia związane z wykorzystaną bazą danych. W przypadku zastosowania większej bazy danych, zawierającej więcej obserwacji z wielu regionów czy państw, zbudowany model znacząco zyskałby na dokładności, ponieważ jego oszacowania opierałby się na danych lepiej reprezentujących ogół populacji.

Eliminacja wspomnianych ograniczeń znacząco wpłynęłaby na jakość modelu, który został zaproponowany. W takim przypadku wnioski wyciągnięte z badania mogłyby przynieść większą wartość merytoryczną, pozwalającą na dokładniejsze porównanie otrzymanych wyników z przytoczonymi przesłankami teoretycznymi.

## BIBLIOGRAFIA

- Dubow Eric F., Boxer Paul, Huesmann L. Rowell. 2009. Long-term Effects of Parents' Education on Children's Educational and Occupational Success: Mediation by Family Interactions, Child Aggression, and Teenage Aspirations. *Merrill Palmer Q (Wayne State Univ Press)* T. 55, nr. 3, s. 224-249.
- Furda Mark, Shuleski Michael. 2019. The Impact of Extracurriculurs on Academic Performance and School Perception. *Excellence in Education Journal* T. 8, nr 1, s. 64-90.
- Gonzalez-DeHass Alyssa R., Willems Patricia P., Doan Holbein Marie F. 2005. Examining the Relationship Between Parental Involvement and Student Motivation. *Educational Psychology Review* T. 17, s. 99-123.
- Jeynes William H. 2007. The Relationship Between Parental Involvement and Urban Secondary School Student Academic Achievement. *Urban Education* T. 42, nr. 1 s. 82-110.
- Raudenská Petra, Hamplová Dana. 2022. The Effect of Parents' Education and Income on Children's School Performance: the Mediating Role of the Family Environment and Children's Characteristics, and Gender Differences. *Polish Sociological Review* T. 218, nr. 2, s. 247-272.

## ZESTAWIENIE SPISÓW

### Spis tabel

Tabela 1.	Zmienne objaśniające wykorzystane w modelu.	7
Tabela 2.	Proces selekcji zmiennych objaśniających do modelu.	11
Tabela 3.	Wyniki testów sprawdzających założenia KMRL.	13
Tabela 4.	Porównanie modelu 5 przed i po zastosowaniu macierzy odpornej.	15
Tabela 5.	Wartości współczynnika współliniowości VIF.	16

### Spis rysunków

Rysunek 1.	Wykres zależności zmiennych absences oraz grade.	8
Rysunek 2.	Korelacja zmiennych wykorzystanych w modelu w formie heatmapy.	9
Rysunek 3.	Analiza obserwacji dla modelu 5.	14