

# PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

# RESEARCH PAPERS

of Wrocław University of Economics

Nr 328

**Taksonomia 23**

**Klasyfikacja i analiza danych –  
teoria i zastosowania**

Redaktorzy naukowci

Krzysztof Jajuga, Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
Wrocław 2014

Redaktor Wydawnictwa: Barbara Majewska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

[www.ibuk.pl](http://www.ibuk.pl), [www.ebscohost.com](http://www.ebscohost.com),

w Dolnośląskiej Bibliotece Cyfrowej [www.dbc.wroc.pl](http://www.dbc.wroc.pl),

The Central and Eastern European Online Library [www.ceeol.com](http://www.ceeol.com),

a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon

[http://kangur.uek.krakow.pl/bazy\\_ae/bazekon/nowy/index.php](http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php)

Informacje o naborze artykułów i zasadach recenzowania znajdują się  
na stronie internetowej Wydawnictwa

[www.wydawnictwo.ue.wroc.pl](http://www.wydawnictwo.ue.wroc.pl)

Tytuł dofinansowany ze środków Narodowego Banku Polskiego  
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Kopiowanie i powielanie w jakiegokolwiek formie  
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu  
Wrocław 2014

**ISSN 1899-3192** (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)

**ISSN 1505-9332** (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

## Spis treści

<b>Wstęp</b> .....	11
<b>Małgorzata Rószkiewicz</b> , Wykorzystanie metaanalizy w budowaniu modelu pomiarowego w przypadku braku niezmienniczości zasad pomiaru na przykładzie pomiaru zadowolenia z życia.....	13
<b>Elżbieta Sobczak</b> , Harmonijność inteligentnego rozwoju regionów Unii Europejskiej .....	21
<b>Ewa Roszkowska, Renata Karwowska</b> , Analiza porównawcza województw Polski ze względu na poziom zrównoważonego rozwoju w roku 2010.....	30
<b>Tadeusz Kufel, Magdalena Osińska, Marcin Błażejowski, Paweł Kufel</b> , Analiza porównawcza wybranych filtrów w analizie synchronizacji cyklu koniunkturalnego.....	41
<b>Marcin Salamaga</b> , Próba konstrukcji tablic „wymierania scenicznego” spektakli operowych na przykładzie Metropolitan Opera.....	51
<b>Iwona Foryś</b> , Wykorzystanie analizy dyskryminacyjnej do typowania rynków podobnych w procesie wyceny nieruchomości niemieszkalnych .....	59
<b>Jerzy Korzeniewski</b> , Selekcja zmiennych w klasyfikacji – propozycja algorytmu .....	69
<b>Sabina Denkowska</b> , Testowanie wielokrotne przy weryfikacji wieloczynnikowych modeli proporcjonalnego hazardu Coxa.....	76
<b>Ewa Chodakowska</b> , Teoria równań strukturalnych w klasyfikacji zmiennych jawnych i ukrytych według charakteru ich wzajemnych oddziaływań .....	85
<b>Iwona Konarzewska</b> , Model PCA dla rynku akcji – studium przypadku .....	94
<b>Katarzyna Wójcik, Janusz Tuchowski</b> , Dobór optymalnego zestawu słów istotnych w opiniach konsumentów na potrzeby ich automatycznej analizy	106
<b>Aleksandra Łuczak</b> , Zastosowanie metody AHP-LP do oceny ważności determinant rozwoju społeczno-gospodarczego w jednostkach administracyjnych .....	116
<b>Aleksandra Witkowska, Marek Witkowski</b> , Klasyfikacja pozycyjna banków spółdzielczych według stanu ich kondycji finansowej w ujęciu dynamicznym .....	126
<b>Adam Depta</b> , Zastosowanie analizy korespondencji do oceny jakości życia ludności na podstawie kwestionariusza SF-36v2 .....	135
<b>Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Marek Marciniak, Jerzy Kołodziej</b> , Indukcja reguł dla danych niekompletnych i niezbalansowanych: modele klasyfikatorów i próba ich zastosowania do predykcji ryzyka operacyjnego w torakochirurgii .....	146

<b>Małgorzata Misztal</b> , Wybrane metody oceny jakości klasyfikatorów – przegląd i przykłady zastosowań.....	156
<b>Anna M. Olszewska</b> , Wykorzystanie wybranych metod taksonomicznych do oceny potencjału innowacyjnego województw .....	167
<b>Iwona Bąk</b> , Porównanie jakości grupowań powiatów województwa zachodniopomorskiego pod względem atrakcyjności turystycznej.....	177
<b>Agnieszka Kozera, Joanna Stanisławska, Romana Głowicka-Wołoszyn</b> , Segmentacja gospodarstw domowych według wydatków na turystykę zorganizowaną.....	186
<b>Agnieszka Wałęga</b> , Podejście syntetyczne w analizie spójności ekonomicznej gospodarstw domowych.....	196
<b>Joanna Banaś, Małgorzata Machowska-Szewczyk, Bożena Mroczek</b> , Zastosowanie analizy korespondencji do badania wpływu elektrowni wiatrowych na jakość życia ludności .....	205
<b>Joanna Banaś, Krzysztof Małecki</b> , Klasyfikacja punktów pomiarów ankietowych kierowców na granicy Szczecina z wykorzystaniem zmiennych symbolicznych.....	214
<b>Aneta Becker</b> , Wykorzystanie informacji granularnej w analizie wymagań rynku pracy.....	222
<b>Katarzyna Cheba, Joanna Holub-Iwan</b> , Wykorzystanie analizy korespondencji w segmentacji rynku usług medycznych.....	230
<b>Adam Depta, Iwona Staniec</b> , Identyfikacja czynników decydujących o jakości życia studentów łódzkich uczelni.....	238
<b>Katarzyna Dębowska, Jarosław Kilon</b> , Reguły asocjacyjne w analizie wyników badań metodą Delphi.....	247
<b>Anna Domagała</b> , O wykorzystaniu analizy głównych składowych w metodzie <i>Data Envelopment Analysis</i> .....	254
<b>Alicja Grześkowiak</b> , Analiza wykluczenia cyfrowego w Polsce w ujęciu indywidualnym i regionalnym.....	264
<b>Anna M. Olszewska, Anna Gryko-Nikitin</b> , Pomiar postrzegania jakości kształcenia uczelni wyższej na danych porządkowych z wykorzystaniem środowiska R.....	273
<b>Karolina Paradysz</b> , Hierarchiczna metoda grupowania powiatów jako podejście benchmarkowe w ocenie bezrobocia według BAEL-u w wybranych typach małych obszarów .....	282
<b>Radosław Pietrzyk</b> , Porównanie metod pomiaru efektywności zarządzania portfelami funduszy inwestycyjnych.....	290
<b>Agnieszka Przedborska, Małgorzata Misztal</b> , Wybrane metody statystyki wielowymiarowej w ocenie skuteczności terapeutycznej głębokiej stymulacji elektromagnetycznej u pacjentów z chorobą zwyrodnieniową stawów.....	299

<b>Wojciech Roszka, Marcin Szymkowiak</b> , Podejście kalibracyjne w statystycznej integracji danych .....	308
<b>Iwona Skrodzka</b> , Zastosowanie wybranych metod klasyfikacji do analizy kapitału ludzkiego krajów Unii Europejskiej .....	316
<b>Agnieszka Stanimir</b> , Wielowymiarowa analiza czynników sprzyjających włączeniu społecznemu .....	326
<b>Dorota Strózik, Tomasz Strózik</b> , Przestrzenne zróżnicowanie poziomu życia w województwie wielkopolskim.....	334
<b>Izabela Szamrej-Baran</b> , Identyfikacja przyczyn ubóstwa energetycznego w Polsce przy wykorzystaniu modelowania miękkiego.....	343
<b>Janusz Tuchowski, Katarzyna Wójcik</b> , Klasyfikacja obiektów w systemie Krajowych Ram Kwalifikacji opisanych za pomocą ontologii .....	353
<b>Aleksandra Matuszewska-Janica</b> , Grupowanie krajów Unii Europejskiej ze względu na poziom feminizacji sektorów gospodarczych .....	361
<b>Monika Rozkrut, Dominik Rozkrut</b> , Identyfikacja strategii innowacyjnych przedsiębiorstw usługowych w Polsce .....	369

## Summaries

<b>Małgorzata Rószkiewicz</b> , The use of meta-analysis in building the measurement model in case of the absence of measurement invariance on the example of measuring of life satisfaction.....	20
<b>Elżbieta Sobczak</b> , Harmonious smart growth of European Union regions.....	29
<b>Ewa Roszkowska, Renata Karwowska</b> , The comparative analysis of Polish voivodeships with respect to sustainable development in 2010.....	40
<b>Tadeusz Kufel, Magdalena Osińska, Marcin Błażejowski, Paweł Kufel</b> , Comparative analysis of chosen filters in business cycles analysis .....	50
<b>Marcin Salamaga</b> , The attempt of construction of the life tables for opera works on the example of the Metropolitan Opera .....	58
<b>Iwona Foryś</b> , Using discriminant analysis to select similar markets in non-residential property valuation process.....	68
<b>Jerzy Korzeniewski</b> , Variable selection in classification – algorithm proposal .....	75
<b>Sabina Denkowska</b> , Multiple testing in the verification process of multifactorial Cox proportional hazards models .....	84
<b>Ewa Chodakowska</b> , The theory of structural equations modelling in the classification of observed variables and latent constructs according to the character of their relationship.....	93
<b>Iwona Konarzewska</b> , Modelling stock market by PCA factor model – case study .....	105

<b>Katarzyna Wójcik, Janusz Tuchowski</b> , Selection of the optimal set of relevant words in consumers opinions in the context of the opinion mining ..	115
<b>Aleksandra Łuczak</b> , Application of AHP-LP to the evaluation of importance of determinants of socio-economic development in the administrative units .....	125
<b>Aleksandra Witkowska, Marek Witkowski</b> , A dynamic approach to the ranking of cooperative banks by their financial condition .....	134
<b>Adam Depta</b> , Application of correspondence analysis for the measurement of quality of life – questionnaire SF-36v2 based research .....	145
<b>Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Marek Marciniak, Jerzy Kołodziej</b> , Classification rules extraction for missing and imbalance data: models of classifiers and initial results in the rules-based thoracic surgery risk prediction.....	155
<b>Małgorzata Misztal</b> , Selected methods for assessing the performance of classifiers – an overview and examples of applications.....	166
<b>Anna M. Olszewska</b> , The application of selected quantitative methods to the evaluation of voivodeship innovation level potential.....	176
<b>Iwona Bąk</b> , The comparison of the quality of groupings of poviats of West Pomeranian Voivodeship in terms of tourism attractiveness .....	185
<b>Agnieszka Kozera, Joanna Stanisławska, Romana Głowicka-Wołoszyn</b> , Household segmentation with respect to the expenditure on organized tourism.....	195
<b>Agnieszka Wałęga</b> , Synthetic approach in the analysis of economic coherence of households .....	204
<b>Joanna Banaś, Małgorzata Machowska-Szewczyk, Bożena Mroczek</b> , Using the correspondence analysis to examine the impact of wind turbines on the quality of life.....	213
<b>Joanna Banaś, Krzysztof Małecki</b> , Classification of measurement survey points of drivers on the boundary of Szczecin using symbolic variables...	221
<b>Aneta Becker</b> , The use granular information in the analysis of the requirements of the labor market.....	229
<b>Katarzyna Cheba, Joanna Hołub-Iwan</b> , The application of the correspondence analysis of patients segmentation on the medical service market .....	237
<b>Adam Depta, Iwona Staniec</b> , Identification of the factors that determine the quality of students life at universities in Lodz.....	246
<b>Katarzyna Dębowska, Jarosław Kilon</b> , Association rules in the analysis of research results the Delphi method .....	253
<b>Anna Domagała</b> , About using Principal Component Analysis in Data Envelopment Analysis .....	263
<b>Alicja Grześkowiak</b> , Analysis of the digital divide in Poland at the individual and regional level .....	272

<b>Anna M. Olszewska, Anna Gryko-Nikitin</b> , Assessment of perception of quality of teaching at an institution of higher learning based on the ordinal data with the utilization of R environment.....	281
<b>Karolina Paradysz</b> , The hierarchical method of grouping poviats as a benchmark approach in the assessment of unemployment by BAEL in selected types of small areas .....	289
<b>Radosław Pietrzyk</b> , Comparison of methods of measuring the performance of investment funds portfolios.....	298
<b>Agnieszka Przedborska, Małgorzata Misztal</b> , Selected multivariate statistical analysis methods in the evaluation of efficacy of deep electromagnetic stimulation in patients with degenerative joint disease .....	307
<b>Wojciech Roszka, Marcin Szymkowiak</b> , A calibration approach in statistical data integration .....	315
<b>Iwona Skrodzka</b> , Application of some methods of classification to the analysis of human capital in the European Union.....	325
<b>Agnieszka Stanimir</b> , Multivariate analysis of social inclusion factors.....	333
<b>Dorota Strózik, Tomasz Strózik</b> , Spatial differentiation of the standard of living in Great Poland Voivodeship .....	342
<b>Izabela Szamrej-Baran</b> , Identification of fuel poverty causes in Poland using soft modelling .....	352
<b>Janusz Tuchowski, Katarzyna Wójcik</b> , Classification of objects in the National Classification Framework described by the ontology.....	360
<b>Aleksandra Matuszewska-Janica</b> , Clustering of European Union states taking into consideration the levels of feminization of economic sectors..	368
<b>Monika Rozkrut, Dominik Rozkrut</b> , Identification of service sector innovation strategies in Poland.....	379

**Ewa Chodakowska**

Politechnika Białostocka

---

## **TEORIA RÓWNAŃ STRUKTURALNYCH W KLASYFIKACJI ZMIENNYCH JAWNYCH I UKRYTYCH WEDŁUG CHARAKTERU ICH WZAJEMNYCH ODDZIAŁYWAŃ**

---

**Streszczenie:** Artykuł przedstawia próbę analizy i klasyfikacji zmiennych determinujących efektywność uczenia w wybranych szkołach gimnazjalnych w Polsce. Za pomocą SEM dokonano weryfikacji hipotetycznych relacji między zmiennymi opisującymi wyniki egzaminacyjne a wybranymi charakterystykami szkoły oraz jej uczniów. Na efektywność w badanych szkołach bezpośrednio oddziałuje tylko czynnik otoczenia reprezentowany przez wybrane charakterystyki uczniów. Czynnik szkolny reprezentujący warunki nauczania bezpośrednio nie wpływa na wyniki egzaminacyjne, ale jest skorelowany z czynnikiem otoczenia.

**Słowa kluczowe:** SEM, CFA, klasyfikacja, efektywność nauczania.

### **1. Wstęp**

Model równań strukturalnych (SEM – *structural equation model*) jest bardzo ogólnym modelem statystycznej analizy danych. Regresję, analizę czynnikową czy metodę ANOVA można traktować jako szczególne przypadki SEM [Garson 2012, s. 13].

SEM jest popularnym narzędziem analizy danych, szczególnie w naukach społecznych. Na świecie z powodzeniem stosowany jest w obszarze edukacji do określenia determinantów osiągnięć szkolnych [Dufur i in. 2013; Rosandera i in. 2011] czy akademickich [McKenzie i Gow 2004; Kyoung i in. 2013].

W artykule za pomocą SEM podjęto próbę analizy i klasyfikacji zmiennych determinujących efektywność uczenia w wybranych szkołach gimnazjalnych w Polsce. Na podstawie danych empirycznych pochodzących z białostockich gimnazjów za pomocą SEM zweryfikowano powstałe na bazie rozważań teoretycznych hipotetyczne relacje między zmiennymi opisującymi wyniki egzaminacyjne a wybranymi charakterystykami szkoły oraz jej uczniów.



## 2. Modele równań strukturalnych

SEM pozwala na specyfikację i test liniowej relacji między zmiennymi obserwowanymi i ukrytymi. Ogólny model SEM składa się z podmodelu strukturalnego oraz podmodelu pomiarowego [Sagan 2003b, s. 87]. Podmodel strukturalny (1) określa wzajemne relacje strukturalne między zmiennymi ukrytymi [Konarski 2009, s. 250]:

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}. \quad (1)$$

Podmodel pomiarowy (2), (3) pozwala na konstrukcję czynników ukrytych oraz na określenie wielkości błędu pomiarowego [Konarski 2009, s. 248; Sagan 2003a, s. 84]:

$$\mathbf{x} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}, \quad (2)$$

$$\mathbf{y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon}. \quad (3)$$

Oznaczenia:

- $\boldsymbol{\eta}$  – endogeniczne czynniki ukryte ( $m \times 1$ ),
- $\boldsymbol{\xi}$  – egzogeniczne czynniki ukryte ( $n \times 1$ ),
- $\boldsymbol{\zeta}$  – reszty modelu ( $m \times 1$ ),
- $\mathbf{y}$  – endogeniczne zmienne obserwowane dla  $\boldsymbol{\eta}$  ( $p \times 1$ ),
- $\mathbf{x}$  – egzogeniczne zmienne obserwowane dla  $\boldsymbol{\xi}$  ( $q \times 1$ ),
- $\boldsymbol{\epsilon}$  – błąd pomiarowy w  $\mathbf{y}$  ( $p \times 1$ ),
- $\boldsymbol{\delta}$  – błąd pomiarowy w  $\mathbf{x}$  ( $q \times 1$ ),
- $\mathbf{B}$  – współczynniki strukturalne  $\boldsymbol{\eta}$  na  $\boldsymbol{\eta}$  ( $m \times m$ ),
- $\boldsymbol{\Gamma}$  – współczynniki strukturalne  $\boldsymbol{\xi}$  na  $\boldsymbol{\eta}$  ( $m \times n$ ),
- $\boldsymbol{\Lambda}_y$  – ładunki czynnikowe  $\boldsymbol{\eta}$  w  $\mathbf{y}$  ( $p \times m$ ),
- $\boldsymbol{\Lambda}_x$  – ładunki czynnikowe  $\boldsymbol{\xi}$  w  $\mathbf{x}$  ( $q \times n$ ).

Założenia:

$E(\boldsymbol{\epsilon}) = E(\boldsymbol{\delta}) = E(\boldsymbol{\zeta}) = E(\boldsymbol{\xi}) = E(\boldsymbol{\eta}) = 0$ ;  $(\mathbf{I} - \mathbf{B})^{-1}$  istnieje;  $\boldsymbol{\zeta}$  nie jest skorelowane z  $\boldsymbol{\xi}$ ;  $\boldsymbol{\epsilon}$  nie jest skorelowane z  $\boldsymbol{\delta}$ ,  $\boldsymbol{\xi}$ ,  $\boldsymbol{\zeta}$ ,  $\boldsymbol{\eta}$ ;  $\boldsymbol{\delta}$  nie jest skorelowane z  $\boldsymbol{\xi}$ ,  $\boldsymbol{\zeta}$ ,  $\boldsymbol{\eta}$ ,  $\boldsymbol{\epsilon}$ .

Szacowanie parametrów modelu polega na takim ich dobraniu, aby teoretyczna macierz kowariancji w maksymalny sposób reprodukowała zaobserwowaną macierz kowariancji [Konarski 2009, s. 287]:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}). \quad (4)$$

Oznaczenia:

- $\boldsymbol{\Sigma}$  – macierz kowariancji obserwowanych  $p$  zmiennych endogenicznych i  $q$  zmiennych egzogenicznych [ $(p + q) \times (p + q)$ ],
- $\boldsymbol{\Sigma}(\boldsymbol{\theta})$  – macierz kowariancji jako funkcja parametrów modelu.

Opracowanie modelu jest procesem iteracyjnym. Analizę za pomocą SEM rozpoczyna się od specyfikacji modelu najczęściej w formie wykresu ścieżkowego,

który jest następnie wyrażony za pomocą równań strukturalnych. W kolejnym kroku należy oszacować parametry SEM. Następnym etapem jest ocena jakości postulowanego modelu na podstawie dwóch kryteriów: relatywnie małego składnika losowego bez struktury relacyjnej oraz małych różnic między oszacowaną a zaobserwowaną macierzą kowariancji [Konarski 2009, s. 65]. Nie ma jednej uniwersalnej normy dopasowania modelu. Zaleca się oparcie oceny na więcej niż jednej mierze dopasowania [Konarski 2009, s. 373]. Ostatni etap to respecyfikacja modelu, jeżeli postulowany model jest nieadekwatny [Konarski 2009, s. 60].

### 3. Analiza

W artykule podjęto próbę zweryfikowania wpływu czynników szkolnych oraz środowiskowych na efektywność uczenia w danej szkole mierzoną jako przyrost wiedzy jej uczniów w stosunku do poprzedniego etapu edukacyjnego –  $y$  [%]. Bez względu na wyniki nauczania określone są przede wszystkim wynikami osiągniętymi na poprzednim etapie nauczania [Chodakowska 2009]. Gdyby przyjąć założenie, że wszystkie szkoły uczą jednakowo dobrze swoich uczniów, to czynniki społeczne i indywidualne cechy ucznia, ujęte pośrednio w jego osiągnięciach szkolnych na poprzednim etapie nauczania, precyzyjnie definiowałyby wyniki na kolejnym szczeblu edukacji. Inaczej mówiąc, zmiana poziomu wiedzy uczniów powinna być w każdej szkole prawie identyczna, a tak nie jest.

W badaniu empirycznym wykorzystano wyniki uczniów wszystkich trzydziestu białostockich gimnazjów z 2005 r., pochodzące z OKE w Łomży oraz z Kuratorium Oświaty w Białymstoku. Budując model, założono *a priori*, że najwyższy przyrost wiedzy powinna uzyskać sprzyjająca szkoła działająca w sprzyjającym otoczeniu. Przyjęto, że sprzyjające otoczenie to uczniowie niesprawiający zasadniczych problemów wychowawczych, nieopuszczający bezzasadnie zajęć, chętnie uczestniczący w różnorodnych zajęciach pozalekcyjnych, pochodzący z rodzin, których stać na współfinansowanie dodatkowych zajęć. Sprzyjająca szkoła to szkoła z indywidualnym podejściem do ucznia, kameralna, z nielicznymi oddziałami, świetnie wyposażona, organizująca dla swoich uczniów wiele zajęć pozalekcyjnych. W tym celu skonstruowano dwie zmienne ukryte: otoczenie ( $\xi_1$ ) i szkołę ( $\xi_2$ ), które reprezentowane były przez 8 obserwowanych standaryzowanych zmiennych egzogenicznych:

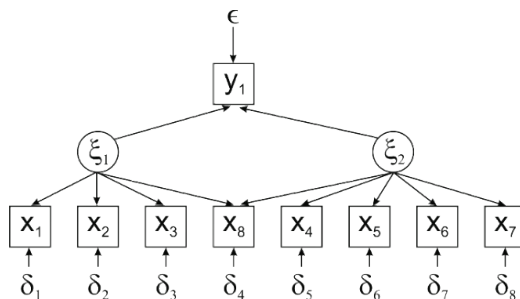
- $x_1$  – liczba uczniów korzystających z posiłków refundowanych na liczbę uczniów ogółem,
- $x_2$  – liczba uczniów powtarzających klasę na liczbę uczniów ogółem,
- $x_3$  – liczba uczniów, którzy opuścili więcej niż 30% zajęć nieusprawiedliwionych, na liczbę uczniów ogółem,
- $x_4$  – liczba uczniów w szkole,
- $x_5$  – średnia liczba uczniów w oddziale,
- $x_6$  – liczba komputerów do użytku uczniów na liczbę uczniów ogółem,

$x_7$  – liczba woluminów w bibliotece na liczbę uczniów ogółem,

$x_8$  – liczba uczniów placówki uczestniczących w zajęciach pozalekcyjnych na liczbę uczniów ogółem.

Klasyczny model SEM zakłada łączną normalność rozkładu zmiennych obserwowanych. W celu spełnienia tego postulatu do analizowanych danych wykorzystano transformacje potęgowe Boxa-Coxa.

Do celów określenia wpływu czynników środowiskowych i szkolnych zaproponowano model SEM ze zmiennymi ukrytymi przedstawiony na rysunku 1. W podmodelu strukturalnym występuje jedna zmienna obserwowana – przyrost wiedzy ( $y$ ) i dwie nieobserwowane: otoczenie ( $\xi_1$ ) i szkoła ( $\xi_2$ ). Część pomiarową modelu reprezentuje model confirmacyjnej analizy czynnikowej (CFA) dla ośmiu wskaźników zwrotnych ( $x_1, x_2, \dots, x_8$ ) dla czynników ukrytych. Ponieważ zmienne ukryte nie wyjaśniają całej zmienności swoich wskaźników cząstkowych przy każ-



Rys. 1. Model SEM

Źródło: opracowanie własne.

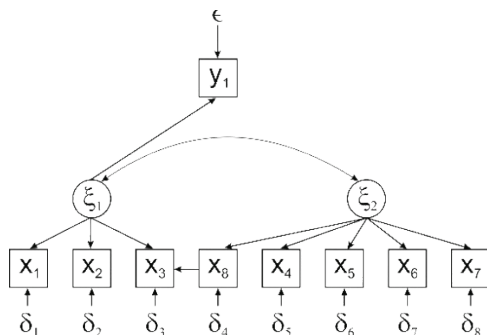
Tabela 1. Oszacowania wybranego modelu SEM

Ścieżka	Współczynniki standaryzowane	Współczynniki niestandaryzowane	SE	CR	P	Zmienne	Procent wyjaśnianej wariancji
$x_1 \leftarrow \xi_1$	,829	5,308	1,362	3,896	<0,001	$x_1$	,687
$x_2 \leftarrow \xi_1$	,822	1,000				$x_2$	,675
$x_3 \leftarrow \xi_1$	,593	,938	,309	3,040	,002	$x_3$	,351
$x_8 \leftarrow \xi_1$	-,307	-,448	,274	-1,631	,103		
$x_4 \leftarrow \xi_2$	-,812	-7,031	1,081	-6,506	<0,001	$x_4$	,659
$x_5 \leftarrow \xi_2$	-,854	-25,416	3,464	-7,337	<0,001	$x_5$	,730
$x_6 \leftarrow \xi_2$	,973	1,000				$x_6$	,948
$x_7 \leftarrow \xi_2$	,714	2,760	,547	5,041	<0,001	$x_7$	,510
$x_8 \leftarrow \xi_2$	,343	,676	,338	2,002	,045	$x_8$	,212
$y \leftarrow \xi_1$	-,463	-2,770	1,132	-2,447	,014		
$y \leftarrow \xi_2$	,289	2,340	1,328	1,763	,078	$y$	,298

Źródło: opracowanie własne za pomocą programu AMOS SPSS, estymator ML.

dym z nich dodano odrębny składnik losowy  $\delta$ . Czynniki otoczenia opisują:  $x_1, x_2, x_3, x_8$ , czynnik szkolny reprezentują:  $x_4, x_5, x_6, x_7$ , oraz ponownie  $x_8$  jako zmienna, której wartość zależy zarówno od zaangażowania szkoły, jak i uczniów. Warto podkreślić, że duże wartości zmiennych:  $x_1, x_2, x_3$  oraz  $x_4, x_5$  wskazują na niesprzyjające warunki działania. W tabeli 1 zamieszczono oszacowania parametrów wybranego modelu SEM otrzymane metodą największej wiarygodności.

Wartości bezwzględne standaryzowanych szacunków parametrów przekraczają 0,7 dla zmiennych  $x_1, x_2, x_4, x_5, x_6, x_7$ , dla zmiennej  $x_3$  przekraczają 0,5, jedynie dla  $x_8$  są mniejsze. Procent wyjaśnianej wariancji dla  $x_3$  i  $x_8$  nie przekracza pożądanej wartości 0,5. Prawdopodobieństwo testowe większe od 0,05 współczynnika ścieżkowego  $x_8 \leftarrow \xi_1$  wskazuje na brak podstaw do odrzucenia hipotezy zerowej o nieistotności tego parametru. Zaskakująco nieistotny okazał się też wpływ zmiennej ukrytej  $\xi_2$  (szkoły) na przyrost wiedzy jej uczniów. Można spróbować zupełnie odrzucić tę zależność – jednak biorąc pod uwagę teorię, byłoby to co najmniej dyskusyjnie. Sformułowany model opisujący przyrost wiedzy uczniów szkoły jako funkcję czynników otoczenia i szkolnych nie znalazł potwierdzenia w empirycznych danych. Bazując na wskazaniach indeksów modyfikacji, dokonano zmian w pierwotnym modelu. Wprowadzono zależność kowariancyjną pomiędzy konstrukcjami ukrytymi, a także zależność regresyjną pomiędzy zmiennymi  $x_3$  i  $x_8$ . Zmodyfikowany model przedstawiono na rysunku 2.



**Rys. 2.** Model SEM zmodyfikowany

Źródło: opracowanie własne.

Oszacowania zmodyfikowanego modelu SEM zamieszczono w tabeli 2. W modelu tym wszystkie współczynniki ścieżkowe są statystycznie istotne.

Warto podkreślić, że dla danego procesu można określić bardzo wiele alternatywnych modeli teoretycznych z różnymi wzorcami relacji. Co więcej, dla prawie każdego procesu można zdefiniować modele ekwiwalentne, które generują takie same wartości statystyk ogólnego dopasowania, ale reprezentują substancyjnie inne teorie [Konarski 2009, s. 405]. Teoria badanego procesu musi rozstrzygać,

**Tabela 2.** Oszacowania zmodyfikowanego wybranego modelu SEM

Ścieżka	Współczynniki standaryzowane	Współczynniki niestandaryzowane	SE	CR	P	Zmienne	Procent wyjaśnianej wariancji
$x_1 \leftarrow \xi_1$	,796	4,852	1,082	4,486	<0,001	$x_1$	,634
$x_2 \leftarrow \xi_1$	,864	1,000				$x_2$	,747
$x_3 \leftarrow \xi_1$	,334	,498	,239	2,083	,037	$x_3$	,502
$x_3 \leftarrow x_8$	-,520	-,530	,146	-3,621	<0,001		
$x_4 \leftarrow \xi_2$	-,818	-7,139	1,075	-6,644	<0,001	$x_4$	,669
$x_5 \leftarrow \xi_2$	-,861	-25,815	3,424	-7,539	<0,001	$x_5$	,741
$x_6 \leftarrow \xi_2$	,966	1,000				$x_6$	,933
$x_7 \leftarrow \xi_2$	,714	2,779	,552	5,034	<0,001	$x_7$	,509
$x_8 \leftarrow \xi_2$	,499	1,046	,352	2,970	,003	$x_8$	,249
$y \leftarrow \xi_1$	-,629	-3,832	1,109	-3,454	<0,001	$y$	,396

Źródło: opracowanie własne za pomocą programu AMOS SPSS, estymator ML.

który z alternatywnych modeli jest najbardziej spójny [Konarski 2009, s. 17]. Model doskonale dopasowany, ale niepasujący do teorii, jest właściwie gorszy niż model dopasowany na granicy akceptowalności, ale zgodny z teorią [Bedyńska, Książek 2012, s. 182].

Uwzględniając powyższe, postanowiono przeprowadzić formalną ocenę jakości oryginalnego modelu i zmodyfikowanego. Test dopasowania modelu opiera się na wartości funkcji rozbieżności (FMIN) pomiędzy obserwowaną, a wynikającą z modelu macierzą kowariancji. Na bazie funkcji rozbieżności opracowano liczne mierniki dopasowania [Konarski 2009, s. 328].

Statystyka testowa ogólnego dopasowania modelu (CMIN) ma asymptotycznie rozkład  $\chi^2$ . Jeżeli jej wartość przekracza wartość krytyczną hipotezę  $H_0$ , mówiącą, że model jest doskonale dopasowany do danych można odrzucić. Wartość CMIN zależy od wielkości próby (gdy próba jest zbyt mała, a taka jest analizowana w przykładzie, statystyka może niesłusznie nie dawać podstaw do odrzucenia  $H_0$ ) [Garson 2012, s. 38-39]. Prosta korektą złożoności jest podzielenie wartości CMIN przez liczbę stopni swobody DF i odrzucenie modelu, gdy stosunek ten przekracza 2 lub mniej restrykcyjnie nawet 10 [Bedyńska, Książek 2012, s. 184]. Wartość FMIN oraz statystyki CMIN modelu pierwotnego (model 1) oraz zmodyfikowanego (model 2) zamieszczono w tabeli 3. Warto podkreślić, że w obu modelach każda z tych miar jest bliższa mierze modelu nasyconego niż modelu niezależności.

W tabeli 4 przedstawiono wybrane miary bezwzględnego oraz relatywnego dopasowania. Miary bezwzględnego dopasowania modelu informują, o ile testowany model jest lepszy od modelu niezależności. GFI (*goodness of fit index*) i skorygowany AGFI (*adjusted goodness of fit index*) inspirowane są analogią do współ-

**Tabela 3.** Statystyki testowe ogólnego dopasowania modeli SEM

Model	FMIN	CMIN	DF	P	CMIN/DF
Model 1	1,394	40,42	25	,026	1,62
Model 2	,637	18,46	25	,822	,74
Nasycony	,000	,00	0		
Niezależności	5,948	172,48	36	,000	4,79

Źródło: opracowanie własne za pomocą programu AMOS SPSS.

czynnika determinacji  $R^2$  i mierzą proporcję wariancji w empirycznej macierzy kowariancji, która została wyjaśniona przez model. Wartości indeksów powinny znajdować się pomiędzy 0 a 1. Bliskie 0 oznaczają kompletny brak dopasowania, bliskie 1 wskazują na bardzo dobre dopasowanie. Wartość co najmniej 0,9 jest dolną granicą akceptacji modelu [Konarski 2009, s. 346]. Zaproponowany model 1 wyjaśnia tylko 79% zmienności macierzy kowariancji, model 2 nieco ponad 88%. Natomiast wartość AGFI nie przekracza nawet 0,7 (model 1) lub 0,8 (model 2). Miary relatywnego dopasowania wskazują z kolei na adekwatność modelu w odniesieniu do modelu niezależności. NFI (*normed fit index*), RFI (*relative fit index*), IFI (*incremental fit index*), CFI (*comparative fit index*) także przeważnie przyjmują wartości z przedziału  $<0, 1>$ . Przyjmuje się, że minimalną wartość NFI, RFI, IFI i CFI uprawniającą do akceptacji modelu jest 0,9 [Bedyńska, Książek 2012, s. 187]. Tylko w wypadku modelu 2 miary te są akceptowalne lub bliskie progu akceptowalności.

**Tabela 4.** Miary bezwzględnego i relatywnego dopasowania

Model	Miary bezwzględnego dopasowania		Miary relatywnego dopasowania			
Model	GFI	AGFI	NFI	RFI	IFI	CFI
Model 1	,790	,622	,766	,663	,895	,887
Model 2	,886	,795	,893	,846	1,044	1,000
Nasycony	1,000		1,000		1,000	1,000
Niezależności	,329	,161	,000	,000	,000	,000

Źródło: opracowanie własne za pomocą programu AMOS SPSS.

Podsumowując, wśród analizowanych mierników prawie wszystkie nakazują model 1 odrzucić. Model 2 jest do zaakceptowania, ale nie można powiedzieć, że jest idealny. W jego przypadku większość miar, dla których są przyjęte normy interpretacji, jeżeli nie sięga 0,9, to przynajmniej przekracza 0,8. Pierwotnie wyodrębnione determinanty przyrostu wiedzy: szkoła i otoczenie okazały się powiązane między sobą. Współczynnik korelacji pomiędzy szkołą ( $\xi_2$ ) a otoczeniem ( $\xi_1$ ) wynosi  $-0,691$ . W modelu 2 na osiągnięty przez szkołę przyrost wiedzy bezpośredni wpływ ma tylko otoczenie. Wyższa ocena niesprzyjającego otoczenia o 1 odchylenie

nie standardowe przekłada się na spadek przyrostu wiedzy o  $-0,629$  odchylenia standardowego. Wszystkie analizowane zmienne wyjaśniają ponad 39% zmienności przyrostu wiedzy uczniów.

#### 4. Podsumowanie

SEM jest bardzo uniwersalnym narzędziem analizy i testowania hipotetycznych zależności liniowych pomiędzy zmiennymi obserwowanymi i/lub ukrytymi [Bedyńska, Książek 2012, s. 198]. Komputerowe pakiety statystyczne, pozwalające na wygodne rysowanie wykresów ścieżkowych, sprawiają wrażenie, że SEM jest techniką niewymagającą ani refleksji, ani wiedzy statystycznej. Stosowanie SEM zmusza jednak do określenia, chociażby w minimalnym stopniu, mechanizmu przyczynowego, który leży u podłoża zaobserwowanych korelacji między zmiennymi obserwowanymi [Konarski 2009, s. 15]. Wymaga też dużych prób danych, a większość metod estymacji parametrów wielowymiarowego rozkładu normalnego.

W artykule przedstawiono wyniki testowania dopuszczalności hipotezy o przyczynowym związku między przyrostem wiedzy a wybranymi wskaźnikami otoczenia szkoły i samej szkoły. Pierwotny prosty model zależności, klasyfikujący obserwowane wskaźniki, nie przeszedł formalnej oceny jakości. Dużo lepiej wykonuje swoje zadanie model zmodyfikowany na podstawie indeksów modyfikacji.

W zaakceptowanym modelu na przyrost wiedzy ma wpływ tylko czynnik otoczenia mierzony wskaźnikiem liczby uczniów korzystających z posiłków refundowanych ( $x_1$ ), wskaźnikiem liczby uczniów powtarzających klasę ( $x_2$ ) oraz wskaźnikiem liczby uczniów, którzy opuścili więcej niż 30% zajęć bez usprawiedliwienia ( $x_3$ ). Czynnik szkolny reprezentowany przez: liczbę uczniów w szkole ( $x_4$ ), średnią liczbę uczniów w oddziale ( $x_5$ ), liczbę komputerów do użytku uczniów na liczbę uczniów ogółem ( $x_6$ ), liczbę woluminów w bibliotece na liczbę uczniów ogółem ( $x_7$ ) oraz liczbę uczniów placówki uczestniczących w zajęciach pozalekcyjnych na liczbę uczniów ogółem ( $x_8$ ) jest skorelowany z czynnikiem otoczenia. Dodatkowo istnieje ujemna zależność między absencją ( $x_3$ ) a udziałem w zajęciach pozalekcyjnych ( $x_8$ ).

Autorka ma świadomość niepełności wykonanej analizy i związanych z tym problemów z generalizacją otrzymanych wyników. Przede wszystkim ocena została przeprowadzona na podstawie nielicznego, zagregowanego dla szkół zbioru danych. Ze względu na fakt, że – jak wykazano – efektywność nauczania determinuje przede wszystkim otoczenie, dużo właściwsza byłaby analiza i uwzględnienie w modelu SEM informacji o statusie materialnym czy ewentualnych problemach poszczególnych uczniów. Jednak dane takie, klasyfikowane jako wrażliwe, są niełatwe do pozyskania, a ich wykorzystanie limitowane prawnie.

## Literatura

- Bedyńska S., Książek M. (2012), *Statystyczny drogowskaz 3. Praktyczny przewodnik wykorzystania modeli regresji raz równań strukturalnych*, Wydawnictwo Akademickie Sedno, Warszawa.
- Chodakowska E. (2009), *Przykłady zastosowań metody Data Envelopment Analysis w badaniu efektywności podmiotów sektora edukacji*, „Problemy Zarządzania”, vol. 7, nr 4, s. 91-112.
- Dufur M.J., Parcel T.L., Troutman K.P. (2013), *Does capital at home matter more than capital at school? Social capital effects on academic achievement*, „Research in Social Stratification and Mobility”, vol. 31, s. 1-21.
- Garson D. (2012), *Structural Equation Modeling*, Statistical Associates Publishers, Asheboro, NC.
- Konarski R. (2009), *Modele równań strukturalnych. Teoria i praktyka*, Wydawnictwo Naukowe PWN, Warszawa.
- Kyounge Ro H., Terenzini P.T., Yin A.C. (2013), *Between-College Effects on Students Reconsidered*, „Research in Higher Education”, vol. 54, no. 3, s. 253-282.
- McKenzie K., Gow K. (2004), *Exploring the first year academic achievement of school leavers and mature-age students through structural equation modeling*, „Learning and Individual Differences”, vol. 14, no. 2, s. 107-123.
- Rosandera P., Bäckströma M., Stenberg G. (2011), *Personality traits and general intelligence as predictors of academic performance: A structural equation modelling approach*, „Learning and Individual Differences”, vol. 21, no. 5, s. 590-596.
- Sagan A. (2003a), *Model pomiarowy satysfakcji i lojalności*, [w:] *Analiza satysfakcji i lojalności klientów*, StatSoft Polska, Warszawa-Kraków, s. 75-85.
- Sagan A. (2003b), *Model strukturalny relacji między satysfakcją i lojalnością wobec marki*, [w:] *Analiza satysfakcji i lojalności klientów*, StatSoft Polska, Warszawa-Kraków, s. 87-92.

### THE THEORY OF STRUCTURAL EQUATIONS MODELLING IN THE CLASSIFICATION OF OBSERVED VARIABLES AND LATENT CONSTRUCTS ACCORDING TO THE CHARACTER OF THEIR RELATIONSHIP

**Summary:** This paper presents an attempt of analysis and classification of variables that determine the effectiveness of teaching in selected secondary schools in Poland. Using SEM hypothetical relationships between variables describing the exam results and selected features of schools and its students were verified. The effectiveness of the schools in the survey was affected directly only by the environmental factor represented by selected characteristics of students. The school's factor represented by teaching conditions had no direct impact on the schools effectiveness, but was correlated with the environmental factor.

**Keywords:** SEM, CFA, classification, effectiveness of education.