

AUTOREFERAT

SPIS TREŚCI

| | |
|---|----|
| A. Dane osobowe..... | 2 |
| B. Przebieg kariery naukowej..... | 3 |
| C. Przebieg pracy zawodowej..... | 5 |
| D. Wskazanie i opis osiągnięcia wynikającego z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz stopniach i tytule w zakresie sztuki (Dz. U. nr 65, poz. 595 ze zm.)..... | 6 |
| D.1. Tytuł osiągnięcia naukowego..... | 6 |
| D.2. Publikacja stanowiąca osiągnięcie naukowe..... | 6 |
| D.3. Syntetyczne omówienie osiągnięcia naukowego..... | 6 |
| E. Omówienie pozostałych osiągnięć naukowo badawczych..... | 23 |
| E.1. Zestawienie dorobku publikacyjnego przed i po uzyskaniu stopnia doktora..... | 23 |
| E.2. Wskaźniki bibliometryczne..... | 23 |
| E.3. Omówienie pozostałych osiągnięć w pracy naukowej..... | 24 |

A. DANE OSOBOWE

Imię/imiona i nazwisko: Dariusz Rafał Mańkowski

Data i miejsce urodzenia: 9 maja 1978 r., Warszawa

Stopień naukowy: doktor inżynier

Miejsce zatrudnienia: Instytut Hodowli i Aklimatyzacji Roślin – Państwowy Instytut Badawczy
Radzików, 05-870 Błonie
Zakład Nasiennictwa i Nasionoznawstwa
Pracownia Ekonomiki Nasiennictwa i Hodowli Roślin

B. PRZEBIEG KARIERY NAUKOWEJ

Studia wyższe:

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
Wydział Rolniczy
Specjalność: Agronomia i Agrobiznes

Praca magisterska realizowana w Katedrze Statystyki Matematycznej i Doświadczalnictwa pod opieką prof. dr hab. Zbigniewa Laudańskiego pt.: „Propozycja metody podziału na podgrupy podobnych populacji jednowymiarowych na podstawie danych z doświadczeń jednoczynnikowych” (Rec.: prof. dr hab. Jan Rozbicki; praca obroniona z wyróżnieniem)

Data ukończenia: czerwiec 2002 r.

Doktorat:

W okresie od października 2003 r. do czerwca 2007 r. udział jako wolny słuchacz w Studium Doktoranckim „Biologiczne uwarunkowania produkcji roślinnej” na Wydziale Rolnictwa i Biologii Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie.

Rozprawa doktorska realizowana w Instytucie Hodowli i Aklimatyzacji Roślin pod opieką prof. dr hab. Zbigniewa Laudańskiego pt.: „Postęp biologiczny w hodowli, nasiennictwie i produkcji ziemniaka w Polsce” (Rec.: prof. dr hab. Ewa Zimnoch-Guzowska, prof. dr hab. Zygmunt Kaczmarek; praca obroniona z wyróżnieniem)

Data nadania stopnia doktora: 16 grudnia 2008 r.

Studia podyplomowe:

Politechnika Łódzka, Instytut Elektroniki
Specjalność: Przetwarzanie i analiza obrazów biomedycznych
Studia finansowane w ramach Europejskiego Funduszu Społecznego

Data ukończenia: marzec 2011 r.

Inne formy edukacji:

Metody statystyczne w pracach badawczych i hodowlanych
Seminarium specjalistyczne / Szkolenie w IHAR — Radzików
(grudzień 2002 r.)

Wprowadzenie do SAS 9.1

Szkolenie w SAS Institute Polska Sp. z o. o.
(maj 2005 r.)

Przetwarzanie danych w systemie SAS 9.1 cz. I

Szkolenie w SAS Institute Polska Sp. z o. o.
(maj 2005 r.)

Przetwarzanie danych w systemie SAS 9.1 cz. II

Szkolenie w SAS Institute Polska Sp. z o. o.
(czerwiec 2005 r.)

Metody analizy danych w systemie SAS 9.1 cz. I

Szkolenie w SAS Institute Polska Sp. z o. o.
(wrzesień 2005 r.)

Metody analizy danych w systemie SAS 9.1 cz. II

Szkolenie w SAS Institute Polska Sp. z o. o.
(marzec 2006 r.)

Komercjalizacja Nauk Innowacyjnych

Warsztaty organizowane przez Fundację na Rzecz Budowy Społeczeństwa Opartego na Wiedzy „Nowe Media” finansowane z Europejskiego Funduszu Strukturalnego, dotyczące zarządzania badaniami naukowymi oraz komercjalizacji wyników badań naukowych
(lipiec 2010 r.)

System R oraz zastosowanie R w genetyce populacyjnej

Szkolenie zorganizowane w IHAR-PIB w Radzikowie przez firmę QuantUp
(sierpień 2012)

STATISTICA kurs podstawowy

Szkolenie w StatSoft Polska Sp. z o. o.
(marzec 2015 r.)

Analiza danych ankietowych - kurs podstawowy

Szkolenie w StatSoft Polska Sp. z o. o.
(wrzesień 2015 r.)

Analiza danych ankietowych - kurs zaawansowany

Szkolenie w StatSoft Polska Sp. z o. o.
(październik 2015 r.)

*Przygotowywanie automatycznych aplikacji/modułów do statystycznej analizy danych i generowania raportów z analiz z wykorzystaniem programu**Statistica i Statistica VB*

Szkolenie w StatSoft Polska Sp. z o. o.
(luty 2016 r.)

C. PRZEBIEG PRACY ZAWODOWEJ

- 2001–2002 Staż asytencki w Katedrze Statystyki Matematycznej i Doświadczalnictwa
Wydziału Rolniczego Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie
- 2002–aktualnie Instytut Hodowli i Aklimatyzacji Roślin — Państwowy Instytut Badawczy
w Radzikowie, Zakład Nasiennictwa i Nasionoznawstwa, Pracownia Ekonomiki
Nasiennictwa i Hodowli Roślin; stanowiska — inżynier stażysta (2002–2003),
inżynier (2003), asystent (2003–2008), adiunkt (2009–aktualnie)

D. WSKAZANIE I OPIS OSIĄGNIĘCIA WYNIKAJĄCEGO Z ART. 16 UST. 2 USTAWY Z DNIA 14 MARCA 2003 R. O STOPNIACH NAUKOWYCH I TYTULE NAUKOWYM ORAZ STOPNIACH I TYTULE W ZAKRESIE SZTUKI (DZ. U. NR 65, POZ. 595 ZE ZM.)

D.1. TYTUŁ OSIĄGNIĘCIA NAUKOWEGO

Modele równań strukturalnych SEM w badaniach rolniczych

D.2. PUBLIKACJA STANOWIĄCA OSIĄGNIĘCIE NAUKOWE

Mańkowski, D. R., 2013. Modele równań strukturalnych SEM w badaniach rolniczych. Monografie i Rozprawy Naukowe IHAR-PIB, Nr. 42. IHAR-PIB, Radzików. [ISBN: 83-891172-70-4].

D.3. SYNTETYCZNE OMÓWIENIE OSIĄGNIĘCIA NAUKOWEGO

Metoda zwana modelowaniem równań strukturalnych (*structural equation modeling*, SEM) znana jest od lat 70. ubiegłego wieku. Jednak nie jest ona powszechnie wykorzystywana w modelowaniu zjawisk i procesów w badaniach przyrodniczych, prawdopodobnie przez swoją złożoność i problemy, jakie może sprawiać niedoświadczonym badaczom w jej zrozumieniu i prawidłowym zastosowaniu. Zwyczajowo stosowane metody, takie jak prosta analiza ścieżek, cechują się jednak poważnymi ograniczeniami, które są związane z rozkładami analizowanych zmiennych lub z wzajemnymi relacjami (niebędącymi bezpośrednim przedmiotem badań) pomiędzy tymi zmiennymi. Tymczasem, to wzajemne powiązania pomiędzy zmiennymi decydują o przebiegu i charakterze modelowanych zjawisk czy procesów. Powszechnie stosowane dziś metody modelowania złożonych zależności przyczynowo-skutkowych nie dają w pełni zadowalających rezultatów w tym zakresie. Uważam, że analiza SEM może być rozwiązaniem tego problemu, o ile zostanie zmodyfikowana stosownie do potrzeb analiz prowadzonych w dziedzinie nauk przyrodniczych. Moja monografia zawiera szczegółowy opis takiego podejścia wraz z przykładami, ukazującymi skuteczność zaproponowanych rozwiązań. Kluczowymi elementami prezentowanego przeze mnie osiągnięcia naukowego jest adaptacja metody SEM do warunków badań rolniczych oraz nowatorskie, autorskie podejście do transformacji danych, a także pokazanie skuteczności tych metod na rzeczywistych przykładach. Prezentowana monografia liczy 159 stron, zatem poniżej przedstawiam w skrócie jej poszczególne części. Należy także podkreślić, że monografia ta jest jedyną pozycją, w skali kraju, a prawdopodobnie również na świecie, zawierającą podstawy merytoryczne metody SEM oraz przykłady jej adaptacji w naukach rolniczych, które wymagają specjalnego podejścia metodologiczno-filozoficznego.

Wstęp

W badaniach przyrodniczych, w tym w naukach rolniczych właściwie nie ma cech niezależnych. Współzależności i zależności przyczynowo-skutkowe są więc bardzo ważnym obiektem badań. Poznanie związków pomiędzy zmiennymi charakteryzującymi zjawiska i procesy przyrodnicze pozwala na wyjaśnienie oraz zrozumienie charakteru i przebiegu tych zjawisk i procesów.

Badanie porównawcze obiektów prowadzone tylko w jednym aspekcie nie jest wystarczające do wyjaśnienia badanego procesu lub zjawiska (Kaczmarek i Mańkowski, 2011). Również badanie porównawcze obiektów pod względem wielu cech, ale każdej rozpatrywanej oddzielnie, nie może przyczynić się do pełnego wyjaśnienia zachodzących zjawisk i procesów. Dopiero równoczesne uwzględnienie wszystkich obserwowanych cech stanowić może podstawę do wyciągnięcia adekwatnych wniosków. Badanie zależności w zbiorowościach wielu

zmiennych wydaje się być zagadnieniem trudnym, ma ono jednak duże znaczenie poznawcze. O ile analiza pojedynczych zmiennych lub par zmiennych prowadzi do jednostronnych wniosków, o tyle analizy wielozmienne pozwalają na szersze spojrzenie na analizowane zjawisko lub proces. Jednak metody, których należy użyć, aby przeprowadzić analizę wielu zmiennych, są często bardzo złożone i sprawiają badaczom wiele problemów w interpretacji uzyskanych wyników. Szeroki zakres informacji jakie można uzyskać stosując metody wielozmienne, jak i wielowymiarowe spojrzenie na analizowany proces lub zjawisko, zdają się jednak jednoznacznie przemawiać na korzyść wykorzystywania analiz wielozmiennych w procesie badawczym (Kaczmarek i Mańkowski, 2011).

W naukach rolniczych informacja o tym, co w badanym procesie lub zjawisku jest przyczyną a co skutkiem, jest bardzo wartościowa, gdyż pozwala na zrozumienie istoty tego procesu lub zjawiska. Błędne określenie przyczyny lub skutku może prowadzić do błędnej interpretacji badanego procesu (zjawiska) i niezrozumienia jego istoty. Uznaje się, że celem statystycznej analizy systemów związków przyczynowo-skutkowych między zmiennymi jest odnalezienie modelu matematycznego jak najlepiej opisującego ten system. Z takiego modelu można wnioskować o tym, które z postulowanych zależności okazały się istotne i jaki jest ich charakter (Kozak, 2008, 2011).

W praktyce doświadczalnej najpopularniejszymi metodami analizy związków przyczynowo-skutkowych są:

- analiza funkcji regresji prostej – opisuje zależność pomiędzy jedną zmienną stanowiącą przyczynę danego zjawiska lub procesu i zmienną stanowiącą skutek; metoda opiera się na estymacji metodą najmniejszych kwadratów (MNK) (Rawlings i in., 2001),
- analiza funkcji regresji wielorakiej – opisuje zależność pomiędzy grupą niezależnych zmiennych przyczynowych i jedną zmienną stanowiącą skutek danego zjawiska lub procesu; również opiera się na estymacji MNK (Rawlings i in., 2001),
- analiza ścieżek (*path analysis*, PA) – jest najpopularniejszą metodą analizy złożonych zależności przyczynowo-skutkowych; jej twórcą był Wright (1921, 1923, 1934); polega na rozbiciu tworzonego diagramu zależności (tzw. diagramu ścieżek) na szereg modeli dla każdej zmiennej endogenicznej (opisywanej przez model) i analizy takich modeli za pomocą analizy funkcji regresji wielorakiej z przeprowadzeniem standaryzacji; pozwala na uwzględnienie w analizie hierarchii zmiennych przyczynowych (Gozdowski i Mądry, 2008; Kozak, 2008),
- inne, bardziej specjalistyczne metody jak sekwencyjna analiza plonu (*sequential yield component analysis*, SYCA) (Mądry i Kozak, 2000; Kozak, 2006, 2008, 2011), dwukierunkowy podział zmienności plonu (*two-dimensional partitioning of yield variation*, TDP) (Eaton, 1986; Gołaszewski, 1996; Gołaszewski i in., 1998) czy też sztuczne sieci neuronowe (*artificial neural networks*, ANN) (Krzyśko i in., 2008; O’Neal i in., 2002; Janaszek i in., 2011).

Celem pracy będącej przedmiotem osiągnięcia naukowego była adaptacja metody SEM do nauk przyrodniczych, a w szczególności nauk rolniczych oraz zaprezentowanie na przykładach sposobu jej wykorzystania jak też interpretacji uzyskanych wyników analiz. Opis metody SEM szczegółowo zaprezentowano w postaci modeli, które zdaniem autora mogą znaleźć praktyczne zastosowanie w doświadczalnictwie rolniczym. Indywidualnym osiągnięciem naukowym jest również nowatorska koncepcja transformacji danych oparta na zastosowaniu modelu liniowego z ważonymi warunkami identyfikacji efektów. Zaproponowana przeze mnie metoda przekształcania danych może znaleźć znacznie szersze zastosowanie, niż wyłącznie w analizie modeli bazujących na równaniach strukturalnych.

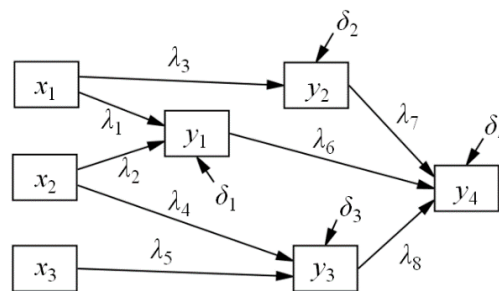
Podstawy teoretyczne SEM

Powstanie metody SEM datuje się na lata 70. XX wieku. Za twórców tej metody uważa się Keesling'a (1972), Jöreskog'a (1973) oraz Wiley'a (1973)— stąd nazwa JKW pierwszych modeli strukturalnych. W późniejszych latach metoda SEM była dalej rozwijana, równolegle ewoluowała nazwa modeli — LINEQS, RAM i LISREL. Na dzień dzisiejszy metoda znalazła powszechne wykorzystanie przede wszystkim w naukach społecznych (Konarski, 2009), w ekonometrii, inżynierii i mechanice; w naukach rolniczych nie jest jeszcze powszechnie stosowana. W literaturze światowej można znaleźć prace, w których do modelowania złożonych

zależności przyczynowo-skutkowych wykorzystano modele SEM, jednak w stosunku do innych dziedzin nauki jest tych prac stosunkowo niewiele. Jeśli chodzi o literaturę polską, takich prac jest bardzo niewiele.

Analiza SEM jest metodą służącą przede wszystkim potwierdzeniu modeli teoretycznych, a nie ich eksploracji lub poszukiwaniu. SEM jest zatem stosowany do sprawdzenia, czy sformułowany *a priori* model ma oparcie w empirycznej próbie danych. Nie służy natomiast do „znalezienia” właściwego modelu opisującego relacje w danym zbiorze danych empirycznych.

Analizę złożonej zależności przyczynowo-skutkowej, zaczyna się z reguły od przedstawienia modelu zależności w postaci diagramu ścieżek. Diagram taki jest graficzną prezentacją wszystkich występujących w badaniach zmiennych i zależności oraz wzajemnych powiązań pomiędzy zmiennymi. Ze względu na rolę w modelu zmienne dzielimy na egzogeniczne, czyli takie które wchodzą do modelu jako zmienne zewnętrzne, oraz zmienne endogeniczne, czyli takie, których kształtowanie się jest w jakimś stopniu opisywane przez postulowany model. Dodatkowo zmienne w analizie SEM możemy podzielić na zmienne obserwowalne — czyli takie, które są fizycznie mierzalne i możliwe do zaobserwowania oraz na zmienne latentne, czyli zmienne, których nie da się bezpośrednio zaobserwować ani zmierzyć. Możemy jedynie obserwować manifestację zmiennych latentnych w postaci zmian zmiennych obserwowalnych zależnych od tych zmiennych latentnych. Dodatkowo w modelu identyfikuje się błędy losowe pojawiające się przy opisie zmiennych endogenicznych. Błędy losowe wyrażają wszystko to, co wpływa na analizowaną zmienną lecz nie jest uwzględnione w modelu. Za pomocą strzałek na diagramie ścieżek przedstawia się relacje pomiędzy zmiennymi. Modele ścieżkowe możemy przedstawić również w postaci układu równań strukturalnych wraz z warunkami identyfikacji. I tak przykładowy model, który przedstawia poniższy diagram:



możemy zapisać w postaci następującego układu równań:

$$\begin{cases} y_1 = \lambda_1 x_1 + \lambda_2 x_2 + \delta_1 \\ y_2 = \lambda_3 x_1 + \delta_2 \\ y_3 = \lambda_4 x_2 + \lambda_5 x_3 + \delta_3 \\ y_4 = \lambda_6 y_1 + \lambda_7 y_2 + \lambda_8 y_3 + \delta_4 \\ Cov(\delta_i; \delta_{j \neq i}) = 0 \end{cases} \quad [1]$$

Podstawowym celem analizy modeli równań strukturalnych jest wyjaśnienie zaobserwowanej zmienności w zbiorze danych, opisanej przez macierz kowariancji, za pomocą mniejszej liczby parametrów postulowanego modelu badanego procesu lub zjawiska. Powyższy cel implikuje podstawową hipotezę SEM mówiącą o tym, że macierz kowariancji zmiennych obserwowalnych jest funkcją zbioru parametrów modelu. Hipotezę tę zapisuje się w postaci:

$$H_0: \Sigma = \Sigma(\theta) \quad [2]$$

gdzie: Σ jest macierzą kowariancji zmiennych obserwowalnych w badanej populacji; θ jest wektorem parametrów postulowanego modelu teoretycznego; $\Sigma(\theta)$ jest macierzą kowariancji zmiennych obserwowalnych, wyrażona jako funkcja parametryczna wektora θ .

Powyższa postać hipotezy zerowej daje dużą swobodę w konstruowaniu badanych modeli, co z kolei pozwala na szerokie zastosowania tej metody w prowadzonych badaniach i analizie zależności przyczynowo-skutkowych. Budowane modele zależności możemy podzielić na:

- proste modele ze zmiennymi obserwowalnymi – w takim modelu wszystkie zmienne to zmienne obserwowalne, jest to najpowszechniejszy typ modeli stosowanych w badaniach rolniczych, układ równań strukturalnych dla takich modeli możemy zapisać w postaci macierzowej

$$\mathbf{y} = \mathbf{B}\mathbf{y} + \mathbf{\Gamma}\mathbf{x} + \boldsymbol{\delta} \quad [3]$$

gdzie \mathbf{y} jest wektorem zmiennych endogenicznych, \mathbf{x} jest wektorem zmiennych egzogenicznych, elementy macierzy \mathbf{B} oraz $\mathbf{\Gamma}$ są współczynnikami strukturalnymi definiującymi model, a $\boldsymbol{\delta}$ to wektor reszt;

- modele ze zmiennymi latentnymi – w takim modelu wszystkie zmienne to zmienne latentne, takich modeli w praktyce prawie nie spotyka się w badaniach rolniczych, układ równań strukturalnych dla takich modeli możemy zapisać w postaci macierzowej

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \mathbf{\Gamma}\boldsymbol{\xi} + \boldsymbol{\varepsilon} \quad [4]$$

gdzie $\boldsymbol{\eta}$ jest wektorem latentnych zmiennych endogenicznych, $\boldsymbol{\xi}$ jest wektorem latentnych zmiennych egzogenicznych, elementy macierzy \mathbf{B} oraz $\mathbf{\Gamma}$ są współczynnikami strukturalnymi definiującymi model, a $\boldsymbol{\varepsilon}$ to wektor reszt;

- modele ogólne – zawierają zarówno zmienne obserwowalne jak i latentne, a w naukach rolniczych są dość rzadkie; ich przykładem są modele konfirmacyjnej analizy czynnikowej (CFA). Układ równań strukturalnych dla tych modeli można zapisać w postaci macierzowej:

$$\begin{aligned} \boldsymbol{\eta} &= \mathbf{B}\boldsymbol{\eta} + \mathbf{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \\ \mathbf{x} &= \mathbf{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta} \\ \mathbf{y} &= \mathbf{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon} \end{aligned} \quad [5]$$

gdzie $\boldsymbol{\eta}$ jest wektorem latentnych zmiennych endogenicznych, $\boldsymbol{\xi}$ jest wektorem latentnych zmiennych egzogenicznych, \mathbf{x} jest wektorem obserwowalnych zmiennych egzogenicznych, \mathbf{y} jest wektorem obserwowalnych zmiennych endogenicznych, elementy macierzy \mathbf{B} , $\mathbf{\Gamma}$, $\mathbf{\Lambda}_x$ oraz $\mathbf{\Lambda}_y$ są współczynnikami strukturalnymi, definiującymi model, a $\boldsymbol{\zeta}$, $\boldsymbol{\delta}$, oraz $\boldsymbol{\varepsilon}$ to wektory reszt.

Najważniejszą rolą analizy SEM jest oszacowanie wartości parametrów postulowanego modelu oraz ocena stopnia dopasowania tego modelu do relacji obserwowanych w zbiorze danych empirycznych. Właściwie nie zdarza się, by postulowany model teoretyczny, opracowany *a priori*, znalazł kompletne odwzorowanie w danych empirycznych. Niedostatecznie dopasowane modele powinny zostać odrzucone. Podobnie jak w innych metodach modelowania statystycznego, również w tym przypadku należy dążyć do tego, aby różnica między postulowanym modelem a danymi empirycznymi była jak najmniejsza.

Proces analizy modeli SEM rozpoczyna się od sformułowania modelu odzwierciedlającego wszystkie możliwe relacje w ramach badanego procesu lub zjawiska. Model ten zapisuje się w postaci diagramu ścieżek, a następnie układu równań strukturalnych. Etap ten nosi nazwę specyfikacji modelu.

W przypadku modelu ze zmiennymi obserwowalnymi, jeżeli istnieje macierz $(\mathbf{I} - \mathbf{B})^{-1}$ to układ równań strukturalnych [3] można zapisać w zredukowanej postaci regresyjnej

$$\mathbf{y} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma}\mathbf{x} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\delta} \quad [6]$$

W nawiązaniu do hipotezy zerowej [2], implikowaną macierz kowariancji w postulowanym modelu SEM dla zmiennych obserwowalnych można zapisać w postaci

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} (\mathbf{I} - \mathbf{B})^{-1}(\mathbf{\Gamma}\boldsymbol{\Phi}\mathbf{\Gamma}^T + \boldsymbol{\Psi})(\mathbf{I} - \mathbf{B}^T)^{-1} & (\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma}\boldsymbol{\Phi}^T \\ \boldsymbol{\Phi}\mathbf{\Gamma}^T(\mathbf{I} - \mathbf{B}^T)^{-1} & \boldsymbol{\Phi} \end{bmatrix} \quad [7]$$

gdzie \mathbf{B} to macierz współczynników regresji zmiennych endogenicznych względem zmiennych endogenicznych, $\mathbf{\Gamma}$ to macierz współczynników regresji zmiennych egzogenicznych względem zmiennych endogenicznych, $\boldsymbol{\Phi}$ to macierz kowariancji zmiennych egzogenicznych, a $\boldsymbol{\Psi}$ to macierz kowariancji reszt. Efekty ogólne x na y można zapisać jako

$$\mathbf{T}_{yx} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma} \quad [8]$$

dla efektów y na x

$$\mathbf{T}_{yy} = (\mathbf{I} - \mathbf{B})^{-1} - \mathbf{I} \quad [9]$$

a efekty pośrednie x na y jako

$$\mathbf{I}_{yx} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma} - \mathbf{\Gamma} \quad [10]$$

dla efektów y na y

$$\mathbf{I}_{yy} = (\mathbf{I} - \mathbf{B})^{-1} - \mathbf{I} - \mathbf{B} \quad [11]$$

Efekty bezpośrednie stanowią różnicę pomiędzy efektami ogólnymi a efektami pośrednimi.

Estymacja parametrów modelu polega na oszacowaniu ich wartości, zawartych w wektorze θ tak, aby implikowana macierz kowariancji $\Sigma(\theta)$ jak najlepiej odpowiadała macierzy kowariancji S zaobserwowanej w empirycznej próbie danych, charakteryzujących dany proces lub zjawisko. Efektem procesu szacowania wartości parametrów postulowanego modelu są ich oceny oraz błędy standardowe tych ocen. Należy więc rozpatrzyć zbiór wszystkich możliwych wartości tych parametrów i odnaleźć w nim te, które pozwolą na spełnienie postawionego warunku. Zbiór możliwych rozwiązań jest ograniczony rozmiarami macierzy modelu, które są zdefiniowane przez liczbę zmiennych endogenicznych y i egzogenicznych x , a także przez liczbę zmiennych latentnych η oraz ζ . Mimo, że część tego zbioru jest wyłączona z rozważań, ponieważ nie spełnia założeń (tzw. warunków identyfikacji) postulowanego modelu, to nadal zawiera on ogromną liczbę możliwych rozwiązań. Dlatego w praktyce do poszukiwania właściwego rozwiązania wykorzystuje się iteracyjne metody optymalizacyjne (np. quasi-Newtona, Levenberga-Marquardta, itp.). Metody te oparte są najczęściej na minimalizacji wartości funkcji dopasowania \hat{F}_{min} , która wyznaczana jest zależnie od wybranego estymatora. Do najczęściej wykorzystywanych w tym celu estymatorów należą: estymator największej wiarygodności ML (*maximum likelihood*), estymator nieważonych najmniejszych kwadratów ULS (*unweighted least squares*), estymator uogólnionych najmniejszych kwadratów GLS (*general least squares*) czy też estymator ważonych najmniejszych kwadratów WLS (*weighted least squares*).

Ostatnim etapem analizy SEM jest ocena dopasowania postulowanego modelu do danych empirycznych. Podstawową hipotezę zerową w modelowaniu SEM [2] testuje się z wykorzystaniem statystyki [12], która ma rozkład chi-kwadrat.

$$\chi^2 = (N - 1)\hat{F}_{min} \quad [12]$$

Jednakże, istnieje kilka powodów, dla których ta statystyka nie powinna być stosowana jako kryterium oceny dobroci dopasowania modelu do danych empirycznych, przynajmniej w badaniach rolniczych. Po pierwsze, jest ona bezpośrednią funkcją liczebności próby (N), co sprawia że wynik ogólnego testu dopasowania jest wysoce wrażliwy na wielkość próby, a czułość tej statystyki rośnie wraz ze wzrostem liczby analizowanych przypadków. Zatem szansa na odrzucenie hipotezy zerowej w przypadku analizy modelu dla wielu badanych przypadków jest większa, podczas gdy analiza dla tego samego modelu dla mniejszej liczby przypadków nie pozwalałaby na odrzucenie H_0 . Po drugie, statystyka ta preferuje modele bardziej złożone, czyli takie, w których występuje więcej relacji. I po trzecie, statystyka ta ma skłonność do dawania fałszywie istotnego wyniku w przypadku, gdy postulowany model nie jest bardzo dokładnym odwzorowaniem wszystkich występujących relacji empirycznych, a taka sytuacja ma miejsce bardzo często w badaniach przyrodniczych, a w szczególności rolniczych.

Alternatywą dla testu ogólnego dopasowania są tzw. opisowe miary dopasowania. Wśród nich możemy wyróżnić miary bezwzględnego dopasowania (np. GFI, AGFI, RMR, SRMR), miary relatywnego dopasowania (np. NFI, RFI, NNFI, IFI, RNI, CFI), czy też miary złożoności modelu (np. PNFI, PGFI).

Jeżeli w końcowym etapie, w modelu występują nieistotne relacje pomiędzy zmiennymi lub jego dopasowanie nie jest zadowalające, należy model zmodyfikować. Proces dostosowywania (strojenia) modelu w przypadku analizy SEM może przebiegać podobnie jak w przypadku analizy ścieżek. Dobór metody strojenia modelu zależy od założeń przyjętych przez badacza. Możliwe są również przypadki, gdy badacza interesuje model ze wszystkimi, wstępnie założonymi relacjami.

Konfirmacyjna analiza czynnikowa jako przykład modelu SEM

Poszukiwanie i analiza zmiennych nieobserwowalnych (latentnych) w badaniach z zakresu nauk przyrodniczych nie jest częste. Tymczasem analiza oparta na takich konstruktach teoretycznych, niemierzalnych w rzeczywistych doświadczeniach, umożliwia gruntowne poznanie zasad rządzących badanym zjawiskiem lub procesem. W praktycznych zastosowaniach zmienne latentne są często określane mianem czynników wspólnych, a definiowane są poprzez obserwowalne i mierzalne zachowania, które reprezentują. Tak więc informacje dotyczące zmiennych latentnych są czerpane pośrednio z zaobserwowanych relacji (kowariancji) pomiędzy zmiennymi obserwowalnymi. Analiza czynnikowa jest najczęściej stosowaną analizą służącą badaniu relacji pomiędzy zbiorem danych obserwowalnych, a zmiennymi latentnymi. Rozróżnia się dwa rodzaje analizy czynnikowej: eksploracyjną analizę czynnikową EFA (*Exploratory Factor Analysis*), której zadaniem jest „odkrywanie” struktury czynników wspólnych oraz konfirmacyjną (potwierdzającą) analizę czynnikową CFA

(*Confirmatory Factor Analysis*), która jest statystycznym modelem, charakteryzującym zależność między czynnikami wspólnymi a ich wskaźnikami — zmiennymi obserwowalnymi.

Mimo, iż EFA jest dobrze poznaną i opisaną metodą, stosowaną od wielu lat, jej zastosowanie w doświadczałnictwie rolniczym nie jest powszechne. Podstawowym celem tej analizy jest ujawnienie nieznanymi, ukrytymi źródłami kowariancji istniejących między zmiennymi obserwowalnymi i zdefiniowanie ich w postaci czynników wspólnych (zmiennych latentnych). Natomiast CFA jest testem *a priori* określonego wzorca relacji między zmiennymi obserwowalnymi a czynnikami wspólnymi.

Analizę CFA, poprzez analogię do ogólnego modelu SEM, można schematycznie zapisać za pomocą układu równań

$$\begin{aligned} \mathbf{x} &= \Lambda_x \boldsymbol{\xi} + \boldsymbol{\delta} \\ \mathbf{y} &= \Lambda_y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \end{aligned} \quad [13]$$

gdzie \mathbf{x} jest wektorem obserwowalnych zmiennych egzogenicznych, \mathbf{y} jest wektorem obserwowalnych zmiennych endogenicznych, $\boldsymbol{\xi}$ jest wektorem zmiennych latentnych egzogenicznych, $\boldsymbol{\eta}$ jest wektorem zmiennych latentnych endogenicznych, Λ_x jest macierzą ładunków czynnikowych $\boldsymbol{\xi}$ na \mathbf{x} , Λ_y jest macierzą ładunków czynnikowych $\boldsymbol{\eta}$ na \mathbf{y} , $\boldsymbol{\delta}$ jest wektorem błędów losowych \mathbf{x} , a $\boldsymbol{\varepsilon}$ jest wektorem błędów losowych \mathbf{y} .

W praktyce najczęściej stosuje się zapis uproszczony postaci

$$\mathbf{x} = \Lambda \boldsymbol{\xi} + \boldsymbol{\delta} \quad [14]$$

gdzie \mathbf{x} jest wektorem zmiennych obserwowalnych, $\boldsymbol{\xi}$ jest wektorem czynników wspólnych (latentnych), Λ jest macierzą ładunków czynnikowych, a $\boldsymbol{\delta}$ jest wektorem błędów losowych.

Analiza CFA polega na odnalezieniu parametrów modelu $\boldsymbol{\theta}$ i budowę, implikowanej przez model, macierzy kowariancji $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, która w przypadku tej analizy jest postaci

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \Lambda \boldsymbol{\Phi} \Lambda^T + \boldsymbol{\Theta}_\delta \quad [15]$$

gdzie $\boldsymbol{\Phi}$ jest macierzą kowariancji czynników wspólnych $\boldsymbol{\xi}$, $\boldsymbol{\Theta}_\delta$ jest macierzą kowariancji błędów pomiarowych $\boldsymbol{\delta}$, a Λ jest macierzą ładunków czynnikowych.

Postulowany model CFA jest więc definiowany poprzez określenie odpowiednich parametrów zawartych w macierzach $\boldsymbol{\Phi}$, $\boldsymbol{\Theta}_\delta$ oraz Λ jako nieznanymi i wolnymi do oszacowania oraz określenie pozostałych parametrów modelu jako nieznanymi, ale ograniczonymi tak, aby odpowiadały funkcji wartości innych parametrów modelu lub znanych i ustalonych wartości. Szacowanie nieznanymi parametrów modelu odbywa się poprzez dopasowanie implikowanej macierzy kowariancji $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ do obserwowanej w zbiorze danych macierzy kowariancji \mathbf{S} za pomocą odpowiedniej funkcji dopasowania.

Zastosowanie modelu liniowego z ważonymi warunkami identyfikacji efektów do przekształcania danych

Założenie normalności rozkładu wielozmiennego, dla cech obserwowanych w populacji, jest typowym założeniem większości statystycznych analiz wielozmiennych. Założenie normalności rozkładu danych w analizie SEM pozwala na korzystanie z takich estymatorów, jak ML lub GLS i ULS. Choć uznaje się, że są one dość odporne na brak spełnienia założeń o normalności rozkładu, to jednak jakość wnioskowania jest zależna od tego czy rzeczywisty rozkład w populacji jest zbliżony do normalnego. Model SEM zakłada łączną normalność rozkładu zmiennych obserwowalnych oraz liniowość relacji między zmiennymi w modelu. Jeżeli te warunki są spełnione, macierz kowariancji dla danych empirycznych (\mathbf{S}) jest nieobciążonym estymatorem macierzy kowariancji dla całej populacji ($\boldsymbol{\Sigma}$). Ponadto cała informacja o nieznanymi elementach wektora $\boldsymbol{\theta}$ (parametrach modelu) jest zawarta w macierzy kowariancji \mathbf{S} . Oznacza to, że macierz \mathbf{S} jest statystyką wyczerpującą, gdyż zawiera tyle samo informacji odnośnie wektora $\boldsymbol{\theta}$ co oryginalna macierz danych obserwowalnych \mathbf{X} . Założenie wielozmiennego rozkładu normalnego implikuje rozkład macierzy kowariancji \mathbf{S} , jako rozkład Wisharta (Johnson i Wichern, 1992; Konarski 2009). Właściwości testu Wisharta są podstawą do testowania dopasowania postulowanego modelu do danych empirycznych. Jeżeli założenie wielozmiennego rozkładu normalnego nie jest spełnione, to macierz \mathbf{S} może nie mieć rozkładu Wisharta, a zatem nie będzie statystyką wyczerpującą. W konsekwencji wyniki analizy SEM dla takich danych mogą nie być optymalne.

Brak normalności rozkładu wielozmiennego stwarza zagrożenie poprawności oszacowania parametrów postulowanego modelu oraz jakości wnioskowania, szczególnie jeśli korzystamy z estymatorów ML, GLS lub

ULS. W takiej sytuacji istnieją trzy możliwości postępowania. Pierwszą jest transformacja zmiennych, tak aby ich rozkłady spełniały przyjęte założenia. Drugą jest wykorzystanie próbkowania „bootstrap”, trzecią zaś jest wykorzystanie estymatorów niewrażliwych na rozkład zmiennych (np. WLS) lub wykorzystanie skorygowanych statystyk testowych i skorygowanych błędów standardowych.

Transformacja danych wpływa przede wszystkim na symetrię rozkładu, a w szczególności na jego skośność i kurtozę. Ponadto dobrze dobrana transformacja może zmienić charakter zależności między cechami z nieliniowej na liniową, jak również stabilizować wariancję i eliminować nietypowe obserwacje. W przypadku analizy SEM najczęściej wykorzystuje się dwa rodzaje transformacji danych źródłowych – transformację potęgową Boxa-Coxa (Box i Cox, 1964) oraz rodzinę transformacji z drabiny potęg i pierwiastków Tukeya (Tukey, 1977).

Powyższe transformacje mają zasadniczą wadę: w praktycznym zastosowaniu mogą jedynie prowadzić do „poprawy” rozkładu, zbliżonego pierwotnie do normalnego. Jeżeli zmienna ewidentnie nie ma rozkładu normalnego, co więcej — jeśli jest to, na przykład, cecha nominalna, powyższe transformacje okazują się nieskuteczne. W takiej sytuacji bardzo trudno znaleźć taką transformację, po której zastosowaniu rozkład cechy był choć trochę zbliżony do rozkładu normalnego. Jedną z możliwości rozwiązania tego problemu jest transformacja, wynikająca z rozwiązania modelu liniowego analizy wariancji z ważonymi warunkami identyfikacji efektów. To rozwiązanie mogłoby znaleźć szczególne zastosowanie w przypadku danych niekompletnych (np. pochodzących z doświadczeń w układach nieortogonalnych). Proponowana transformacja wywodzi się z definicji „przekształcenia”, którą można zapisać jako matematyczną projekcję wiedzy (poglądów) o wpływie danej cechy na inną (np. plon) i o jej ilościowym opisie.

Cecha jest to właściwość, która umożliwia rozróżnianie zbiorowości statystycznych (populacji) i może posiadać różne warianty (wartości). Statystyka jest zainteresowana takimi zbiorami wartości, w których występuje zmienność, to znaczy zróżnicowanie jednostek statystycznych ze względu na rozpatrywane cechy. To zróżnicowanie można analizować jedynie w przypadku cech ilościowych (mierzalnych). Metoda analizy wariancji jednoczynnikowej w układzie całkowicie losowym pozwala na analizę porównawczą poziomów badanego czynnika dla wybranej cechy mierzalnej, a model liniowy dla tego zadania jest postaci

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad [16]$$

gdzie: μ — jest średnią ogólną zmiennej ilościowej Y dla wszystkich wariantów czynnika A ; α_i — jest efektem i -tego poziomu czynnika A ; ε_{ij} — jest składnikiem losowym (błędem losowym). Całkowitą zmienność cechy Y można zapisać jako

$$\text{var } Y = \text{var } A + \text{var } E \quad [17]$$

co oznacza, że składają się na nią dwa komponenty: $\text{var } A$ – zmienność wśród wartości średnich dla rozpatrywanych poziomów (obiektów) oraz $\text{var } E$ – zmienność resztową (wewnątrz obiektów). Różnice międzyobiektywne są nieobciążonymi ocenami różnic między wartościami oczekiwanymi badanych obiektów, a istotność tych różnic może być oceniana testem F Fishera (Laudański, 1981). Należy zauważyć, że dla dalszych analiz porównawczych, dla każdej średniej obiektowej, wyznaczany jest odpowiedni przedział ufności, czyli zakres możliwych wartości, dla których przy ustalonym poziomie ufności, wartość średniej obiektowej jest wyrażeniem tendencji wartości tego przedziału.

Zatem rozwiązania modelu liniowego umożliwiają przewartościowanie wartości jakościowych (niemierzalnych) tak, aby spełniały podstawowy warunek cech mierzalnych, mówiący o tym, że między dwiema wartościami cechy można oczekiwać wartości pośrednich. Jednocześnie, zróżnicowanie wartości cechy niemierzalnej w danych warunkach, wyrażane jest przez oczekiwaną ocenę różnic wartości tych wariantów, które reprezentowane są przez odpowiednie średnie wartości cechy Y .

Należy również zwrócić uwagę na fakt, że z równania [17] wynika równość

$$\frac{\text{var } A}{\text{var } Y} + \frac{\text{var } E}{\text{var } Y} = 1 \quad [18]$$

przy czym $\frac{\text{var } A}{\text{var } Y}$ oznacza determinację badanego czynnika (jego wariantów) w oddziaływaniu na zmienność cechy

Y , a $\sqrt{\frac{\text{var } A}{\text{var } Y}}$ wyraża korelację tegoż czynnika z wartościami cechy mierzalnej Y .

Powyższy opis przedstawia pewną koncepcję transformacji cech niemierzalnych na mierzalne. W praktyce możemy jednak mieć do czynienia z wieloma cechami (czynnikami) niemierzalnymi, co może znacznie utrudniać analizę danych. W takim przypadku można zastosować podejście bazujące na liniowym modelu analizy wariancji z ważonymi warunkami identyfikacji efektów, który został szczegółowo zaprezentowany przez Wójcika i Laudańskiego (1989) oraz Laudańskiego (1996). Podejście to polega na wyznaczeniu najlepszego ważonego liniowego nieobciążonego estymatora BWLUE (*Best Weighted Linear Unbiased Estimator*) efektów badanego czynnika postaci (wzory dla układu dwuczynnikowego z interakcją)

$$\hat{a}_i = \sum_{t=1}^a c^{it} q_t \text{ dla } i = 1, 2, \dots, a \quad [19]$$

gdzie $q_t = Y_t - \sum_{j=1}^b n_{tj} \bar{y}_{.j}$, a c^{it} to elementy macierzy odwrotnej do macierzy $\mathbf{C} = [c_{it}]$ o elementach

$$c_{it} = \begin{cases} n_i + \frac{n_i^2}{n_{..}} - \sum_{j=1}^b \frac{n_{ij}^2}{n_j} & \text{dla } i = t \\ \frac{n_i n_t}{n_{..}} - \sum_{j=1}^b \frac{n_{ij} n_{tj}}{n_j} & \text{dla } i \neq t \end{cases} \quad [20]$$

Istotą tego przekształcenia jest przypisanie, do każdego poziomu zmiennej wyrażonej w formie dyskretnej, transformowanej wartości średniej nieobciążonej, wyrażonej jako

$$x_i^{(t)} = \hat{m} + \hat{a}_i, \quad [21]$$

gdzie $x_i^{(t)}$ jest wartością transformowanej zmiennej dla i -tego poziomu jej dyskretnej formy; \hat{m} to średnia ogólna; natomiast \hat{a}_i to BWLUE wyznaczony według wzoru [19].

Rezultatem tej transformacji jest zmienna, posiadająca nowe, praktyczne cechy. Po pierwsze, wszelkie relacje pomiędzy zmiennymi zostają zachowane, podobnie jak wszelkie relacje występujące w ramach zmiennej transformowanej. Jest to możliwe, ponieważ metoda analizy wariancji odpowiada za ocenę relacji pomiędzy poziomami czynnika i wyrażenie jej w postaci obiektywnej miary odległości, zapisanej w formie statystyki F . Po drugie, po transformacji wartości zmiennej są wartościami średnimi, co zgodnie z twierdzeniem Lindeberga-Lévy'ego oznacza, że po transformacji ma ona charakter ciągły i rozkład normalny. Koncepcja tej transformacji, choć ciekawa z numerycznego punktu widzenia, stwarza poważny problem z punktu widzenia merytorycznej interpretacji wyników. Stosowanie tej transformacji, szczególnie przy okazji analizy relacji między cechami, może prowadzić do łączenia pewnych subpopulacji i szacowania dla nich łącznych związków między cechami ilościowymi. Takie podejście, zwłaszcza w obszarze nauk biologicznych, może prowadzić do błędów we wnioskowaniu, znanych pod nazwą paradoksu Simpsona (Simpson, 1951).

Przykład 1 – Badanie struktury składowych postępu biologicznego w hodowli ziemniaka w Polsce w latach 1957-2003

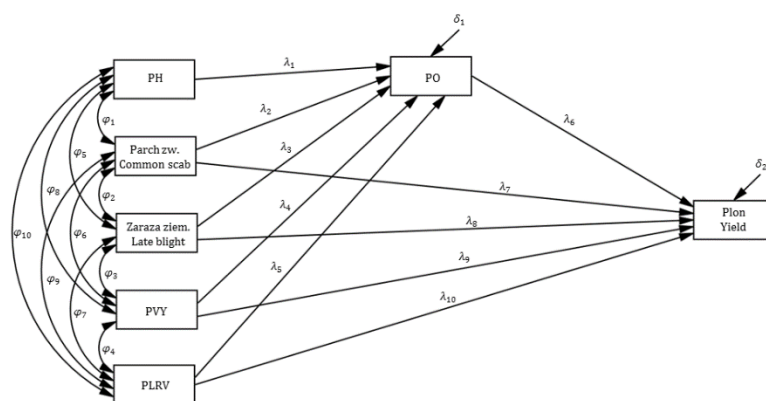
Postęp biologiczny można zdefiniować jako całokształt zmian samoistnych (naturalnych), bądź wynikających z celowej działalności człowieka (antropogenicznych) w organizmach roślinnych i zwierzęcych, wpływających na cechy indywidualne tych organizmów. W sensie rolniczym, jest to ogół zmian, wpływających na wartość technologiczno-użytkową roślin uprawnych i zwierząt gospodarskich (Mańkowski, 2008). Z punktu widzenia rolnika, podstawowym miernikiem postępu biologicznego jest plon roślin uzyskiwany z jednostki powierzchni uprawy. Wielkość plonu decyduje o przychodach rolnika, a więc z praktycznego punktu widzenia, najważniejsze są te składowe postępu biologicznego, które w sposób pośredni lub bezpośredni wpływają na kształtowanie się plonu.

Próbie oceny struktury i wpływu wybranych składowych postępu biologicznego na plon ziemniaków przedstawił Mańkowski (2008). Z metodycznego punktu widzenia ocena ta była obarczona bardzo dużym błędem. Do jej przeprowadzenia wykorzystano model regresji wielorakiej, jednak jak udowodniono, dane nie spełniały założeń analizy (niezależność zmiennych przyczynowych — współliniowość i wzajemne interakcje), co mogło skutkować znacznymi błędami w oszacowaniach wartości parametrów modelu regresyjnego. W takiej sytuacji najlepszym rozwiązaniem było wykorzystanie metody SEM do analizy struktury składowych postępu biologicznego.

W analizie wykorzystano wyniki badań nad postępowaniem biologicznym w hodowli, nasiennictwie oraz produkcji ziemniaka w Polsce, realizowanych w Pracowni Ekonomiki Nasiennictwa i Hodowli Roślin IHAR-PIB w Radzikowie (Mańkowski, 2008). Analizowano między innymi postęp hodowlany (PH), wyrażony ilościowym

wskaźnikiem potencjału plonowania wyhodowanych odmian, postęp odmianowy (PO), wyrażony ilościowym wskaźnikiem potencjału plonowania odmian wpisanych przez COBORU do rejestru krajowego (NLI) oraz postęp jakościowy wyrażony jako średnia odporność odmian, badanych w doświadczeniach COBORU w latach 1957–2003, na zarzę ziemniaka, parcha zwykłego, wirus Y (PVY) oraz wirus liściozwoju (PLRV). Dodatkowo, jako wskaźnik łącznego postępu biologicznego uwzględniono średnie plony ziemniaków, uzyskiwane w doświadczeniach odmianowych z zarejestrowanymi odmianami ziemniaka, prowadzonych przez Ministerstwo Rolnictwa (w latach 50. XX wieku) oraz Stacje Doświadczalnej Oceny Odmian COBORU w latach 1957–2003.

W pierwszym etapie analizy zbudowano model, opisujący strukturę badanych zależności. W analizowanym zbiorze cech, postęp odmianowy uzależniony był zarówno od ilościowej oceny efektów prac hodowlanych (postęp hodowlany), jak i postępu związanego ze zmianami odporności ziemniaka na najważniejsze patogeny. Natomiast średnie plony ziemniaka były uzależnione od postępu odmianowego oraz postępu w hodowli odpornościowej (minimalizowanie strat, spowodowanych przez patogeny). Taki układ zależności przedstawiono za pomocą diagramu ścieżek:



Postulowany model zawierał pięć zmiennych egzogenicznych (PH, odporność na parcha zwykłego, zarzę ziemniaka, wirus Y oraz wirus liściozwoju) i dwie zmienne endogeniczne (PO oraz średnie plony ziemniaka). Wskazano również błędy losowe oceny zmiennych endogenicznych. Relacje między zmiennymi egzogenicznymi i endogenicznymi oraz wzajemne relacje między zmiennymi endogenicznymi oznaczono na diagramie prostymi, jednokierunkowymi strzałkami, których kierunek oznaczał kierunek zależności. Wzajemne relacje pomiędzy zmiennymi egzogenicznymi oznaczono zakrzywionymi, dwukierunkowymi strzałkami.

Rozwiązanie układu równań strukturalnych, czyli elementy wektora θ , wyznaczono metodą iteracyjną z zastosowaniem optymalizacji Levenberga-Marquardta i estymatora największej wiarygodności – ML. Zbieżność osiągnięto już po dwóch iteracjach, przy wartości funkcji dopasowania $F_{ML} = 0,01029$. Przy tak dobranych parametrach modelu, uzyskano wartość statystyki testowej $\chi^2 = 0,4735^{NS}$, co nie dawało podstaw do odrzucenia hipotezy zerowej postaci $H_0: \Sigma = \Sigma(\theta)$. Tak więc nie było podstaw do odrzucenia twierdzenia, że macierz kowariancji, uzyskana dla modelu z parametrami, będącymi elementami wektora θ , jest tożsama z macierzą kowariancji dla badanej populacji. Należy jednak wyjaśnić, że analizę przeprowadzono na stosunkowo niewielkiej liczbie obserwacji ($n = 47$), co z kolei sprzyja uzyskiwaniu nieistotnej wartości statystyki testowej. Dlatego prześledzono również inne wskaźniki dopasowania modelu. Wartości $RMR = 0,0448$ oraz $SRMR = 0,0050$ były bardzo niskie i spełniały kryteria określone przez Hu i Bentlera (1998, 1999). Wartości współczynników $GFI = 0,9971$ oraz $AGFI = 0,9184$ były bardzo wysokie, co także świadczy o dobrym dopasowaniu postulowanego modelu do rzeczywistych danych. Inne statystyki dopasowania również uzyskiwały zadowalające wartości, np. $NFI = 0,9991$ oraz $RFI = 0,9802$. Uznano więc, że postulowany model bardzo dobrze oddaje strukturę relacji zachodzących w obrębie składowych postępu biologicznego w hodowli ziemniaka w Polsce, w latach 1957–2003.

Spśród wyznaczonych współczynników ścieżek nie wszystkie były istotne statystycznie. Na wartości indeksów opisujących postęp odmianowy istotny i bezpośredni wpływ miały postęp hodowlany oraz odporność na wirus Y, a na uzyskiwane średnie plony istotnie wpływały postęp odmianowy oraz odporność na wirus Y. Mimo, iż część analizowanych zmiennych okazała się nie mieć istotnych współczynników ścieżek, postanowiono pozostawić je w modelu, ponieważ wszystkie, z wyjątkiem plonu, uważane są za składowe postępu biologicznego

(Mańkowski, 2008). Zatem, jeśli zadaniem prowadzonej analizy jest określenie struktury wspomnianych składowych względem miernika postępu biologicznego, jakim jest plon, to model nie powinien być redukowany.

Kowariancje między pozostawionymi w modelu zmiennymi egzogenicznymi również poddano testowaniu, które potwierdziło, że wszystkie były istotne statystycznie. Fakt występowania istotnych relacji pomiędzy wszystkimi zmiennymi egzogenicznymi potwierdził dodatkowo zasadność pozostawienia w postulowanym modelu ścieżek nieistotnych statystycznie. Wykrycie istotnych efektów oddziaływania ścieżki, pomimo braku istotności jej współczynnika, może być skutkiem oddziaływań pośrednich (o czym świadczą mogą istotne kowariancje) pomiędzy zmiennymi egzogenicznymi.

Postulowany model determinował postęp odmianowy w około 91%, przy czym największy wpływ na jego kształtowanie się miała odporność na zarazę ziemniaka. Odporność na wirusa Y oraz wirusa liściozwoju miały mniejszy wpływ na kształtowanie się postępu odmianowego, przy czym jego wartość obniżała się wraz ze wzrostem odporności na te wirusy. Najmniejszy wpływ na kształtowanie się postępu odmianowego miała odporność na parcha zwykłego. Wielkość średnich plonów uzyskiwanych w doświadczeniach odmianowych była determinowana przez postulowany model SEM w ponad 65%. Największy wpływ na plony ziemniaka miał postęp odmianowy (efekt bezpośredni), a następnie odporność na wirusa Y (głównie efekt bezpośredni). Mniejszym wpływem na plon cechowała się odporność na zarazę ziemniaka (efekt pośredni). Znacznie mniejszy wpływ na uzyskiwane plony miały postęp hodowlany (efekt pośredni) oraz odporność na wirus liściozwoju (efekt bezpośredni). Odporność na parcha zwykłego wpływała na wysokość uzyskiwanych plonów w najmniejszym stopniu (wpływ pośredni).

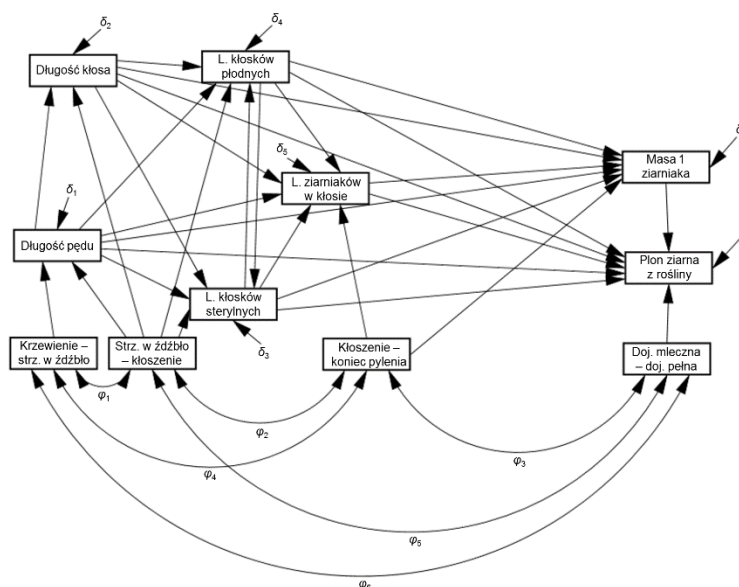
Uzyskane wyniki świadczą o tym, że postulowany model dobrze opisywał strukturę zależności pomiędzy plonem oraz składowymi postępu biologicznego. Analiza funkcji regresji wielokrotnej nie dałaby możliwości odkrycia struktury zależności, a jej rezultaty byłyby obciążone znacznym błędem, wynikającym z wzajemnego powiązania zmiennych egzogenicznych (przyczynowych). Analiza SEM wykazała natomiast, że nie bez znaczenia w ocenie wpływu poszczególnych składowych postępu biologicznego na plon oraz w ocenie wagi poszczególnych składowych postępu biologicznego, są wzajemne relacje pomiędzy jego składowymi (np. kowariancje pomiędzy zmiennymi egzogenicznymi). Za kształtowanie się postępu odmianowego odpowiadał przede wszystkim wzrost odporności na zarazę ziemniaka (postęp w hodowli odpornościowej) oraz postęp ilościowy (wzrost plenności) w hodowli ziemniaka. Za kształtowanie się średnich plonów ziemniaka w doświadczeniach odpowiadał przede wszystkim postęp odmianowy oraz wzrost odporności na wirus Y ziemniaka.

Przykład 2 – Analiza modelu ścieżkowego opisującego zależności przyczynowo-skutkowe pomiędzy plonem ziarna linii podwojonych haploidów pszenicy ozimej a cechami z nim powiązanyymi

Przedmiotem badań były rośliny podwojonych haploidów (DH) pszenicy ozimej, a także rośliny odmian 'Izolda', 'Zyta' oraz 'Astron'. Badaną populacją DH uzyskano na drodze androgenezy w kulturze *in vitro* z izolowanych mikrospor odmian 'Izolda', 'Zyta' oraz 'Astron'. Doświadczenie polowe założono w trzech kolejnych sezonach: 2004/05, 2005/06 oraz 2006/07 na glebie zbielicowanej, czarnej ziemi, klasy bonitacyjnej IIIb/IVa, kompleksu glebowego żyniego bardzo dobrego, na polach doświadczalnych IHAR-PIB w Radzikowie. Ziarniaki kolejnej generacji populacji DH i odmian źródłowych, wysiano po 100 sztuk na poletko o powierzchni 1m². Przedsięwzięcie wykonano nawożenie mineralne wieloskładnikowym nawozem mineralnym Polifoska, pogłównie natomiast rośliny nawożono saletrą amonową. W fazie krzewienia zastosowano chemiczne odchwaszczanie. W okresie wegetacji roślin wykonano opisy ich stadiów rozwojowych wg dziesięcnej skali Zadoksa (Zadoks i in., 1974). W fazie dojrzałości pełnej, z każdego poletka z osobnikami z populacji DH pobierano losowo po 15 roślin. Łącznie przeanalizowano 825 roślin, w tym 225 w sezonie 2004/2005, 570 w sezonie 2005/2006 oraz 30 roślin w sezonie 2006/2007. Dla każdego pojedyńka zmierzono długość najdłuższego pędu w roślinie oraz długości kłosów, ogólną liczbę kłosów w kłosie z uwzględnieniem liczby kłosów sterylnych i płodnych, liczbę ziarniaków w kłosie oraz masę ziarniaków ze wszystkich kłosów w roślinie. Jednostkową masę ziarniaka stanowił iloraz masy ziarniaków z rośliny i liczby ziarniaków z rośliny.

Ze względu na złożoną strukturę relacji między analizowanymi zmiennymi oraz na podstawie przesłanek merytorycznych i logicznych, zbudowano model ogólny, który uwzględniał potencjalne relacje występujące

w analizowanym zbiorze cech oraz ontogenetyczny porządek kształtowania się poszczególnych cech. Model ten zapisano w postaci diagramu:

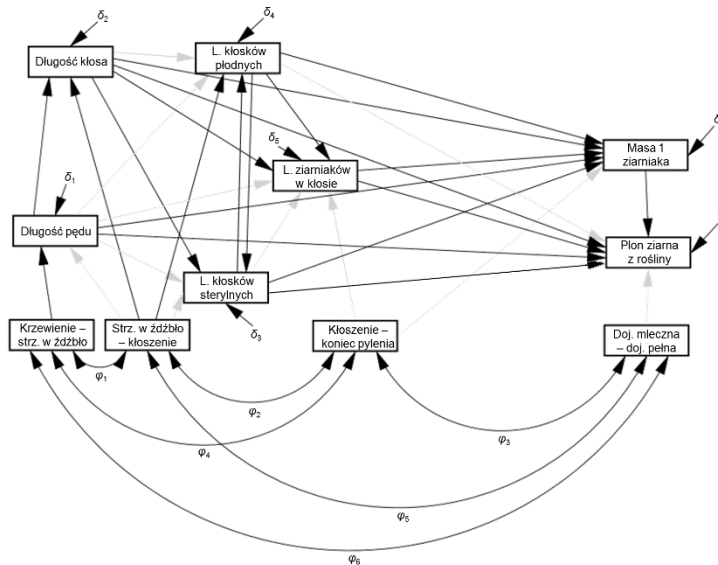


W tym eksperymencie nie zastosowano klasycznego układu czynnikowego, ponieważ w każdym roku badano inną grupę roślin. Ten specyficzny układ doświadczalny wymusza zastosowanie specjalnego podejścia do analizy SEM, ponieważ w modelu nie można pominąć czynnika lat, który odpowiadał za losowy układ warunków pogodowych, co przekładało się na kształtowanie poszczególnych cech (np. czas trwania poszczególnych faz rozwojowych) i ich wzajemnych relacji. W tym przypadku uzasadnionym wydaje się wykorzystanie podejścia opisanego przez Shipleya (2002), zwanego wielogrupową analizą SEM (*multigroup SEM analysis*). Jeżeli dane pochodzą z jednej populacji ogólnej, lecz z różnych subpopulacji oraz podejrzewa się, że układ relacji przyczynowo-skutkowych w tych subpopulacjach może choćby nieznacznie się różnić, bądź układ pozostaje ten sam, lecz różna jest waga poszczególnych relacji, wówczas łączna analiza takich danych za pomocą jednego wspólnego modelu SEM może prowadzić do fałszywych wniosków, w tym do odrzucenia całego modelu. Stąd, bardziej uzasadnionym podejściem jest analizowanie modelu SEM oddzielnie dla każdej subpopulacji. Jeżeli różnice między subpopulacjami nie zostaną wykazane, wówczas uzyskane dla nich modele będą identyczne, w przeciwnym wypadku każdej subpopulacji będzie odpowiadał inny model. W przypadku stwierdzenia występowania w modelu nieistotnych statystycznie relacji, model dla danego roku był redukowany tak, aby wyrażał jedynie istotne relacje przyczynowo-skutkowe. Ze względu na liczbę dostępnych obserwacji w każdym roku badań, analizie poddano jedynie dane z lat 2004/05 i 2005/06.

W pierwszym etapie, analizę SEM przeprowadzono dla kompletnego modelu zależności, osobno dla danych z sezonów 2004/05 oraz 2005/06. W obliczeniach wykorzystano odchylenia od wartości średnich analizowanych cech. W analizie zastosowano metodę optymalizacji Levenberga-Marquardta oraz estymator ML do oszacowania wartości parametrów modelu. Wartości początkowe dla algorytmu optymalizacji dobrano metodą McDonalda (McDonald i Hartmann, 1992). W przypadku danych z sezonu 2004/05 zbieżność osiągnięto po 31 iteracjach, a w przypadku danych z sezonu 2005/06 po 46 iteracjach. Wstępna analiza statystyk dopasowania postulowanego modelu wykazała, że ogólny model zależności wymaga korekty i dostosowania do struktury zależności, specyficznej dla każdej z analizowanych subpopulacji.

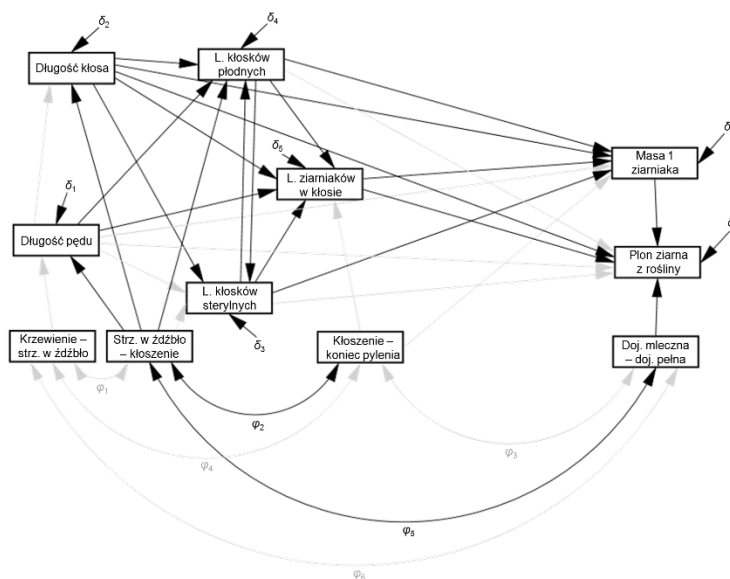
Kolejnym etapem analizy była więc krokowa redukcja postulowanego modelu, którą realizowano w identyczny sposób, osobno dla każdej subpopulacji. W każdym kroku z modelu usuwano relację, cechującą się najmniejszą bezwzględną wartością statystyki testowej. Następnie ponownie analizowano zaktualizowany model. Czynności te powtarzano do momentu uzyskania modelu, w którym wszystkie wartości statystyk testowych wskazywały na występowanie istotnych wartości współczynników ścieżek, wariacji błędów losowych oraz kowariancji między zmiennymi.

Ostateczny, zredukowany model zależności, dla obserwacji z sezonu wegetacyjnego 2004/05, przedstawiono w formie diagramu ścieżek, w którym szare strzałki oznaczają relacje usunięte z wyjściowego modelu.



W przypadku tego modelu, kryterium konwergencji uzyskano po 24 iteracjach, statystyki dopasowania były wyraźnie lepsze w porównaniu do modelu wyjściowego, jednak wciąż nie w pełni satysfakcjonujące. Wszystkie oszacowane współczynniki ścieżek były istotne statystycznie, podobnie jak wariancje dla błędów losowych i kowariancje pomiędzy zmiennymi egzogenicznymi. Wyniki uzyskane z doświadczenia w sezonie wegetacyjnym 2004/05 wskazywały, że głównymi czynnikami, kształtującymi wysokość plonu ziarna z rośliny pszenicy ozimej były długość kłosa i liczba ziarniaków w kłosie. Liczba ziarniaków w kłosie była determinowana przede wszystkim przez długość kłosa oraz liczbę kłosków płodnych w kłosie. Długość kłosa była zależna od czasu między początkiem fazy strzelania w źdźbło a końcem fazy kłoszenia. Stwierdzono również występowanie istotnych relacji między czasem trwania poszczególnych faz rozwojowych. Owe relacje wynikały z układu warunków pogodowych, więc były ściśle powiązane z rokiem badań. Ponadto zaobserwowano, że relacja zwrotna pomiędzy liczbą kłosków płodnych i sterylnych w kłosie pełni ważną rolę w modelu. Relacja ta była bardzo silna, o czym świadczyć może bardzo wysoka wartość standaryzowanego współczynnika ścieżek oraz stosunkowo wysoka wartość wariancji błędu losowego. Analiza uzyskanych współczynników determinacji wykazała, że postulowany model ścieżkowy opisywał zmienność plonu ziarna z rośliny pszenicy ozimej w sezonie 2004/05 roku w około 64%.

Ostateczny, zredukowany model zależności, dla obserwacji z sezonu wegetacyjnego 2005/06, przedstawiono w formie diagramu ścieżek :



W przypadku tego modelu, kryterium konwergencji uzyskano po 39 iteracjach. Statystyki dopasowania były lepsze niż w przypadku modelu ogólnego, jednak wciąż nie w pełni satysfakcjonujące. Wszystkie oszacowane

współczynniki ścieżek były istotne statystycznie, podobnie jak wariancje dla błędów losowych i kowariancje pomiędzy zmiennymi egzogenicznymi. Wyniki uzyskane z doświadczenia w sezonie wegetacyjnym 2005/06 wskazywały, że cechą, która w największym stopniu wpływała na kształtowanie się plonu ziarna z rośliny, była masa jednego ziarniaka. O wielkości masy pojedynczego ziarniaka decydowały głównie liczba kłosek sterylnych w kłosie, liczba kłosek płodnych w kłosie oraz liczba ziarniaków w kłosie. Stwierdzono również występowanie istotnych współzależności między czasem trwania fazy od strzelania w źdźbło do kłoszenia i długością fazy od kłoszenia do końca pylenia, a także między czasem trwania fazy od strzelania w źdźbło do kłoszenia i długością fazy od dojrzałości mleczej do dojrzałości pełnej. Analiza uzyskanych współczynników determinacji wykazała, że postulowany model ścieżkowy opisywał zmienność plonu ziarna z rośliny pszenicy ozimej w sezonie 2005/06 roku w około 66%.

Wykonanie analizy modelu dla danych z sezonu wegetacyjnego 2006/07 nie było możliwe, ponieważ w tym roku badano tylko trzy linie, co dało w sumie 45 obserwacji. Liczba obserwacji była zatem niższa niż liczba współczynników, szacowanych w analizie wyjściowego modelu. Przy tak małej liczbie obserwacji nie było również możliwe opracowanie poprawnego merytorycznie i numerycznie modelu zredukowanego.

Powyższe rezultaty wyraźnie wskazują na znaczący wpływ roku doświadczeń na wyniki, uzyskane z analizy modeli ścieżkowych zależności plonu ziarna z rośliny pszenicy ozimej i cech plonotwórczych. W przypadku modelu dla danych, uzyskanych w sezonie 2004/05, głównymi czynnikami wpływającymi na kształtowanie się plonu ziarna z rośliny były długość kłosa oraz liczba ziarniaków w kłosie. Tymczasem w przypadku modelu dla danych, uzyskanych w sezonie 2005/06 czynnikiem odpowiedzialnym za kształtowanie się plonu ziarna z rośliny była przede wszystkim masa jednego ziarniaka. Zatem uzyskane wyniki nie były w pełni zgodne, co prowadzi do pewnych problemów z ich interpretacją i określeniem, które cechy plonotwórcze w zdecydowanym największym stopniu odpowiadały za kształtowanie się plonu ziarna z rośliny. Wyniki te mogą jednak prowadzić do wnioskowania na temat kształtowania się cech, warunkowanych wpływem środowiska.

Przykład 3 – Wykorzystanie konfirmacyjnej analizy czynnikowej (CFA) do oceny oddziaływania głównych czynników produkcji na plon w uprawie pszenicy ozimej

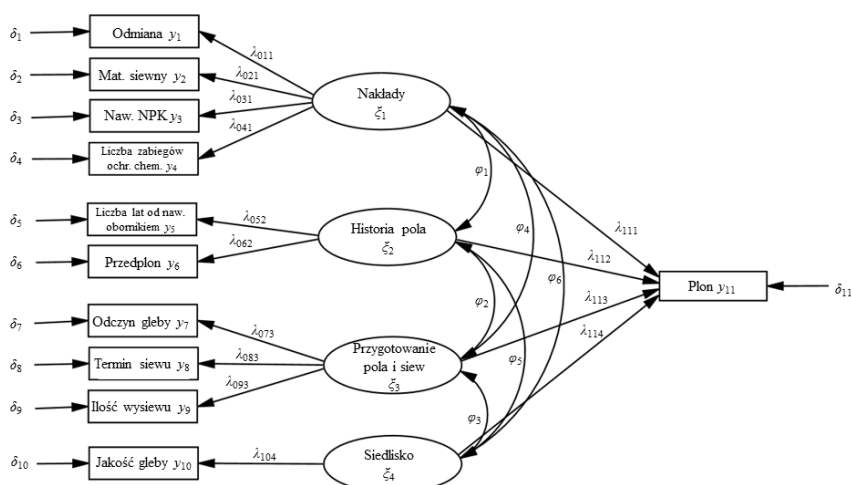
Materiał do badań stanowiły dane pochodzące z badań ankietowych indywidualnych gospodarstw rolniczych, prowadzących rachunkowość rolną dla potrzeb Instytutu Ekonomiki Rolnictwa i Gospodarki Żywnościowej. Badania zostały przeprowadzone przez Pracownię Ekonomiki Nasiennictwa i Hodowli Roślin Instytutu Hodowli i Aklimatyzacji Roślin w Radzikowie. Corocznie inspektorzy rachunkowości rolnej IERiGŻ ankietowali od 420 do 620 indywidualnych gospodarstw rolniczych. Analizowane dane obejmowały lata 1992–2003 i dotyczyły pól uprawnych pod pszenicą ozimą. W badanym okresie było to łącznie 6 290 pól, ale z tej puli usunięto rekordy niepewne, w których wartości cech nie były wiarygodne lub zawierające braki danych oraz dotyczące odmian reprezentowanych nielicznie. W dalszych analizach wykorzystano przypadki opisujące łącznie 3 980 pól uprawnych pszenicy ozimej. Spośród wszystkich 53 cech, charakteryzujących każde pole uprawne, do analizy wybrano dziesięć cech, które na podstawie oceny merytorycznej i analiz wstępnych uznano za najważniejsze determinanty plonu ziarna pszenicy ozimej. Do tej grupy należały: odmiana, kategoria i stopień kwalifikacji materiału siewnego, rodzaj przedplonu, liczba lat od ostatniego nawożenia obornikiem, jakość stanowiska – klasa gleby, odczyn gleby, nawożenie mineralne NPK, termin siewu, ilość wysiewu oraz liczba zabiegów chemicznej ochrony roślin. Wybrane czynniki agrotechniczne miały charakter cech jakościowych oraz ilościowych. Inne cechy (odmiany, klasa gleby, stopień kwalifikacji oraz rodzaj przedplonu) przyjmowały wartości (poziomy czynnika) o konkretnych nazwach. Cechy te, określane jako nominalne, były ewidentnie jakościowe skategoryzowane. Część cech (liczba lat od ostatniego nawożenia obornikiem, odczyn gleby, liczba zabiegów chemicznej ochrony roślin) była cechami ilościowymi, ale przyjmowała tylko określone wartości. Cechy te były cechami ilościowymi skokowymi. Pozostałe cechy ilościowe (nawożenie mineralne, ilość wysiewu, termin siewu) były reprezentowane przez wartości (kategorie) wyraźnie uporządkowane, które występowały w takiej ilości różnych wariantów, że można było je uznać za cechy o charakterze ciągłym. Z racji takiej różnorodności cech oraz z powodu ich wyraźnej zmienności w latach badań ankietowych, dane poddano transformacji z wykorzystaniem modelu liniowego z ważonymi warunkami identyfikacji efektów. Każdą cechę ilościową traktowano jak jakościową przez dyskretyzację czynnika ilościowego, tzn. kolejne wartości tej cechy stawały się kolejnymi poziomami czynnika. W ten sposób każdy czynnik uzyskał charakter cechy jakościowej skategoryzowanej. Za ocenę wartości plonotwórczej poziomu (kategorii) czynnika przyjęto odpowiadającą mu

wielkość średniego plonu pszenicy, uzyskanego na podstawie nieobciążonych, ważonych estymatorów efektów tego czynnika, przy czym drugi czynnik stanowiły lata badań. Innymi słowy ocenami wartości plonotwórczej poziomów czynnika były nieobciążone wartości analizowanych kategorii tego czynnika, wyrażone w relatywnym, nieobciążonym średnim plonie pszenicy ozimej, uzyskiwanym przy danym poziomie czynnika. W wyniku takiej transformacji wszystkie cechy były wyrażone w sposób ilościowy i mogły być uznane za cechy o charakterze ciągłym.

Do oceny wpływu rozpatrywanych cech na plon pszenicy ozimej, wykorzystano analizę regresji wielorakiej, której wynik wykazał, że wszystkie cechy pozostawały w istotnej relacji z obserwowanym plonem. Współczynnik determinacji plonu pszenicy ozimej przez cechy plonotwórcze wynosił 41%, należy jednak pamiętać, że wyniki badań ankietowych obarczone są dużym błędem losowym, wynikającym z charakteru i specyfiki tych badań. Zatem uzyskany wynik należy uznać za zadowalający.

W celu wyodrębnienia głównych czynników produkcji przeprowadzono eksploracyjną analizę czynnikową metodą składowych głównych Hotelling'a z normalizacją Kaiser'a i rotacją ortogonalną *varimax*. W wyniku analizy wyodrębniono 4 główne czynniki produkcji, odpowiadające czynnikom wspólnym, które wyjaśniały kolejno 19,5%, 12,7%, 12,6% oraz 10,3% zmienności obserwowanej w zbiorze danych. Rodzaj i charakter wyodrębnionych czynników zidentyfikowano na podstawie wartości ładunków czynnikowych dla analizowanych cech. Pierwszy czynnik wspólny budowały, przede wszystkim, cztery cechy plonotwórcze: odmiana, rodzaj materiału siewnego, nawożenie mineralne NPK oraz liczba zabiegów chemicznej ochrony roślin. Ponieważ wszystkie te cechy są powiązane z bezpośrednimi kosztami, ponoszonymi przez rolnika, pierwszy główny czynnik produkcji w uprawie pszenicy ozimej określono jako „nakłady”. Drugi czynnik budowały cechy, związane z przeszłą działalnością rolniczą, prowadzoną na danym polu uprawnym: rodzaj przedplonu oraz liczba lat od ostatniego nawożenia obornikiem. Zatem ten czynnik produkcji określono mianem „historii pola”. Na trzeci czynnik wspólny składały się przede wszystkim: odczyn gleby, termin siewu oraz ilość wysiewu. Czynnik ten określono jako „przygotowanie pola i siew”. Czwarty główny czynnik produkcji określono jako „siedlisko”, ponieważ był najsilniej powiązany z jakością gleby. Główne czynniki produkcji poddano analizie regresji wielorakiej, aby określić ich wpływ na kształtowanie się plonu ziarna pszenicy ozimej, w ankietowanych gospodarstwach. Wyniki analizy wykazały, że plon determinowany był najsilniej przez czynnik „nakłady”, nieco słabiej przez czynnik „przygotowanie pola i siew”, a najsłabiej przez dwa pozostałe czynniki. Łączny współczynnik determinacji plonu pszenicy ozimej przez główne czynniki produkcji wynosił 36,8%.

Z uwagi na występowanie wzajemnych powiązań pomiędzy głównymi czynnikami produkcji oraz w celu lepszego określenia ich oddziaływania na plon pszenicy ozimej, dane ankietowe zostały poddane analizie CFA. Model CFA, uwzględniający wzajemne relacje pomiędzy czynnikami wspólnymi oraz relacje pomiędzy nimi a plonem, przedstawiono w formie diagramu ścieżek:



Wartości parametrów modelu oszacowano podobnie, jak w przykładzie drugim, a zbieżność algorytmu optymalizacyjnego osiągnięto po 11 iteracjach.

W wyniku tej analizy stwierdzono, że czynnik „nakłady” najmocniej wpływał na nawożenie mineralne NPK oraz liczbę zabiegów chemicznej ochrony roślin, a nieco słabiej na dobór odmiany oraz kategorię i stopień

kwalifikacji materiału siewnego. Czynniki „historia pola” wiązały się najsilniej z przedplonem, a nieco słabiej z liczbą lat od ostatniego nawożenia obornikiem. Czynniki „przygotowanie pola i siew” wpływały najsilniej na ilość wysiewu, słabiej na termin siewu, a najsłabiej na odczyn gleby. Relacja pomiędzy czynnikiem „siedlisko” a klasą gleby również okazała się istotna. Plon ziarna pszenicy ozimej kształtował się najsilniej pod wpływem czynnika „nakłady”, nieco słabszy efekt na plon miał czynnik „przygotowania pola i siew”, a najsłabiej na plon wpływały czynniki „siedlisko” oraz „historia pola”. Ponadto, w przypadku pierwszych trzech czynników, relacja miała charakter wprost proporcjonalny, a w przypadku „historii pola” relacja była odwrotnie proporcjonalna. Wszystkie kowariancje, wskazane w diagramie ścieżek, były istotne statystycznie, co oznacza, że wszystkie relacje między czynnikami wspólnymi były znaczące i mogą mieć duże znaczenie w poznaniu struktury czynnikowej procesu produkcji pszenicy ozimej. Powyższe wyniki były podstawą do stwierdzenia, że niebagatelną rolę w analizie strukturalnej czynników produkcji pszenicy ozimej pełnią wzajemne oddziaływania pomiędzy nimi.

W strukturze opisanej przez postulowany model CFA, determinacja plonu ziarna przez główne czynniki produkcji pszenicy ozimej, wynosiła ponad 68%. Zatem różnica (wynosząca nieco ponad 31%) w determinacji plonu wynikająca z bezpośredniego oddziaływania ortogonalnych czynników produkcji, jest wynikiem występowania wzajemnych powiązań pomiędzy tymi czynnikami. Patrząc ogólniej na postulowany model, można wnioskować, że cechy wyjściowe (10 czynników plonotwórczych) są zewnętrzną, obserwowalną manifestacją działania czynników wspólnych, a skoro czynniki te kształtują plon, to znaczy, że jest on pośrednio determinowany przez te cechy. Ponadto ich oddziaływanie nie jest addytywne, lecz bardziej złożone, gdyż wynika z wzajemnych relacji i oddziaływań pomiędzy samymi cechami plonotwórczymi.

Takie podejście do analizy struktury czynnikowej wskazuje, że analiza EFA umożliwia pogrupowanie analizowanych zmiennych, a następnie wyodrębnienie i zdefiniowanie głównych czynników produkcji pszenicy ozimej. Analiza CFA umożliwia natomiast wykonanie pełnej charakterystyki i opisu tych czynników oraz określenie relacji pomiędzy nimi a plonem ziarna.

Podsumowanie

Jak wskazują Konishi i Kitagawa (2008) – metody modelowania statystycznego są kluczowym narzędziem, wykorzystywanym w badaniach naukowych. Modele statystyczne są formą zrozumiałego opisu struktury złożonych i często nie do końca poznanych zjawisk oraz umożliwiają odkrywanie występujących w nich zależności przyczynowo-skutkowych. Narzędzia te nie są jednak doskonałe, a każda metoda ma wady i znane ograniczenia. Metoda analizy modeli równań strukturalnych (SEM), mimo iż nie jest narzędziem uniwersalnym, zdaje się być bardzo wartościową alternatywą dla powszechnie stosowanych metod statystycznego modelowania zjawisk i procesów przyrodniczych. Metoda ta daje możliwości

- budowy i analizy bardzo złożonych modeli zależności przyczynowo-skutkowych, uwzględniających relacje różnego typu, w tym relacje rekurencyjne oraz pętle sprzężeń zwrotnych;
- uwzględnienia zmiennych latentnych, nie tylko do modelu confirmacyjnej analizy czynnikowej, ale również w klasycznym modelu ścieżkowym;
- uwzględnienia w modelu powiązań między zmiennymi egzogenicznymi, latentnymi oraz błędami losowymi dla zmiennych endogenicznych;
- zastosowania, innych niż metoda najmniejszych kwadratów, sposobów aproksymacji parametrów modelu;
- analizowania danych, które z powodu niespełnienia założeń co do rozkładu, nie mogą być poddane analizie innymi metodami;
- oceny dopasowania modelu, nie tylko na podstawie testu statystycznego, ale również w oparciu o różne indeksy dopasowania.

Oprócz oczywistych zalet, metoda analizy SEM posiada również istotne wady. Podstawowym problemem, związanym z praktycznym zastosowaniem tej metody, jest jej złożoność, zarówno w sensie numerycznym jak i użytkowym. Zaawansowany aparat matematyczny, towarzyszący tej metodzie wymaga od badacza dobrego przygotowania matematyczno-statystycznego. Z kolei prawidłowy dobór modelu, wybór strategii analizy oraz właściwa interpretacja uzyskanych wyników, wymagają nie tylko świetnej znajomości analizowanego procesu lub zjawiska, lecz również dobrego rozeznania w metodyce analizy SEM. Kolejnym istotnym problemem jest

ontologiczny charakter prawdziwości testowanego modelu. Jak podaje Konarski (2009), w przypadku badań empirycznych należy przyjąć, że postulowany w analizie SEM model, jedynie w przybliżeniu reprezentuje procesy stochastyczne, co oznacza, że hipoteza ścisłego dopasowania modelu jest *a priori* fałszywa. W efekcie jedyną realistyczną hipotezą, jaką powinno się testować, jest hipoteza bliskiego dopasowania. Częstym następstwem tego problemu jest odrzucenie hipotezy o dopasowaniu postulowanego modelu do danych, opisujących badane zjawisko lub proces, na podstawie testu chi-kwadrat. Dlatego należy przyjąć, że w przypadku badań z zakresu nauk rolniczych, to statystyki (indeksy) dopasowania powinny być podstawą do wnioskowania o stopniu dopasowania postulowanego modelu do danych empirycznych, natomiast test szczegółowy powinien być traktowany jedynie jako argument dodatkowy, ale nie wymagany.

Przedstawione przykłady zastosowania analizy modeli strukturalnych, w połączeniu z jej teoretycznym opisem, stanowią opracowanie metodyczno-aplikacyjne, przybliżające metodę badania zależności przyczynowo-skutkowych, zwłaszcza o złożonym charakterze, która jest mało znana na gruncie doświadczalnictwa rolniczego.

Na znaczenie poruszanego zagadnienia uwagę zwrócili również recenzenci wydawniczy monografii, prezentującej moje osiągnięcie naukowe. Prof. dr hab. Zygmunt Kaczmarek zaznaczył w swojej recenzji, że „podjęty temat jest bardzo interesujący i niezwykle poznawczy dla doświadczalnictwa rolniczego, ze względu na możliwość zastosowania proponowanej metody do analizy danych z doświadczeń rolniczych. Oprócz wnikliwego opisu samej metody i jej weryfikacji eksperymentalnej, Autor przedstawia również oryginalne rozwiązanie dotyczące metody przekształcania danych, nie spełniających warunków koniecznych do zastosowania analizy SEM. Praktyczne znaczenie proponowanej transformacji, opartej na modelu liniowym z ważonymi warunkami identyfikacji efektów, znacznie wykracza poza przydatność w analizie modeli równań strukturalnych. Może ona znaleźć szczególne zastosowanie w przypadku danych niekompletnych.” Ponadto Recenzent zaznaczył, że „modele równań strukturalnych mogą stać się, obok analizy regresji i analizy ścieżek, podstawową metodą badania zależności przyczynowo-skutkowych w naukach rolniczych”. Natomiast dr hab. Marcin Kozak zaznaczył, że „choć zawsze byłem sceptykiem możliwości zastosowania analizy czynnikowej w naukach rolniczych, przyznaję, że pomysł w przykładzie trzecim jest ciekawy. Ten kierunek mi się podoba i odnoszę wrażenie, że z takiego modelu można by wyciągnąć bardzo ciekawe wnioski.”

Do najważniejszych elementów przedstawionego osiągnięcia naukowego można zaliczyć:

1. Przedstawienie metodyki SEM oraz sposobów wykorzystania tej metody w analizie złożonych zależności przyczynowo-skutkowych, w badaniach z dziedziny nauk rolniczych.
2. Przybliżenie sposobu budowy wyjściowego modelu badanych zależności oraz przedstawienie sposobów ulepszania (strojenia) badanego modelu w trakcie analizy.
3. Wykazanie, że test chi-kwadrat, stosowany do testowania hipotezy $H_0: \Sigma = \Sigma(\theta)$ nie powinien być traktowany jako podstawowe kryterium oceny poprawności postulowanego modelu w przypadku analizy zależności przyczynowo-skutkowych w naukach rolniczych.
4. Przedstawienie praktycznego wykorzystania metodyki SEM w analizie danych, pochodzących z różnych doświadczeń i badań rolniczych, w tym w szczególności:
 - wykorzystania metodyki SEM w badaniu struktury postępu biologicznego w hodowli ziemniaka i porównanie wyników z ocenami parametrów modelu, uzyskanego w analizie regresji wielorakiej;
 - budowa i analiza modelu ścieżkowego zależności plonu ziarna pszenicy ozimej od czynników warunkujących plon, z wykorzystaniem relacji nierekurencyjnej i analizy wielogrupowej;
 - wykorzystanie metody SEM do przeprowadzenia analizy CFA na danych rolniczych i wskazanie głównych czynników produkcji pszenicy ozimej oraz określenie wpływu tych czynników na kształtowanie się plonu ziarna.
5. Zaproponowanie nowatorskiego podejścia do transformowania danych empirycznych w oparciu o model liniowy z ważonymi warunkami identyfikacji efektów, bazujący na estymatorze BWLUE.

Literatura cytowana w autoreferacie

- Box G. E. P., Cox D. R. 1964. An analysis of transformation (with discussion). *Journal of the Royal Statistical Society, Series B*, 26: 211—252.
- Eaton G. W. 1986. Two-dimensional partitioning of yield variation. *HortScience*, 21(4): 1052—1053.
- Gołaszewski J. 1996. A method of yield component analysis. *Biometrical Letters*, 33(2): 79—88.
- Gołaszewski J., Idżkowska M., Milewska J. 1998. The TDP method of seed yield component analysis in grain legume breeding. *Journal of Applied Genetics*, 39(4): 299—308.
- Gozdowski D., Mądry W. 2008. Charakterystyka i empiryczne porównanie prostej oraz złożonej analizy ścieżek w ocenie determinacji plonu roślin przez ich cechy plonotwórcze Część I. Prezentacja stosowanych metod. *Biuletyn IHAR*, 249: 109—124.
- Hu L.-T., Bentler P. M. 1998. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4): 424—453.
- Hu L.-T., Bentler P. M. 1999. Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1): 1—55.
- Janaszek M., Mańkowski D. R., Kozdój J. 2011. Sieci neuronowe MLP w prognozowaniu plonu jęczmienia jarego. *Biuletyn IHAR*, Nr 259, 93—112.
- Johnson R. A., Wichern D. W. 1992. *Applied multivariate statistical analysis*. 2nd edition. Prentice Hall, Englewood Cliffs, NJ.
- Jöreskog K. G. 1973. A general method for estimating a linear structural equation system. In: Goldberger A. S., Duncan O. D. (eds.). *Structural Equation Models in the Social Sciences*. Academic Press, New York: 83—112.
- Kaczmarek Z., Mańkowski D. R. 2011. Wprowadzenie do statystycznych analiz wielozmiennych. Część I. Podstawy teoretyczne. *Biuletyn IHAR*, 259: 23—34.
- Keesling W. 1972. Maximum likelihood approaches to causal flow analysis. Ph.D. thesis, University of Chicago.
- Konarski R. 2009. *Modele równań strukturalnych — teoria i praktyka*. Wydawnictwo Naukowe PWN, Warszawa.
- Konishi S., Kitagawa G. 2008. *Information Criteria and Statistical Modeling*. Springer Science + Business Media, NY, USA.
- Kozak M. 2006. Two-dimensional partitioning of yield variation: A critical note. *Plant Breeding and Seed Science*, 53: 37—42.
- Kozak M. 2008. Analiza związków przyczynowo-skutkowych w agronomii. *Rozprawa habilitacyjna*, SGGW, Warszawa.
- Kozak M. 2011. Analiza związków przyczynowo-skutkowych w agronomii i hodowli roślin. *Biuletyn IHAR*, 259: 3—21.
- Krzyško M., Wołyński W., Górecki T., Skorzybut M. 2008. *Systemy uczące się*. WNT, Warszawa.
- Laudański Z. 1981. Analiza statystyczna plonów pszenicy ozimej w warunkach produkcyjnych na podstawie danych ankietowych. *Hodowla Roślin Aklimatyzacja i Nasiennictwo*, T.25, Z.5/6: 275—288.
- Laudański Z. 1996. Zastosowania operatorów rzutowania w analizowaniu danych nieortogonalnych.
- Mańkowski D. R. 2008. Postęp biologiczny w hodowli, nasiennictwie oraz produkcji ziemiaka w Polsce. *Rozprawa doktorska*, IHAR-PIB, Radzików.
- Mądry W., Kozak M. 2000. Analiza ścieżek i sekwencyjna analiza plonu w badaniach zależności plonu od cech łanu. Cz. I. Opis metod. *Roczniki Nauk Rolniczych, Seria A*, Nr 115: 143—157.
- McDonald R. P., Hartmann W. 1992. A Procedure for Obtaining Initial Values of Parameters in the RAM Model. *Multivariate Behavioral Research*, 27: 57—176.
- O'Neal M. R., Engel B. A., Ess D. R., Frankenberger J. R. 2002. Neural network prediction of maize yield using alternative data coding algorithms. *Biosystems Engineering*, 83: 31—45.
- Rawlings J. O., Pantula S. G., Dickey D. A. 2001. *Applied regression analysis — a research tool*. 2nd Edition. Springer-Verlag, New York.
- Shipley B. 2002. *Cause and correlation in biology. A user's guide to path analysis, structural equations and causal inference*. Cambridge University Press, Cambridge.
- Simpson E. H. 1951. The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 13, No. 2: 238—241.
- Spearman C. 1904. 'General intelligence' objectively determined and measured. *American Journal of Psychology*, 15: 201—293.
- Teoria, algorytmy i przykłady. Wydawnictwo SGGW, Warszawa.
- Tukey J. W. 1977. *Exploratory data analysis*. Addison-Wesley, Reading, MA.
- Wiley D. E. 1973. The identification problem for structural equation models with unmeasured variables. In: Goldberger A. S., Duncan O. D. (eds.). *Structural Equation Models in the Social Sciences*. Academic Press, New York: 69—83.
- Wójcik A. R., Laudański Z. 1989. *Planowanie i wnioskowanie statystyczne w doświadczalnictwie*. PWN, Warszawa.
- Wright S. 1921. Correlation and causation. *Journal of Agricultural Research*, 20: 557—585.
- Wright S. 1923. The theory of path coefficients — a reply to Niles's criticism. *Genetics*, 8: 239—255.
- Wright S. 1934. The method of path coefficients. *Annals of Mathematical Statistics*, 5: 161—215.
- Zadoks J. C., Chang T. T., Konzak C. F. 1974. A decimal code for the growth stages of cereals. *EUCARPIA Bulletin*, Vol. 7: 42—52.

E. OMÓWIENIE POZOSTAŁYCH OSIĄGNIĘĆ NAUKOWO BADAWCZYCH

E.1. ZESTAWIENIE DOROBKU PUBLIKACYJNEGO PRZED I PO UZYSKANIU STOPNIA DOKTORA

| | Przed uzyskaniem stopnia doktora | Po uzyskaniu stopnia doktora | Łącznie |
|---|----------------------------------|------------------------------|---------|
| Prace opublikowane w recenzowanych czasopismach krajowych | 15 | 24 | 39 |
| Prace opublikowane w recenzowanych czasopismach zagranicznych | - | 8 | 8 |
| Monografie i rozprawy naukowe | - | 2 | 2 |
| Rozdziały w monografiach naukowych | 1 | 1 | 2 |
| Prace w wydawnictwach pokonferencyjnych | 5 | 28 | 33 |
| Publikacje popularnonaukowe | - | 2 | 2 |
| Redakcja wydawnictw naukowych i materiałów konferencyjnych | - | 1 | 1 |

E.2. WSKAŹNIKI BIBLIOMETRYCZNE

| | Punktacja wg roku wydania | Punktacja wg ostatniej listy |
|--|---------------------------|------------------------------|
| Sumaryczna liczba punktów MNiSW po uzyskaniu stopnia doktora | 345 | 396 |
| Sumaryczny IF po uzyskaniu stopnia doktora | 11,658 | 11,116 |
| <u>Liczba cytowań publikacji</u> | | |
| - wg Web of Science | | 17 |
| - wg Google Scholar | | 120 |
| <u>Indeks h</u> | | |
| - wg Web of Science | | 3 |
| - wg Google Scholar | | 6 |

E.3. OMÓWIENIE POZOSTAŁYCH OSIĄGNIĘĆ W PRACY NAUKOWEJ

Działalność naukową rozpocząłem w 2002 roku, podejmując pracę w Pracowni Ekonomiki Nasiennictwa i Hodowli Roślin (PENiHR), Zakładu Nasiennictwa i Nasionoznawstwa, Instytutu Hodowli i Aklimatyzacji Roślin w Radzikowie. Początkowo wdrażałem się w zagadnienia związane z szeroko pojętym postępowaniem biologicznym w roślinach uprawnych (publikacje – Zał.4, II.D.1., poz.: 1, 2, 4, 11, 12, 13, 14). Wraz z nowopoznanymi metodami analizy danych, wprowadzałem do tematyki realizowanej w Pracowni nowe zagadnienia statystyczne, jak np. zagadnienie interakcji genotypowo-środowiskowej oraz związane z nią zagadnienia stabilności odmian w sensie rolniczym (publikacje – Zał.4, II.D.1., poz.: 3, 6, 8, 9, 16). W tym okresie moje zainteresowanie wzbudziły dwa zagadnienia: analizy wielozmienne – jako całościowe podejście do badania złożonych zjawisk i procesów, opisywanych wieloma zmiennymi (publikacje – Zał.4, II.D.1., poz.: 12, 13) oraz zagadnienia, związane z analizą obrazów cyfrowych i możliwością wykorzystania tej techniki w badaniach rolniczych (publikacje – Zał.4, II.D.1., poz.: 7). Badania związane z powyższymi zagadnieniami realizowałem w ramach tematów działalności statutowej oraz programu wieloletniego w PENiHR (projekty badawcze – Zał.4, II.I., poz.: 1, 2, 4). Uczestniczyłem również jako wykonawca w projekcie zamawianym nr: PBZ-MNiSW-06/I/2007 (projekty badawcze – Zał.4, II.I., poz.: 3). Byłem także zaangażowany we współpracę z naukowcami ze Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie (Wydział Ogrodnictwa i Architektury Krajobrazu, Wydział Inżynierii Produkcji oraz Wydział Nauk o Zwierzętach), czego efektem były publikacje (Zał.4, II.D.1., poz.: 7, 9, 15) i rozdział w monografii (Zał.4, II.D.1., poz.: 5). Przed uzyskaniem stopnia doktora prowadziłem zajęcia dla studentów Wydziału Rolnictwa i Biologii oraz Wydziału Nauk o Żywności SGGW w Warszawie (ćwiczenia z przedmiotów informatycznych, matematycznych i statystycznych oraz metod badań rolniczych i doświadczalnictwa), co dało łącznie 909 godzin zajęć dydaktycznych (Zał.4, III.I.1.). W tym czasie przeprowadziłem również dwa szkolenia z zakresu statystyki matematycznej i doświadczalnictwa dla pracowników naukowych IHAR-PIB i instytucji współpracujących. W 2008 roku byłem członkiem Komitetu organizacyjnego XXXVIII Międzynarodowego Colloquium Biometrycznego organizowanego przez SGGW w Warszawie.

Po uzyskaniu stopnia doktora moje zainteresowania i praca naukowa koncentrowały się na siedmiu głównych działach:

1) Postęp w hodowli i uprawie odmian roślin uprawnych

Zadania związane z tą tematyką na początku skupiły się na przygotowaniu i opublikowaniu wyników mojej pracy doktorskiej, zgodnie z zaleceniami recenzentów. Efektem był cykl sześciu publikacji naukowych, prezentujących całość zagadnień poruszanych w rozprawie doktorskiej (publikacje – Zał.4, II.D.2., poz.: 2, 3, 4, 5, 6, 7) oraz jedna publikacja popularnonaukowa (Zał.4, III.Q.2., poz.: 1).

Tematyką związaną z postępowaniem w hodowli roślin rolniczych i jego wykorzystaniem w praktyce zajmowałem się również, uczestnicząc w realizacji kilku zadań w ramach działalności statutowej Instytutu (projekty badawcze – Zał.4, II.I., poz.: 1, 2, 5) oraz w ramach tematu, wchodzącego w skład programu wieloletniego (projekty badawcze – Zał.4, II.I., poz.: 4). W ramach tych prac, uczestniczyłem w organizacji badań ankietowych gospodarstw indywidualnych (opracowywanie instrukcji, przygotowywanie dokumentacji, gromadzenie i czyszczenie danych, itp.), które były i są nadal podstawą oceny efektów prac hodowlanych w praktyce rolniczej. Efektem tych prac była między innymi publikacja naukowa (publikacje – Zał.4, II.D.2., poz.: 19). Tą samą tematyką zajmowałem się również uczestnicząc w temacie zamawianym nr PBZ-MNiSW-06/I/2007 (projekty badawcze – Zał.4, II.I., poz.: 3).

2) Badanie złożonych zależności przyczynowo-skutkowych

Zagadnienie analizy złożonych zależności przyczynowo-skutkowych zainteresowało mnie w trakcie opracowywania wyników mojej rozprawy doktorskiej. W trakcie prowadzonych analiz doszedłem do etapu oceny wpływu poszczególnych składowych postępu biologicznego na ten postęp, mierzony bezpośrednim efektem ważnym dla rolnika – czyli wzrostem plonowania. W pracy doktorskiej do opisanie tej zależności wykorzystałem metodę analizy regresji wielorakiej. Zdawałem sobie jednak sprawę, że stosowana przeze mnie metoda zakłada pełną niezależność zmiennych przyczynowych, a w przypadku składowych postępu biologicznego nie miało to miejsca, zatem uzyskane oceny były obciążone. Niestety wszystkie powszechnie stosowane w badaniach rolniczych metody (jak np. klasyczna analiza ścieżek) opierają się albo bezpośrednio na modelu funkcji regresji wielorakiej, albo na estymacji metodą najmniejszych kwadratów, przy założeniu niezależności zmiennych przyczynowych. Zwróciłem zatem uwagę na rozwiązania, stosowane często w psychometryce, czyli na analizę modeli równań strukturalnych (SEM). Zgłębiając to zagadnienie stwierdziłem, że mimo wielu zalet, metoda ta nie znalazła jeszcze powszechnego zastosowania w badaniach rolniczych lub przyrodniczych. Dlatego też postanowiłem opracować monografię naukową, poświęconą tej metodzie. Monografia ta stanowi prezentację osiągnięcia naukowego w postępowaniu habilitacyjnym.

3) Analizy wielozmienne i możliwości ich wykorzystania w badaniach rolniczych

Innym interesującym mnie zagadnieniem, są metody wielozmiennej analizy danych. Tematyka ta jest częściowo powiązana z wyżej opisanymi zagadnieniami i jednocześnie stanowi kontynuację prac rozpoczętych przed uzyskaniem stopnia doktora. Moja praca w ramach tego zagadnienia skupiała się przede wszystkim na teoretycznych podstawach wielozmiennego podejścia do analizy danych doświadczalnych (publikacje – Zał.4, II.D.2., poz.: 13, 14, 16) oraz na możliwości praktycznego wykorzystania eksploracyjnej analizy czynnikowej (EFA) w badaniach rolniczych (publikacje – Zał.4, II.D.2., poz.: 8, 18). Wnikliwe poznanie i zrozumienie analiz EFA i SEM zainspirowało mnie do wprowadzenia i wykorzystania konfirmacyjnej analizy czynnikowej (CFA) w badaniach rolniczych. Zagadnienie to stanowi część prezentowanego osiągnięcia naukowego.

Ponadto moje zainteresowanie wzbudziło wykorzystanie systemów sztucznej inteligencji w analizie danych rolniczych, szczególnie w zakresie opracowywania modeli predykcyjnych (publikacje – Zał.4, II.D.2., poz.: 15).

Dwa referaty dotyczące tej tematyki wygłoszono również na warsztatach szkoleniowych (Zał.4, III.I.2., poz.: 8, 9).

4) Planowanie doświadczeń rolniczych i analiza danych pochodzących z tych doświadczeń

W zakresie tematyki realizowanej w PENiHR znajdowały się zagadnienia, związane bezpośrednio z podstawami doświadczalnictwa rolniczego, czyli z zasadami planowania eksperymentu i metodyką analiz danych doświadczalnych (projekty badawcze – Zał.4, II.I., poz.: 1, 3, 7). Realizowane przeze mnie prace wiązały się z metodyką planowania doświadczeń hodowlanych, nasiennych i uprawowych oraz z różnymi sposobami analizy danych, pochodzących z tych doświadczeń. W ramach tych zagadnień prowadziłem szereg szkoleń (Zał.4, III.I.2., poz.: 5, 12, 14, 15, 16) i wygłosiłem referat na ogólnopolskim seminarium naukowym organizowanym przez firmę StatSoft Polska Sp. z o. o. (Zał.4, III.I.2., poz.: 13). Wyniki prowadzonych prac zostały opublikowane w monografii naukowej, której jestem współautorem oraz jako rozdział w innej monografii (Zał.4., II.D.2., poz.: 10, 24).

5) Współpraca z zespołami naukowców prowadzącymi badania w zakresie nauk rolniczych

W obecnej chwili jestem jedynym *statystykiem* zatrudnionym w IHAR-PIB. Z tego też powodu mam możliwość nawiązywania współpracy z różnymi zespołami naukowymi, działającymi w Instytucie oraz z pojedynczymi naukowcami.

Współpracuję z zespołem naukowców z Zakładu Fitopatologii oraz Zakładu Genetyki i Hodowli Roślin IHAR-PIB, w ramach badań, związanych z odpornością zbóż na choroby grzybowe. Do moich zadań należy planowanie części eksperymentalnej oraz analiza statystyczna, uzyskanych danych. W ramach tej współpracy zbadano wirulencję izolatów *Mycosphaerella graminicola*, otrzymanych z roślin pszenicy, posiadających znane źródła odporności na *Septoria triticii* (Załącznik 4., II.A.2., poz.: 1). Główny problem w niniejszych badaniach stanowiło wskazanie izolatów o nietypowej, znacząco odstającej od oczekiwań, reakcji w kontakcie z roślinami testowymi. Występowanie takich nietypowych reakcji (interakcji) potwierdzono dwuczynnikową analizą wariancji, a następnie każdemu z poziomów czynników przypisano ich przeciętną wartość porażenia (w postaci mediany) i na tak uzyskanych danych przeprowadzono analizę regresji wielorakiej. Zakładając, że hiperpłaszczyzna odpowiadająca uzyskanej regresji reprezentuje typową reakcję na efekty główne, poszukiwano obserwacji odstających (z pomocą statystyki D-Cooka), które powinny odpowiadać za występowanie interakcji izolat \times roślina.

W innych badaniach zajmowano się izolatami *Blumeria graminis* porażającymi rośliny pszenżyta (Załącznik 4., II.A.2., poz.: 3). Dotychczas podejrzewano, że szczepy specyficzne dla pszenicy, porażają również pszenżyto. W wyniku przeprowadzonych doświadczeń stwierdzono, wbrew wcześniejszym przekonaniom, że niektóre szczepy są specyficzne w stosunku do pszenżyta. W badaniach wykorzystano analizę wariancji oraz analizę składowych głównych PCA.

Współpracując z zespołem badaczy z Zakładu Biochemii i Fizjologii Roślin IHAR-PIB, brałem udział w badaniach roślin o zwiększonej tolerancji na glin (Załącznik 4., II.A.2., poz.: 2, 6). Moim zadaniem było porównanie badanych genotypów. W tym celu wykorzystałem analizę wariancji, współczynniki odziedziczalności oraz testy normalności Shapiro-Wilka, analizę komponentów wariancyjnych metodą REML, test homogeniczności wariancji Bartlett'a oraz test t-Studenta dla dwóch populacji z modyfikacją Satterwhite'a. W innych badaniach analizowałem zmienność somaklonalną w kulturach *in vitro* pszenżyta (Załącznik 4., II.A.2., poz.: 4). W tym celu zastosowałem, przede wszystkim, analizę wariancji oraz analizę kontrastów metodą Tukeya.

Współpracuję również z zespołem badaczy z Samodzielnej Pracowni Oceny Jakości Produktów Roślinnych IHAR-PIB. W ramach tej współpracy badano odmiany i rody jęczmienia jarego pod względem zmienności genotypowej i środowiskowej, występującej w ramach cech jakościowych i browarnianych (Załącznik 4., II.D.2., poz.: 20, 23). W badaniach wykorzystano analizę wariancji, procedurę porównań wielokrotnych Tukeya oraz analizę współczynników korelacji liniowych Pearsona.

W ramach współpracy z oddziałami IHAR-PIB w Młochowie i Jadwisinie zajmowałem się zagadnieniami związanymi z hodowlą ziemniaka, a w szczególności wielocephową zmiennością mieszańców i rodów ziemniaka (Załącznik 4., II.D.2., poz.: 21, 22) oraz analizą współzależności wybranych cech użytkowych ziemniaka, uprawianego w różnych warunkach (Załącznik 4., II.D.2., poz.: 17). W badaniach tych wykorzystałem analizę wariancji, analizę interakcji genotypowo-środowiskowej w oparciu o model mieszany Scheffégo-Calińskiego i model regresji łącznej Calińskiego-Kaczmarka, analizę składowych głównych PCA oraz eksploracyjną analizę czynnikową EFA. Zajmowałem się również analizą zmian wybranych parametrów fizjologicznych, charakteryzujących przebieg fotosyntezy, w czasie rozwoju rośliny (Załącznik 4., II.A.2., poz.: 5). W tych badaniach, oprócz analizy wariancji, wykorzystałem analizę linearyzowanego modelu regresji nieliniowej.

Podjąłem również współpracę z zespołem naukowym z Zakładu Biotechnologii i Cytogenetyki Roślin IHAR-PIB. W ramach tej współpracy zajmowałem się analizą zmienności cech powiązanych z kształtowaniem się plonu ziarna z rośliny, pomiędzy liniami DH (podwojonych haploidów) pszenicy, jęczmienia i pszenżyta (Załącznik 4., II.D.2., poz.: 9, 12, 25). W badaniach tych wykorzystałem analizę wariancji i procedurę porównań wielokrotnych Tukeya, analizę współczynników korelacji, analizę funkcji regresji wielorakiej i klasyczną analizę ścieżek.

Ponadto efektem mojej współpracy z badaczami z i spoza IHAR-PIB było uczestnictwo w badaniach nad zdrowotnością ziarna zbóż, pochodzącego z upraw ekologicznych (Załącznik 4., II.D.2., poz.: 1, 27) oraz uwarunkowaniami ekonomiczno-społecznymi korzystania przez rolników z funduszy UE (Załącznik 4., II.D.2., poz.: 11). W latach 2012 i 2013 kierowałem również zespołem ekspertów z zakresu statystyki i modelowania matematycznego, w ramach realizacji projektu POIG nr: UDA-POIG.01.03.01-00-204/09-06.

6) Udział w projekcie LIDER pt. „Pieczywo pszenżytnio-owsiane jako nowy polski produkt żywieniowy o podwyższonej wartości prozdrowotnej”

Od stycznia 2014 roku jestem wykonawcą zadań realizowanych w projekcie LIDER nr: LIDER/019/519/L-4/12/NCBR/2013, którego kierownikiem jest dr inż. Anna Fraś z Samodzielnej Pracowni Oceny Jakości Produktów Roślinnych (SPOJPR). Celem projektu jest opracowanie receptury nowego rodzaju pieczywa, wytwarzanego z mąki pszenżytniej z dodatkiem komponentów owsianych. Do moich obowiązków należało kierowanie dwoma zadaniami. Pierwsze zadanie zostało zrealizowane w 2014 roku i dotyczyło przygotowania i przeprowadzenia ankiet wśród konsumentów pieczywa oraz zebranych analiza danych ankietowych. Celem badania ankietowego było rozpoznanie preferencji i oczekiwań konsumentów względem pieczywa. Zbieraniem danych zajmował się podmiot zewnętrzny (konsorcjum BIOSTAT z Rybnika). W wyniku realizacji tego zadania przygotowano i opublikowano artykułu popularnonaukowego w prasie branżowej (Załącznik 4., III.Q.2., poz.: 2). Drugie zadanie, którym kieruję obecnie, dotyczy statystycznego opracowania wyników, uzyskanych przez wykonawców pozostałych zadań, przewidzianych w projekcie. Odpowiadam za planowanie poszczególnych doświadczeń, dobór właściwych metod analizy danych oraz za przeprowadzenie analiz i wnioskowanie statystyczne na podstawie ich wyników. W ramach tego zadania dotychczas opublikowano dwa artykuły naukowe (publikacje – Załącznik 4., II.A.2., poz.: 7; II.D.2., poz.: 26).

7) Organizacja doświadczeń i analiza danych z serii zespołowych doświadczeń hodowlanych

Tematyką tą zająłem się od połowy 2014 roku, gdy Rada Ministrów RP przyjęła program wieloletni dla IHAR-PIB, na lata 2014-2020. W tym programie jestem wykonawcą zadania, związanego z opracowywaniem i doskonaleniem metodyki doświadczalnej w badaniu i selekcji nowych odmian zbóż, pod kątem ich cech użytkowych. Do moich obowiązków należy analiza i ocena metod, stosowanych dotychczas w planowaniu zespołowych doświadczeń hodowlanych oraz w analizie uzyskanych w nich danych. Doświadczenia te są organizowane corocznie przez polskie firmy hodowlane. W tych doświadczeniach hodowcy porównują wzajemnie swoje materiały hodowlane, a także dokonują ich ostatecznej selekcji do zgłoszeń rejestrowych do COBORU. Po przeanalizowaniu stosowanych metod, moim zadaniem jest poprawienie systemu planowania doświadczeń oraz usprawnienie i poprawienie jakości stosowanych narzędzi analitycznych. Do tej pory opracowałem narzędzie (w postaci makroprogramu w Statistica Visual Basic) do przygotowywania schematów i planowania doświadczeń oraz generowania plików do wprowadzania danych, podlegających dalszym analizom. Narzędzie to zostanie udostępnione na stronie internetowej programu wieloletniego, pod koniec roku 2016.

Pozostała działalność naukowo-dydaktyczna i organizacyjna

W ramach pozostałej działalności naukowej prowadziłem zajęcia (wykłady i ćwiczenia) dla studentów Wydziału Rolnictwa i Biologii oraz Wydziału Nauk o Żywności SGGW w Warszawie (Załącznik 4., III.I.2., poz.: 1, 2, 3, 4). W ramach współpracy z firmą StatSoft Sp. z o. o. (producent programu Statistica) przeprowadziłem trzy szkolenia z zakresu statystyki matematycznej i doświadczalnictwa, skierowane do pracowników nauki, zajmujących się badaniami rolniczymi (Załącznik 4. III.I.2., poz.: 12, 14, 15, 16). Byłem również członkiem komitetu organizacyjnego 44 Międzynarodowego Colloquium Biometrycznego i IV Polsko-Portugalskich Warsztatów z Biometrii, współorganizowanych przez IHAR-PIB w Krakowie. Ponadto byłem sekretarzem komitetu organizacyjnego Warsztatów Biometrycznych, organizowanych w 2010 roku w IHAR-PIB w Radzikowie. Od 2015 roku jestem członkiem komitetu redakcyjnego czasopisma *Plant Breeding and Seed Science* (ISSN: 0018-

3040), w którym pełnię funkcje redaktora statystycznego. W tym samym roku podjąłem współpracę z firmą StatSoft Polska Sp. z o. o. w roli konsultanta z zakresu wykorzystania metod statystycznych w badaniach rolniczych.

14 lipca 2015 roku zostałem wyznaczony przez Radę Naukową IHAR-PIB w Radzikowie na promotora pomocniczego w przewodzie doktorskim mgr inż. Joanny Jankowskiej dla uzyskania stopnia doktora nauk rolniczych, dziedzina Nauki Rolnicze, dyscyplina Agronomia.

A handwritten signature in blue ink, reading "Dariusz Mańkowski". The signature is written in a cursive style with a large initial 'D'.