

## NOWOŚCI W STATISTICA WSPIERAJĄCE PLANOWANIE BADAŃ I ANALIZĘ ICH WYNIKÓW W NAUKACH PRZYRODNICZYCH

*Dariusz R. Mańkowski, Instytut Hodowli i Aklimatyzacji Roślin — PIB, Radzików;  
Paweł Januszewski, StatSoft Polska Sp. z o.o.*

Każda dziedzina nauki charakteryzuje się specyficznym warsztatem narzędzi wykorzystywanych do prowadzenia badań. Różne są podejścia do planowania badań, różne do prowadzenia eksperymentów, w końcu różne są podejścia do interpretacji wyników, wnioskowania i tworzenia wiedzy naukowej. Aby stosowane przez badaczy narzędzia badawcze sprawdzały się w danej dziedzinie muszą być one do niej właściwie dostosowane. W programie Statistica, dzięki zespołowi analityków StatSoft Polska, sukcesywnie implementowane są dodatkowe rozwiązania dla badaczy z poszczególnych dziedzin nauki: dla nauk ekonomicznych, medycznych, dla pracowników laboratoriów. Teraz pojawiają się rozwiązania dedykowane naukowcom związanym z naukami przyrodniczymi.

Przedmiotem badań w naukach przyrodniczych jest otaczający nas świat. Prowadzone doświadczenia są związane z obserwacją zjawisk i procesów zachodzących wokół nas. Dlatego też stosuje się tu czasem wręcz unikalne układy doświadczalne, niespotykane w innych badaniach. Przykładem takich rozwiązań mogą być doświadczenia związane z naukami rolniczymi. Doświadczalnictwo rolnicze ma szeroki wachlarz mniej oraz bardziej złożonych układów doświadczalnych, takich jak na przykład: układy bloków losowych, układy hierarchiczne split-plot i split-blok, układy o charakterze hierarchiczno-krzyżowym, a także całą gamę układów doświadczalnych związanych z hodowlą roślin: układy dialleliczne,

top-cross czy też układy kratowe. Złożone jest tu zarówno planowanie doświadczeń, jak i analiza ich wyników. W naukach przyrodniczych szczególną rolę odgrywa również badanie zależności o charakterze przyczynowo-skutkowym, czyli takich, w których określa się, jakie zmienne są przyczyną kształtowania się danego zjawiska lub procesu, a jakie jego skutkiem (Mańkowski, 2013). Do analizy tych zależności wykorzystuje się narzędzia związane z modelowaniem statystycznym za pomocą prostych i złożonych modeli regresyjnych oraz innych bardziej skomplikowanych narzędzi. Stosowane tu rozwiązania nie zawsze są szablonowe i czasem wymagają specjalnego podejścia. Zupełnie nowym wyzwaniem dla istniejących od lat pakietów do analizy danych są zagadnienia związane z tzw. bioinformatyką i wykorzystaniem technik obliczeniowych i statystyki matematycznej w biologii molekularnej. Tu często mamy do czynienia z zupełnie innym rodzajem danych poddawanych analizie, co zmusza do wykorzystywania innych podejść do analizy, gdyż standardowe metody statystycznej analizy danych nie mają tu zastosowania.

## Planowanie doświadczeń

Pozyskanie danych empirycznych wymaga przeprowadzenia eksperymentu. Zadaniem pracy naukowo-badawczej jest uzyskanie z badanego materiału możliwie dużego zasobu informacji przy jak najmniejszym nakładzie sił i środków, i to w jak najkrótszym czasie (Trajer i in., 2016). Sprawia to, że właściwe zaplanowanie eksperymentu jest kluczowym elementem prowadzonych badań. W naukach przyrodniczych stosuje się bardzo wiele różnych układów doświadczalnych, których dobór uzależniony jest od liczby badanych czynników, charakteru obserwowanych zmiennych oraz warunków, w jakich prowadzone będzie doświadczenie. Najlepiej jest to widoczne przy planowaniu doświadczeń polowych z roślinami. Gama dostępnych układów doświadczalnych jest bardzo duża. Od układów najprostszych (doświadczenia jednopowtórzeniowe, bloki losowe), poprzez układy hierarchiczne (split-plot, split-blok), aż do bardzo złożonych i wyspecjalizowanych układów doświadczalnych (split-split-plot, czy też układy wykorzystywane w hodowli roślin: dialleliczne, top-cross, kratowe).

Planując jakiegokolwiek doświadczenie, dążymy do uzyskania danych niezbędnych do charakterystyki i opisu badanego zjawiska lub procesu, bądź też rozstrzygnięcia pewnych nowych hipotez lub potwierdzenia hipotez stawianych przez innych badaczy, w odmiennych, specyficznych warunkach (Trajer i in., 2016). W ścisłych doświadczeniach naukowych zwykle bada się wpływ konkretnych czynników kontrolowanych na występowanie lub natężenie zjawisk będących przedmiotem badań, z możliwym włączeniem czynników niebędących przedmiotem zainteresowania, ale mogących zakłócić proces i poprawność zbieranych informacji (Wójcik i Laudański, 1989).

Jednak nawet po wybraniu określonego układu, w którym zamierzamy założyć doświadczenie, pozostaje problem właściwego rozlosowania obiektów doświadczalnych i ich kombinacji wg tego układu. Naprzeciw tym problemom wychodzą rozwiązania do planowania doświadczeń zawarte w *Zestawie Przyrodnika* (StatSoft, 2017).

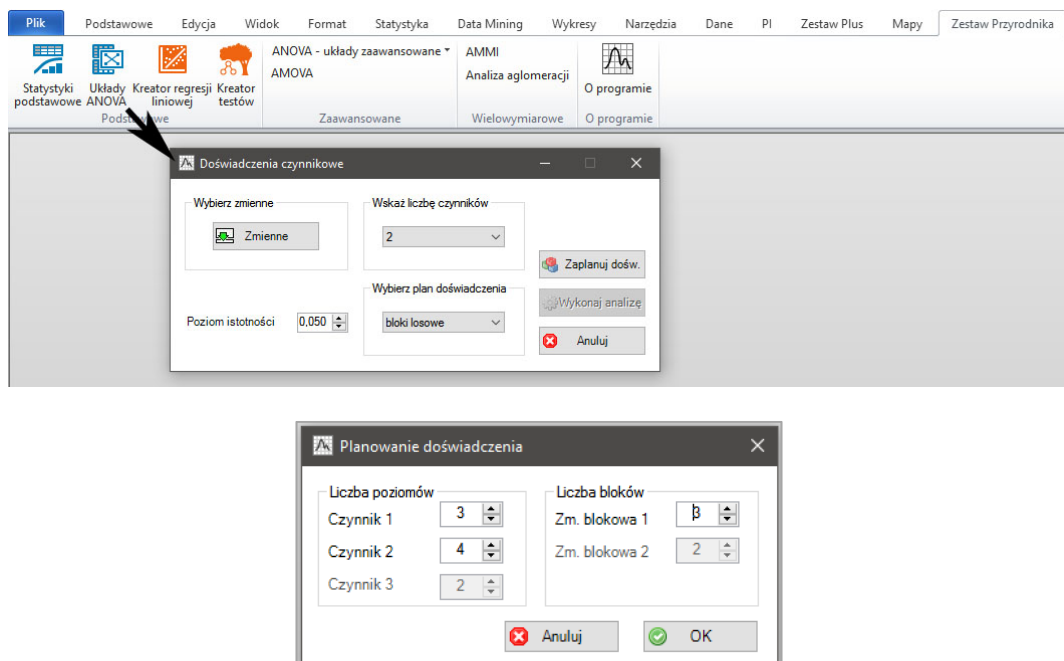
### ***Przykład – planowanie dwuczynnikowego doświadczenia w układzie bloków losowych***

Układ bloków losowych (*completely randomized blocks design, CRB design*) jest jednym z najprostszych i najbardziej podstawowych układów doświadczalnych. Wykorzystuje się go, gdy na obszarze doświadczalnym (np. na polu doświadczalnym) występuje ukierunkowana zmienność systematyczna wpływająca na wyniki prowadzonego doświadczenia. Wówczas obszar doświadczenia dzieli się na prostokątne bloki, które położone są prostopadłe dłuższym bokiem do kierunku zmienności systematycznej. Następnie w każdym bloku oddzielnie rozlosowuje się wszystkie warianty (kombinacje) badanych czynników.

Zmienność systematyczna w doświadczeniu założonym w układzie bloków losowych odpowiada zmienności pomiędzy blokami uwzględnionymi w planie doświadczenia. Zakłada się, że zmienność ta wewnątrz pojedynczego bloku jest nieznacząca, a jedyna obserwowana zmienność jest wynikiem działania badanych czynników.

W celu rozlosowania doświadczenia w układzie bloków losowych z narzędzi dostępnych w *Zestawie Przyrodnika* wybieramy **Układy ANOVA**, w oknie dialogowym wybieramy

liczbę czynników (2) oraz układ doświadczenia (**bloki losowe**), następnie wybieramy przycisk **Zaplanuj dośw.** Następnie wskazujemy liczbę poziomów dla planowanych czynników doświadczenia (np. 3 oraz 4 odpowiednio dla pierwszego oraz drugiego czynnika) oraz liczbę planowanych bloków (np. 3).



Rys. 1. Planowanie doświadczeń czynnikowych w *Zestawie Przyrodnika*

Wynikiem rozplanowania doświadczenia jest plan doświadczenia pokazujący w sposób schematyczny, jak wyglądać będzie doświadczenie, oraz w jaki sposób rozlosowane są poszczególne poziomy czynnika lub kombinacje poziomów czynników badanych w doświadczeniu. Plan wskazuje rozmieszczenie przestrzenne wszystkich jednostek doświadczalnych.

Kolejnym elementem uzyskanych wyników jest arkusz danych przygotowany do wprowadzania wyników zaplanowanego doświadczenia. Dzięki wygenerowanemu arkuszowi badacz nie będzie miał problemów z odpowiednim przygotowaniem arkusza danych.

A

Analiza dwuczynnikowa, bloki losowe # bloków: 3 # poziomów czynnika I: 3 # poziomów czynnika II: 4			
	1	2	3
	Blok 1	Blok 2	Blok 3
1	A2B2	A2B4	A2B4
2	A3B3	A3B3	A1B2
3	A3B4	A1B4	A3B3
4	A1B1	A1B3	A3B1
5	A2B4	A1B2	A1B4
6	A3B2	A2B2	A1B1
7	A1B4	A2B1	A2B3
8	A2B1	A3B4	A2B1
9	A3B1	A2B3	A2B2
10	A1B2	A3B2	A3B2
11	A1B3	A3B1	A1B3
12	A2B3	A1B1	A3B4

B

	1	2	3	4
	Blok	Czynnik I	Czynnik II	Zmienna zależna
1	1	1	1	
2	1	1	2	
3	1	1	3	
4	1	1	4	
5	1	2	1	
6	1	2	2	
7	1	2	3	
8	1	2	4	
9	1	3	1	
10	1	3	2	
11	1	3	3	
12	1	3	4	
13	2	1	1	
14	2	1	2	
15	2	1	3	
16	2	1	4	
17	2	2	1	
18	2	2	2	
19	2	2	3	
20	2	2	4	

Rys. 2. Plan doświadczenia dwuczynnikowego (A) w układzie bloków losowych oraz arkusz danych (B) przygotowany do wprowadzania wyników doświadczenia

## Analiza danych z doświadczeń czynnikowych

Wyniki przeprowadzonego doświadczenia staramy się zapisać w postaci liczb – zmiennych ilościowych charakteryzujących wybrane cechy badanych obiektów. Te obserwacje zmiennych ilościowych stanowią dane, które poddajemy następnie analizie. Podstawowym testem statystycznym wykorzystywanym do analizy wyników pochodzących z doświadczeń czynnikowych jest test F analizy wariancji (*ANOVA*). Metoda ta została opracowana przez R. A. Fishera, uznawanego za prekursora statystyki matematycznej i doświadczalnictwa przyrodniczego (Mądry, 2003). Z matematycznego punktu widzenia analiza wariancji wymaga uwzględnienia wszystkich elementów statystycznej charakterystyki doświadczenia oraz wykorzystanego układu doświadczalnego. Wyrażone jest to w postaci równania zwanego modelem liniowym analizy wariancji. Model ten traktuje każdą obserwację w doświadczeniu czynnikowym jako realizację zmiennej losowej o rozkładzie normalnym

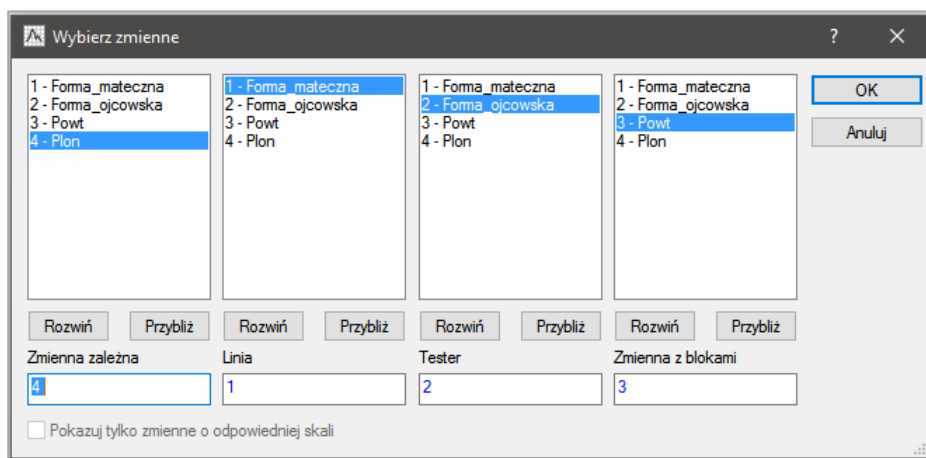
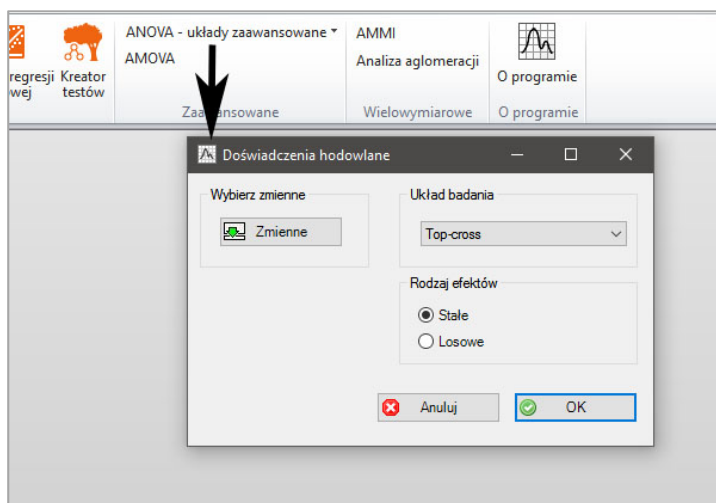
z jednakową wariancją  $\sigma_e^2$  i wartością oczekiwaną zależną od poziomów badanych czynników, bloków i efektów interakcyjnych (Mądry, 2003). Analiza wariancji polega na matematycznym rozdzieleniu zmienności zmiennej zależnej na zmienności odpowiadające wyodrębnionym źródłom zmienności pokrywającym się z efektami uwzględnionymi w modelu liniowym analizy. Każdemu układowi doświadczalnemu odpowiada stosowny model liniowy, a co za tym idzie – układ wyodrębnianych źródeł zmienności (Cochran i Cox, 1992; Montgomery 2005).

Bardziej skomplikowane układy doświadczalne wymagają od badacza właściwego ich zdefiniowania i właściwego przeprowadzenia analizy. Im bardziej złożony układ doświadczenia, tym proces przeprowadzenia analizy jest trudniejszy. Dlatego też w *Zestawie Przyrodnika* przygotowano gotowe modele do analizy najczęściej spotykanych układów doświadczalnych w badaniach przyrodniczych.

***Przykład – analiza danych z doświadczenia hodowlanego z krzyżowań linia  $\times$  tester z formami rodzicielskimi traktowanymi jako czynniki stałe w układzie top-cross***

W ramach prowadzonych prac hodowlanych wykonano 25 kombinacji krzyżowań pomiędzy 10 odmianami pszenicy ozimej w układzie czynnikowym top-cross (linia  $\times$  tester – *North Carolina II*), gdzie pięć odmian było formami matecznymi, a pozostałe pięć formami ojcowskimi (Ubysz-Borucka i in., 1985; Mądry i in., 2010). Uzyskane potomstwo z pokolenia F1 badano w 3 powtórzeniach (poletkach), z każdego poletka wybrano po 5 roślin do badań szczegółowych. Analizowaną zmienną ilościową stanowił plon ziarna w gramach z jednej rośliny, a model analizy był stały.

Analizę wyników tego doświadczenia przeprowadzono z wykorzystaniem narzędzi dostępnych w *Zestawie Przyrodnika*.



Rys. 3. Okno analizy dla zaawansowanych układów ANOVA – doświadczenia hodowlane oraz okno wyboru zmiennych dla układu linia  $\times$  tester (top-cross)

W analizie wyników doświadczeń hodowlanych z roślinami, w których badanymi obiektami są krzyżówki dwóch form rodzicielskich, efekty główne (efekty genetyczne form matecznych i ojcowskich) noszą nazwę efektów GCA, czyli efektów ogólnej zdolności kombinacyjnej, natomiast efekty interakcyjne form rodzicielskich noszą nazwę efektów SCA, czyli efektów specyficznej zdolności kombinacyjnej.

W wyniku przeprowadzonej analizy uzyskana tabela analizy wariancji była postaci:

Źródło zmienności	Doświadczenie linia x tester Efekty stałe				
	SS	Stopnie swobody	MS	F	p
Linia	30,028	4	7,507	7,713	0,000006
Tester	61,663	4	15,416	15,838	0,000000
Linia * Tester	40,521	16	2,533	2,602	0,000752
Blok	5,014	2	2,507		
Błąd losowy	338,724	348	0,973		

Rys. 4. Tabela analizy wariancji dla doświadczenia w układzie linia × tester (top-cross)

Jak widać, wszystkie efekty badane w doświadczeniu były istotne statystycznie. Wyznaczone wartości efektów ogólnej zdolności kombinacyjnej GCA linii i testerów były postaci:

	Wartości GCA			
	1 Linia	2 GCA	3 Tester	4 GCA
1	Bezostaja	0,0082666667	1/59 St.	-0,4944
2	C-48	-0,317066667	Capelle DespreZ	-0,0410666667
3	R-1099	-0,307733333	Miranowska	0,6616
4	R-1298	0,394933333	Perdrix	-0,3184
5	R-4013	0,2216	Vermilian	0,192266667

Rys. 5. Tabela efektów GCA dla doświadczenia w układzie linia × tester (top-cross)

Natomiast wartości efektów specyficznej zdolności kombinacyjnej SCA kombinacji badanych w doświadczeniu linii i testerów były postaci:

Linia/Tester	Wartości SCA				
	1 1/59 St.	2 Capelle DespreZ	3 Miranowska	4 Perdrix	5 Vermilian
Bezostaja	-0,0722667	-0,3256	-0,494933333	0,5184	0,3744
C-48	-0,1602667	0,453066667	-0,136266667	0,0104	-0,1669333
R-1099	-0,4562667	-0,076266667	0,821066667	-0,3989333	0,1104
R-1298	0,40773333	0,2544	-0,314933333	-0,1682667	-0,1789333
R-4013	0,28106667	-0,3056	0,125066667	0,0384	-0,1389333

Rys. 6. Tabela efektów SCA dla doświadczenia w układzie linia × tester (top-cross)

Ostatnim elementem doświadczeń hodowlanych jest wyznaczenie współczynników odziedziczalności w szerokim lub wąskim sensie. **Współczynnik odziedziczalności w szerokim**



**sensie** jest miarą korelacji (bliskości) między wartościami fenotypowymi danej zmiennej (w postaci średniej z poletek) a wartościami genotypowymi badanych jednostek selekcyjnych (w postaci średnich dla genotypów określonych w losowych warunkach środowiska, charakteryzowanych przez błąd doświadczalny), natomiast **współczynnik odziedziczalności w wąskim sensie** jest miarą korelacji (bliskości) między wartościami fenotypowymi danej zmiennej (w postaci średniej z poletek), a genetycznymi efektami addytywnymi (średnimi efektami dla potomstwa) (Mądry i in., 2010).

Uzyskane wartości współczynników odziedziczalności w wąskim i szerokim sensie były postaci:

	Współczynniki odziedziczalności	
	1 W szerokim sensie [%]	2 W wąskim sensie [%]
Krzyżówka	84,05	58,51
Jednostka	26,00	18,10

Rys. 7. Tabela współczynników odziedziczalności dla doświadczenia w układzie linia × tester (top-cross)

## Analiza serii doświadczeń

Analiza statystyczna wyników doświadczenia pojedynczego umożliwia jedynie uzyskanie informacji o zachowaniu się badanych czynników (wpływie tych czynników na analizowaną zmienną zależną) w konkretnych warunkach doświadczenia. Dla doświadczeń przyrodniczych realizowanych w konkretnym miejscu i czasie są to konkretne warunki związane z położeniem doświadczenia i warunkami pogodowymi panującymi w trakcie trwania doświadczenia.

W ramach doświadczeń z obiektami roślinnymi (konkretnymi genotypami roślin) wpływ warunków prowadzenia doświadczenia jest bardzo istotny ze względu na tak zwaną interakcję genotypu ze środowiskiem (warunkami prowadzonego doświadczenia). Interakcja ta polega na tym, że reakcja fenotypowa roślin na zmiany warunków środowiska nie jest

jednakowa dla wszystkich badanych genotypów. W rezultacie różnice obserwowane pomiędzy porównywanymi genotypami zależą od środowiska, w którym prowadzone jest porównanie (Mądry i in., 2010). To wszystko sprawia, że kompleksowa analiza zróżnicowania pomiędzy porównywanymi genotypami wymaga przeprowadzenia nie jednego, a całej serii (trzy lub więcej) identycznych doświadczeń założonych w różnych warunkach środowiska.

Sam sposób analizy serii doświadczeń będzie uzależniony od charakteru czynników i pytań, jakie postawimy na etapie planowania eksperymentu. W przypadku gdy czynniki traktujemy jako efekty stałe, możemy sięgnąć po modele analizy, takie jak split-split-plot lub po model hierarchiczno-krzyżowy. Jeżeli natomiast jeden z badanych czynników (najczęściej czynnik środowiskowy) lub wszystkie analizowane w doświadczeniu czynniki mają charakter losowy, to analizę opieramy odpowiednio na modelu mieszanym lub losowym.

Metody statystyczne wykorzystywane w analizie interakcji genotypowo-środowiskowej ( $G \times E$ ) są bardzo liczne i różnorodne w treści matematycznej i metodologicznej. Najczęściej wykorzystywane są mieszane modele analizy wariancji wraz z analizą regresji łącznej, metody oparte na analizie biplotów GGE oraz analizy oparte na modelu addytywno-multiplikatywnym – AMMI.

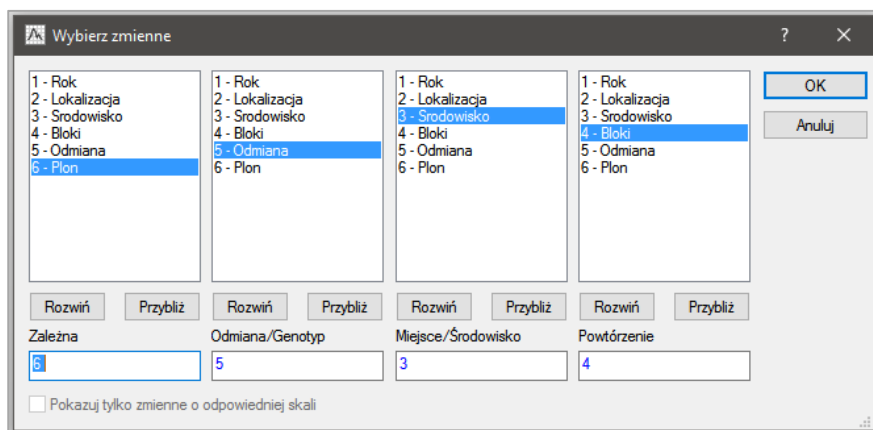
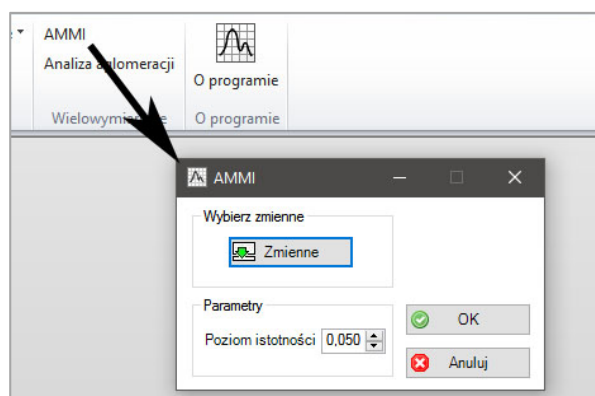
Model AMMI zakłada addytywność efektów głównych genotypów i środowisk oraz efektu interakcji  $G \times E$ , a efekty interakcji  $G \times E$  są przedstawione w tym modelu w postaci multiplikatywnych parametrów dla genotypów i środowisk. Takie podejście pozwala opisać wiernie i w sposób dość prosty każdą reakcję genotypów na zmienne warunki środowiska (Paderewski, 2008; Paderewski i in., 2011).

Kolejnym elementem blisko powiązanim z analizą interakcji  $G \times E$  jest ocena adaptacji genotypów w szerokim sensie. Genotypami o szerokiej adaptacji są nazywane te genotypy, które dają efekty wyższe niż średnia środowiskowa w znaczącej większości badanych środowisk (Paderewski, 2008). Do oceny stopnia szerokiej adaptacji genotypów wykorzystuje się ilościowe miary statystyczne stanowiące funkcje matematyczne dwóch powszechnie

uznanych determinantów stopnia szerokiej adaptacji, to jest średniej genotypowej oraz wariacji stabilności Shukli (Mądry i Iwańska, 2011).

### ***Przykład – analiza interakcji genotypowo-środowiskowej z wykorzystaniem modelu AMMI***

W ramach organizowanych przez Centralny Ośrodek Badania Odmian Roślin Uprawnych (COBORU) doświadczeń Porejstrowego Doświadczalnictwa Odmianowego (PDO) co roku w pewnej liczbie lokalizacji prowadzi się doświadczenia porównawcze z odmianami roślin dopuszczonymi do uprawy na terenie Polski. W ramach tych doświadczeń w latach 2011-2014 prowadzono w 4 lokalizacjach (Białogard, Chrzastowo, Krzyżewo, Jadwisin) doświadczenia porównawcze z 18 odmianami ziemiaka.



Rys. 8. Okno analizy dla modelu AMMI oraz okno wyboru zmiennych do analizy

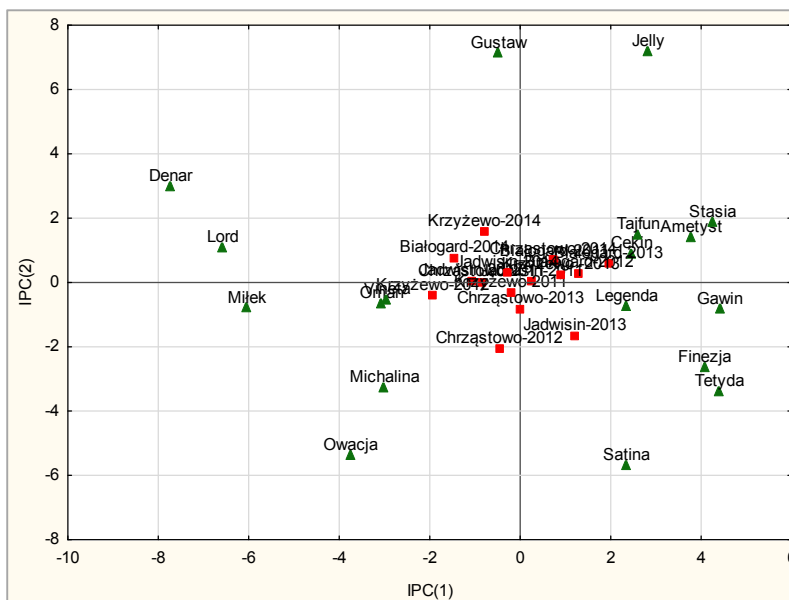
Przeprowadzono analizę interakcji  $G \times E$ , gdzie jako środowiska potraktowano kombinacje lata  $\times$  lokalizacje. Analizę wykonano z wykorzystaniem modelu AMMI w *Zestawie Przyrodnika* (zob. rys. 8).

Po przeprowadzonej analizie otrzymano tabelę analizy wariancji postaci:

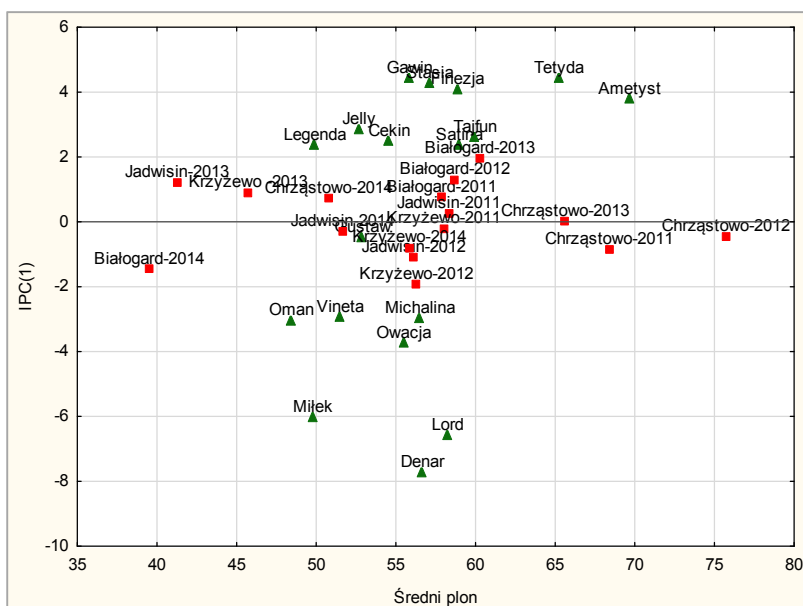
Źródło zmienności	Liczba stopni swobody	Suma kwadratów	Średni kwadrat	Odsetek zmienności	Względem interakcji	Statystyka F	Wartość p
Ogół	863	136482	158	100,0			
Miejsce	15	70958	4731	52,0		85,89	0,00
Powtórzenie(Miejsce)	32	1762	55	1,3		6,87	0,00
Odmiana	17	23374	1375	17,1		171,49	0,00
Odmiana x Miejsce	255	36026	141	26,4		17,62	0,00
Składowa nr 1	31	15933	514	11,7	44,2	17,62	0,00
Składowa nr 2	29	8053	278	5,9	22,4	11,19	0,00
Składowa nr 3	27	2854	106	2,1	7,9	7,70	0,00
Składowa nr 4	25	2176	87	1,6	6,0	6,82	0,00
Składowa nr 5	23	1888	82	1,4	5,2	6,12	0,00
Składowa nr 6	21	1313	63	1,0	3,6	5,32	0,00
Składowa nr 7	19	1182	62	0,9	3,3	4,80	0,00
Składowa nr 8	17	1027	60	0,8	2,8	4,10	0,00
Składowa nr 9	15	654	44	0,5	1,8	3,17	0,00
Składowa nr 10	13	487	37	0,4	1,4	2,46	0,00
Składowa nr 11	11	210	19	0,2	0,6	1,64	0,01
Składowa nr 12	9	151	17	0,1	0,4	1,31	0,15
Składowa nr 13	7	71	10	0,1	0,2	0,83	0,65
Składowa nr 14	5	23	5	0,0	0,1	0,45	0,89
Składowa nr 15	3	6	2	0,0	0,0	0,26	0,86
Błąd losowy	544	4361	8	3,2			

Rys. 9. Tabela analizy wariancji dla analizowanego modelu AMMI

Efekty główne dla odmian oraz środowisk (pozycja „Miejsce”) były istotne statystycznie (podobnie bloki w środowiskach). Udział odmiany w wytłumaczeniu obserwowanej zmienności wynosił 17,1%, a środowiska – 52,0%. 26,4 % zmienności przypadało na istotną interakcję  $G \times E$ . Wynika z tego, że największy wpływ na uwarunkowanie zmienności plonów ziemniaka miały środowiska, mniejszy interakcje  $G \times E$ , efekt odmiany był najmniej znaczący. Suma kwadratów odchyleń dla interakcji  $G \times E$  została rozłożona na sumy kwadratów odchyleń dla składowych interakcyjnych. Pierwszych jedenaście składowych interakcyjnych było istotne statystycznie. Z racji, że pierwsza składowa interakcyjna odpowiadała za 44,2%, a druga za 22,4% sumy kwadratów dla interakcji  $G \times E$ , dalszą interpretację ograniczono do tych dwóch składowych interakcyjnych. Należy mieć na uwadze, że efekty główne są oceniane z dużo większą precyzją niż efekty interakcyjne.



Rys. 10. Wykres IPC(1)/IPC(2)



Rys. 11. Wykres IPC(1)/Plon

Wykres PCI(1)/PCI(2) (zob. rys. 10) interpretujemy jak klasyczny biplot GGE. Badane genotypy możemy podzielić na kilka grup (tu 3), a na podstawie długości rzutów prostopadłych na proste łączące środek układu współrzędnych z punktami dla środowisk możemy ocenić adaptację poszczególnych odmian do warunków panujących w środowiskach biorących udział w doświadczeniu.

Wykres PCI(1)/Plon (zob. rys. 11) przedstawia oceny odmian i lokalizacji względem średnich (genotypowych i środowiskowych) oraz równocześnie względem pierwszej składowej interakcyjnej. Obiekty (odmiany i środowiska) leżące po prawej stronie wykresu cechowały się przeciętnie wyższymi plonami od średniej, a obiekty leżące po lewej stronie wykresu cechowały się przeciętnymi plonami poniżej wartości średniej. Z kolei położenie względem osi składowej interakcyjnej świadczy o udziale poszczególnych obiektów (odmian i środowisk) w kształtowaniu się efektu interakcyjnego. Obiekty położone blisko osi (o małych wartościach składowej interakcyjnej) cechowały się małym udziałem (można je uznać za stabilne).

Odmiana	Miara nadrzędności plonowania	Miara niezawodności przewagi plonowania	Miara Kanga stabilności
<b>Ametyst</b>	<b>11</b>	1,00	21
Cekin	171	0,38	6
Denar	199	0,56	4
Finezja	117	0,69	15
Gawin	148	0,56	8
Gustaw	232	0,44	5
Jelly	232	0,38	2
Legenda	254	0,06	1
Lord	145	0,56	14
Michalina	154	0,56	11
Milek	294	0,19	0
Oman	304	0,00	-1
Owacja	175	0,38	7
Satina	111	0,69	16
Stasia	135	0,63	13
Tajfun	92	0,75	17
Tetyda	35	0,94	19
Vineta	244	0,13	3

Rys. 12. Wyznaczone miary szerokiej adaptacji badanych odmian

Ostatnim elementem przeprowadzonej analizy była ocena badanych genotypów (odmian) pod względem miar szerokiej adaptacji. Wyznaczone zostały trzy takie miary. Wartość

**miary nadrzędności plonowania**  $i$ -tej odmiany jest tym mniejsza (bliższa zeru), im adaptacja danej odmiany w szerokim sensie jest wyższa. Im wartość **miary niezawodności przewagi plonowania** jest bliższa 1, tym dana odmiana ma wyższy stopień adaptacji w szerokim sensie, a odmiana z wartością równą 1 zawsze plonuje powyżej średnich środowiskowych. Ostatnią wyznaczoną miarą jest **miara Kanga stabilności**  $YS_i$ . Im wartość tej miary jest wyższa dla danej odmiany w zbiorze badanych odmian, tym ta odmiana cechuje się relatywnie (w stosunku do pozostałych odmian) wyższym stopniem szerokiej adaptacji.

## Badania molekularne

Do lat 70. XX wieku ocena zróżnicowania pomiędzy obiektami opierała się w głównej mierze na markerach morfologicznych, fizjologicznych i cytologicznych, izoenzymach oraz na analizach porównawczych z doświadczeń ścisłych, ocenie heterozji i analizie zmienności w krzyżowaniach (Rief i in., 2005). Obecnie najpowszechniej stosowane i uznane za najbardziej dokładne są metody oparte na markerach molekularnych. Wykorzystuje się je w badaniu genetycznych relacji pomiędzy różnymi genotypami w bankach genów oraz w procesach hodowlanych.

Dane pochodzące z analiz molekularnych są najczęściej matematycznym zapisem obrazów żeli elektroforetycznych, z najczęściej binarnym (zero-jedynkowym) zapisem pasm uzyskiwanych w wykonywanych analizach molekularnych. Fakt wystąpienia prążka o danej masie na elektroforegramie jest zwyczajowo zapisywany jako 1, brak prążka o tej masie jako 0. W ten sposób matematyczny obraz wszystkich pasm dla analizowanej puli genotypów ma postać zero-jedynkowej macierzy. Dane te mają charakter nominalny (skategoryzowany), to znaczy, że rzeczywista różnica pomiędzy wystąpieniem i brakiem danego prążka nie może być zapisana matematycznie jako różnica (1–0). Z tego powodu do analizy takich danych nie powinno się wykorzystywać klasycznych metod analizy danych ilościowych, takich jak analiza wariancji czy analiza funkcji regresji (Mańkowski i in., 2011).

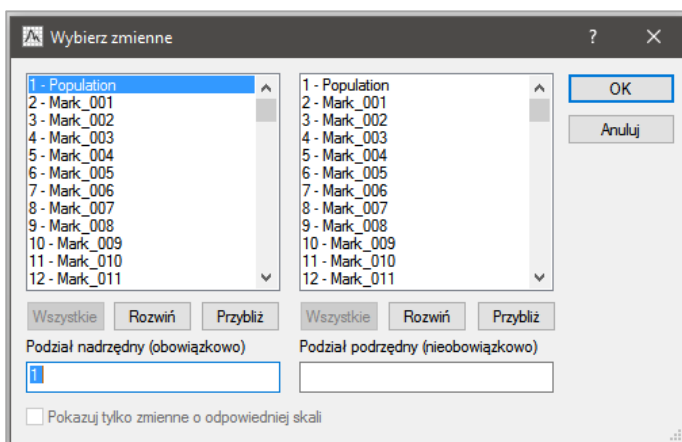
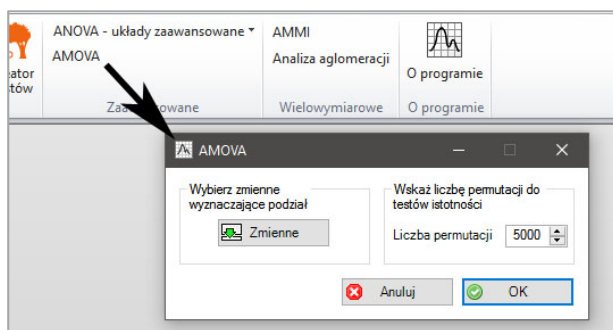
Najczęściej wykonywanymi analizami dla danych pochodzących z badań molekularnych są analizy oparte na podejściach taksonomicznych – analizy struktury badanych populacji (np. ocena miar podobieństwa genetycznego, analiza skupień, PCoA, analizy filogenetyczne, itp.), analizy zmienności w obrębie badanych populacji i subpopulacji (np. analiza wariancji molekularnej AMOVA) oraz analizy związane z mapowaniem genomu (np. mapowanie asocjacyjne, QTL). Narzędzia te wymagają często dość złożonego aparatu numerycznego. Dlatego bardzo ważny jest tu rozwój odpowiedniego oprogramowania wspierającego badaczy w ich pracy.

### ***Przykład – analiza wariancji molekularnej (AMOVA) dla populacji wybranych odmian pszenicy ozimej***

W ramach prowadzonych badań grupę 46 genotypów pszenicy ozimej poddano analizie molekularnej z wykorzystaniem 100 markerów genetycznych. W wyniku przeprowadzonych analiz uzyskano macierz 0-1 oddającą uzyskany obraz elektroforetyczny. Dodatkowo tę grupę genotypów podzielono z wykorzystaniem technik filogenetycznych na 6 populacji. Dla takich danych przeprowadzono analizę wariancji molekularnej w celu potwierdzenia występowania różnic pomiędzy wyodrębnionymi populacjami. Do przeprowadzenia analizy wykorzystano *Zestaw Przyrodnika*.

Analiza wariancji molekularnej służy do identyfikacji różnic pomiędzy populacjami i subpopulacjami z wykorzystaniem markerów molekularnych (Excoffier i in., 1992). AMOVA dokonuje rozdziału ogólnej zmienności pomiędzy badanymi genotypami na zmienność obserwowaną pomiędzy populacjami, pomiędzy subpopulacjami w obrębie populacji oraz wewnątrz populacji/subpopulacji.





Rys. 13. Okno analizy AMOVA oraz okno wyboru zmiennych

Źródło zmienności	1	2	3	4	5	6
	% zmienności	# stop.swob.	Suma kwadratów	Średni kwadrat	Statystyka $\Phi$	Istotność
(Population)	45,9	40	465,5500	11,6388	0,5411	****
Population	54,1	5	564,0804	112,8161		
Ogół		45	1029,6304	22,8807		

Rys. 14. Wyniki AMOVA

Zmienność wewnątrz wyznaczonych populacji (zmienności pomiędzy obiektami należącymi do tej samej populacji) stanowiła 45,9% ogółu zmienności, a 54,1% stanowiła zmienność pomiędzy wyodrębnionymi populacjami. Wyznaczona statystyka  $\Phi$  wskazywała na istotne różnicowanie wewnątrz wyodrębnionych populacji.

## Podsumowanie

Nauki przyrodnicze otrzymały nowe wsparcie w zakresie analizy danych empirycznych w postaci zestawu narzędzi zawartych w *Zestawie Przyrodnika*. Znajdują się tu narzędzia przydatne zarówno w pracy badawczej, jak i w nauczaniu przedmiotów związanych z analizą danych empirycznych w badaniach przyrodniczych.

Przedstawione przykłady to tylko kilka spośród dostępnych w *Zestawie Przyrodnika* analiz i kreatorów. Znajdują się tam narzędzia podstawowe związane z wyznaczaniem podstawowych statystyk opisowych, planowaniem i analizą wyników (wraz z porównaniami szczegółowymi) uzyskanych z doświadczeń w klasycznych układach czynnikowych (jedno-, dwu- i trójczynnikiowe układy całkowicie losowe, bloki losowe, kwadrat łaciński, split-plot, split-blok, split-split-plot, split-plot-blok, split-blok-plot itd.), z budowaniem modeli regresyjnych (kreator regresji liniowej) oraz z doborem właściwego testu istotności (kreator testów). Wśród dostępnych narzędzi zaawansowanych należy wymienić doświadczenia kratowe, hodowlane, jednopowtórzeniowe z wzorcem systematycznym, a także analizę serii doświadczeń oraz analizę wariancji molekularnej AMOVA. Dostępne narzędzia wielowymiarowe obejmują analizę AMMI oraz analizę aglomeracyjną (analizę skupień).

Takie narzędzia jak *Zestaw Przyrodnika* są bardzo potrzebne zarówno dla pracowników naukowych, jak i nauczycieli akademickich zajmujących się naukami przyrodniczymi. Powinny być również dalej rozwijane o kolejne potrzebne naukowcom analizy i narzędzia.

## Literatura

1. Cochran, W.G., Cox, G.M., 1992. Experimental design. John Wiley & Sons Inc., Hoboken, USA.
2. Excoffier L., Smouse PE., Quattro JM., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131:479–491.

3. Mądry W., Mańkowski D.R., Kaczmarek Z., Krajewski P., Studnicki M., 2010. Metody statystyczne oparte na modelach liniowych w zastosowaniach do doświadczalnictwa, genetyki i hodowli roślin. Monografie i Rozprawy Naukowe IHAR-PIB Nr 34, Radzików.
4. Mądry, W., 2003. Doświadczalnictwo – doświadczenia czynnikowe. Fundacja Rozwój SGGW, Warszawa.
5. Mądry, W., Iwańska, M., 2011. Ilościowe miary szerokiej adaptacji odmian i ich zastosowanie w doświadczeniach wstępnych z pszenicą ozimą. Biuletyn IHAR 260/261, 81–95.
6. Mańkowski D.R., 2013. Modele równań strukturalnych SEM w badaniach rolniczych. Monografie i Rozprawy Naukowe IHAR-PIB Nr 42, Radzików.
7. Mańkowski D.R., Laudański Z., Janaszek M., 2011. Przydatność wybranych miar podobieństwa dla danych binarnych do analiz wielocechowych w badaniach molekularnych. Biuletyn IHAR, Nr 262: 155–173.
8. Montgomery, D.C., 2005. Design and analysis of experiment. 6th edition. John Wiley & Sons Inc., Hoboken, NJ, USA.
9. Paderewski J. 2008. Przydatność modelu AMMI do badania reakcji roślin rolniczych na warunki środowiskowe. Rozprawa doktorska, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie.
10. Paderewski, J., Gauch, H.G., Mądry, W., Drzazga, T., Rodrigues, P.C., 2011. Yield Response of Winter Wheat to Agro-Ecological Conditions Using Additive Main Effects and Multiplicative Interaction and Cluster Analysis. Crop Science 51, 969–980. doi:10.2135/cropsci2010.05.0278.
11. Reif J. C., Melcinger A. E., Frisch M. 2005. Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. Crop Science, 45: 1–7.
12. StatSoft Polska Sp. z o.o. 2017. Zestaw Przyrodnika wersja 1.0. [www.statsoft.pl](http://www.statsoft.pl).

13. Trajer J.T., Janaszek-Mańkowska M., Mańkowski D.R., 2016. Komputerowa analiza danych w badaniach naukowych. Wydawnictwo SGGW, Warszawa.
14. Ubysz-Borucka L., Mądry W., Muszyński S., 1985. Podstawy statystyczne genetyki cech ilościowych w hodowli roślin. Wydawnictwo SGGW-AR, Warszawa.
15. Wójcik, A.R., Laudański, Z., 1989. Planowanie i wnioskowanie statystyczne w doświadczalnictwie. PWN, Warszawa.