

Ocena skuteczności działania środków ochrony roślin

Evaluation biologique des produits phytosanitaires

Projekt i analiza badań oceniających skuteczność działania środków ochrony roślin

Zakres

Niniejsza norma została przygotowana dla zastosowania w związku z Normami EPPO z pakietu PP1 (Ocena skuteczności działania środków ochrony roślin) i zawiera szczegółowe wskazówki odnośnie do projektowania i analizy badań oceniających skuteczność.

Zatwierdzenie normy i poprawki

Po raz pierwszy zatwierdzono we wrześniu 1989 r.
Pierwsza poprawka zatwierdzona we wrześniu 1998 r.
Druga poprawka zatwierdzona we wrześniu 2006 r.

Wprowadzenie

Zadaniem niniejszej normy jest przedstawienie ogólnych informacji związanych z projektowaniem i analizą badań oceniających skuteczność. Normy EPPO dotyczące oceny skuteczności działania środków ochrony roślin zawierają bardziej szczegółowe wskazówki dla tego typu badań dla indywidualnych układów żywiciel/agrofag. Pierwszym ustalonym czynnikiem jest układ badania (projekt badania, rozmiar i układ poletka, rola i lokalizacja poletek kontrolnych). Następnie dokonywany jest przegląd charakteru obserwacji (rodzaje zmiennych, sposoby obserwacji). Wreszcie, wysuwane są sugestie dotyczące analizy statystycznej wyników badania oraz serii badań (oceny efektów, wybór testu statystycznego, przekształcanie zmiennych). W załączniku 1 zawarte zostały przykłady skal wykorzystywanych w normach EPPO.

Następne działania mają za zadanie naszkicować zarys dobrej praktyki statystycznej podczas analizowania danych. Nie jest to, gdyż nie może być, recepta, którą można by zastosować przy wszystkich analizach lub która obejmowałaby wszystkie sytuacje. Praktycy nie powinni nigdy przeceniać potrzeby uzyskania profesjonalnych porad statystycznych. Ważne jest, aby praktycy

rozumieli wskazówki, które uzyskują od specjalistów. Nierzadko lepiej jest, aby przeprowadzali oni proste analizy, które mogą opisać oraz bronić ich pewną argumentacją, niż aby akceptowali porady, na podstawie których mieliby przeprowadzić analizy zrozumiałe przez nich jedynie częściowo. Pomocna może okazać się bibliografia zawarta na końcu niniejszej normy. Obejmuje ona dobrej jakości publikacje, które mają na celu zaprezentowanie zasad dobrej praktyki statystycznej, zamiast wskazywania statystycznych recept do stosowania bez zastanowienia.

1. Projekt doświadczenia

1.1 Zakres i cele doświadczenia

Przed rozpoczęciem projektowania badań należy w jasny sposób określić ich zakres i cele, ponieważ zawęża to ilość dostępnych sposobów projektowania.

W praktyce często stosowany jest proces powtarzalny: zakres i cele są stopniowo dostosowywane do dostępnych zasobów doświadczalnych. Ważne jest, aby zakres i cele były uaktualniane w celu odzwierciedlenia przez nie decyzji podjętych w trakcie procesu.

Zakres badania odzwierciedla zasięg praktycznych wyników generowanych przez badanie, które są zgodne z jego celami. Część zakresu badania ma związek z populacją, z której pochodzi próbka badania. Inna część decyduje o zakresie warunków środowiskowych, roślin, produktów chemicznych stosowanych podczas zabiegów, metod stosowania oraz zwalczanych agrofagów, określenie czego ma na celu dane badanie. Zakres określa kontekst, w którym badane są jednostki i obserwacje doświadczalne.

Cele badania powinny przybrać formę pytań dotyczących zabiegów, wraz z pożądanymi odpowiedziami na te pytania. Typowymi odpowiedziami będą „tak” lub „nie”, klasyfikacja zabiegów lub oszacowana wartość.

Zakres i cele powinny tworzyć część protokołu doświadczenia, co zostało określone w Normie EPPO PP 1/181 *Prowadzenie i opis doświadczeń oceniających skuteczność, w tym dobrej praktyki eksperymentalnej*. Planowane metody doświadczalne, projekt i analiza opisane poniżej powinny również stanowić część protokołu.

1.2 Rodzaje projektów

Normy EPPO dotyczące oceny skuteczności stosowania środków ochrony roślin przewidują badania, w których badaniom doświadczalnym są poddawane „produkty badane, produkty porównawcze oraz poletka kontrolne, uporządkowane wg odpowiedniego planu statystycznego”. Przewidują one także, że produkty mogą być poddawane badaniom w różnych dawkach i/lub terminach stosowania. Ma to zastosowanie w szczególności przy wyższych dawkach w ogólnych badaniach selektywności oraz badaniach dotyczących reakcji na dawkę.

Rys. 1. Układ całkowicie zrandomizowany. Każdy zabieg (znakowany 1-8) jest powtarzany czterokrotnie; poszczególne oznakowania zabiegów przydzielane są w sposób całkowicie losowy do 32 poletek.

2	7	3	7	8	3	5	4
1	2	6	2	2	3	4	6
8	4	5	4	6	8	1	5
1	5	7	8	1	7	3	6

Rys. 2 Możliwe rozmieszczenie bloków i poletek w zrandomizowanych blokach podczas badań w warunkach polowych. Pod uwagę bierze się gradient środowiskowy, albo poprzez rozmieszczenie bloków w dół pochylenia lub poprzez ustawienie bloków jeden przy drugim. W każdym z tych przypadków, na każde z poletek w obrębie bloków rozmieszczonych w dół pochylenia, zmienna środowiskowa ma równy wpływ.

Blok 1	3	8	7	2	5	4	6	1
Blok 2	4	7	5	1	6	2	8	3
Blok 3	5	6	7	2	8	3	1	4
Blok 4	8	4	1	3	5	6	7	2

Blok 1				Blok 2				Blok 3				Blok 4																			
5	7	1	2	8	4	3	6	4	6	1	5	3	8	2	7	3	8	2	5	4	7	6	1	2	3	1	8	5	6	7	4

Rys. 3 Możliwe rozmieszczenie bloków i poletek w zrandomizowanych blokach podczas badań w warunkach polowych. Alternatywna forma układu losowego rozmieszczenia bloków w sytuacji, w której nie występuje jednoznaczny gradient środowiskowy, natomiast ich heterogeniczność musi zostać domniemana, ponieważ maksymalne odległości pomiędzy poletkami w bloku są stosunkowo duże.

W tym schemacie osiem poletek jest rozmieszczonych w stosunkowo niewielkiej odległości od siebie w prostokącie 4 x 2, zaś bloki są rozmieszczone jeden obok drugiego.

Blok 1		Blok 2		Blok 3		Blok 4	
3	1	8	1	8	2	3	7
6	4	2	6	6	5	1	6
8	5	7	5	3	1	5	8
7	2	3	4	7	4	4	2

Rys. 4 Kolejny przykład rozmieszczenia bloków i poletek w przypadku, gdy podobnie jak na rys. 3, heterogeniczność jest domniemana, lecz nie występuje oczywisty gradient środowiskowy.

Blok 1	2	7	3	1	8	5	2	7	Blok 2
	8	5	4	6	6	3	4	1	
Blok 3	3	6	8	7	6	3	5	2	Blok 4
	1	4	5	2	7	4	8	1	

W tym schemacie osiem poletek jest również rozmieszczonych w stosunkowo niewielkiej odległości od siebie w prostokącie 4 po 2, jednakże bloki są rozmieszczone w siatce 2 po 2.

Plany jednoczynnikowe są odpowiednie dla badań, jeżeli badane produkty, produkty porównawcze oraz poletko kontrolne mogą być uznane za różne poziomy tego samego parametru oraz jeżeli nie występują inne czynniki wymagające zbadania. Jednakże, jeśli przykładowo, efekt każdego środka ochrony roślin w badaniu skuteczności ma zostać zbadany przy różnych dawkach, wówczas można zastosować projekt czynnikowy wraz z wszelkimi kombinacjami zabiegów dla obu reprezentowanych czynników. W ten sposób można uzyskać informacje o ważnym współoddziaływaniu pomiędzy czynnikami, które można poddać ocenie.

Główne układy losowe możliwe do zastosowania są następujące: układ całkowicie losowy oraz układ losowanych bloków. Zostały one zilustrowane poniżej, na podstawie przykładu jednoczynnikowego zakładającego zastosowanie ośmiu poletek, tj. pięć różnych badanych produktów, dwa produkty porównawcze oraz poletko kontrolne, przy czym każde poletko poddawane zabiegowi jest w czterech egzemplarzach.

1.2.1 Układ całkowicie losowy

Zabiegi przeprowadzane w układzie całkowicie losowym (rys. 1) są przypisywane losowo do jednostki doświadczalnej. Układ ten jest potencjalnie najskuteczniejszy statystycznie (w sensie najwyższego prawdopodobieństwa wykrycia znacznej różnicy, jeśli takowa występuje), ponieważ umożliwia on zachowanie maksymalnej liczby stopni swobody wariancji resztkowej. Jednakże jest on odpowiedni jedynie wówczas, jeśli obszar poddany badaniu jest obszarem jednorodnym środowiskowo. Jeżeli występuje znaczna heterogeniczność pomiędzy różnymi częściami obszaru poddanego badaniu, wariancja resztkowa może być niedopuszczalnie wysoka, w związku z tym lepiej jest zastosować układ biorący to pod uwagę, taki jak układ losowanych bloków.

1.2.2 Układ bloków kompletnie zrandomizowanych

Blok składa się z grupy poletek, w obrębie którego środowisko właściwe do prowadzenia obserwacji jest jednorodne. W tym układzie bloki są rozmieszczone rozmyślnie, aby poletka w ich obrębie były jak najbardziej jednolite przed przeprowadzeniem zabiegów. Zwykle każdy zabieg stosowany jest tylko raz w każdym z bloków. Zabiegi są ordynowane w sposób losowy na poletkach w obrębie bloków, co działa niczym powtórzenia. Układ zabiegów w każdym z bloków powinien być losowy dla każdego z poszczególnych bloków. Poniższe przykłady (rys. 2-4) zakładają wystąpienie czterech bloków i ośmiu zabiegów. Rozmieszczenie bloków ma na celu kontrolę heterogeniczności miejsca doświadczenia (np. nachylenie, kierunek prac przy zasiewie lub sadzeniu, narażenie na ryzyko, stopień zagrożenia inwazją agrofagów itp.), roślin (rozmiar, wiek, żywotność) lub warunków występujących podczas przeprowadzania doświadczenia (stosowanie zabiegów, oceny).

W związku z powyższym, rozmieszczenie bloków wymaga uprzedniego zgromadzenia informacji na temat obszaru poddanego badaniu. Na układ poletek w obrębie bloków może mieć wpływ ich kształt: długie, wąskie poletka są często rozmieszczone jedno obok drugiego, zaś poletka kwadratowe mogą być rozmieszczane w innych układach.

Jednakże bloki nie muszą być rozmieszczone jeden obok drugiego. W przypadku wstępnego dobrego zapoznania się z obszarem poddanym badaniu, można wykorzystać ten fakt poprzez rozproszenie bloków na obszarze badanego pola, w celu uwzględnienia uprzednio zaobserwowanej heterogeniczności (rys. 5 i 6). Mimo iż istnieje niewielka możliwość, że w losowym układzie podczas powtórzenia, zabiegi mogą odbywać się w kolejności stosowania, należy tego w miarę możliwości unikać, aby dokonać obiektywnej oceny. Jeżeli wstępna znajomość terenu jest doskonała i można ocenić, że warunki pozostaną identyczne przez cały okres trwania doświadczenia, można uwzględnić złożoną heterogeniczność, i nie jest nawet konieczne, aby poletka w obrębie jednego bloku przylegały do siebie. Dla przykładu, układ bloków może zostać rozbity, aby wyjaśnić wiadome, niejednolite zainfekowanie nicieniami. Na rys. 6 poletka w obrębie bloku 1 zostały celowo rozmieszczone w punktach o wyraźnie niskim poziomie zainfekowania, zaś poletka w obrębie bloku 2 w punktach wyraźnie wysokiego poziomu zainfekowania.

Oczywiście wybór układu, wymiarów i ustawienia zastosowanych bloków, jeżeli takowe zostały zastosowane, zależy od heterogeniczności zaobserwowanej na obszarze poddanym badaniu (np. gleby, nachylenia, narażenia na ryzyko, zainfekowania agrofagami, odmiany uprawnej, itp.). Zmienne takie nigdy nie są jednorodne, zaś układ losowanych bloków na umiarkowanie jednorodnym obszarze pozwala zwykle na uzyskanie bardziej użytecznych informacji dotyczących wydajności produktu niż badanie całkowicie losowe na obszarze, który wydaje się być jednorodny, choć nim nie jest. Rozmieszczenie bloków zależy także od rozmiaru i

kształtu poletka (rys. 5 i 6). Ogólnie rzecz biorąc, mniejsze bloki są bardziej skuteczne pod względem zmniejszania poziomu heterogeniczności. W badaniach o wysokiej liczbie zabiegów należy rozważyć zastosowanie innych układów (np. kwadrat łaciński, układy o blokach niekompletnych).

Badania opierające się na zastosowaniu losowanych bloków, przeprowadzane w różnych regionach o odmiennych warunkach środowiskowych i/lub w różnych latach mogą być uznane w odpowiednich przypadkach jako serie badań. Jest więc konieczne, aby w analizie statystycznej rozdzielić dodatkowe wariancje pomiędzy miejscami badań od wariancji pomiędzy blokami oraz aby oszacować współoddziaływanie zabiegów miejscowych, co może być przedmiotem szczególnego zainteresowania. Należy zauważyć, że w każdym poszczególnym badaniu należy od nowa randomizować zabiegi w każdym z bloków.

1.2.3 Układ z dzielonymi poletkami (Split plot)

Przy przeprowadzaniu doświadczeń wieloczynnikowych najczęściej stosowanym układem jest układ bloków kompletnie zrandomizowanych, przy czym każda kombinacja zabiegów występuje jednorazowo w każdym z bloków. Jednakże czasami jeden z czynników nie może zostać w pełni zrandomizowany do poletek w obrębie danego bloku. Przykładowo, przyjmijmy, że w doświadczeniu występują dwa czynniki: preparat (o czterech poziomach, oznaczonych 1-4) oraz narzędzia wykorzystywane do uprawy roślin (o trzech poziomach, oznaczonych A, B, C) oraz, że poletka są stosunkowo niewielkie. Wówczas rozmiar sprzętu wykorzystywanego do przeprowadzenia zabiegów przy uprawie może uniemożliwić przeprowadzenie pełnej randomizacji na 12 poletkach w obrębie każdego z bloków. W takim przypadku zalecany jest układ z dzielonymi poletkami, w którym, w każdym poletku, podpoletka pogrupowane są po cztery, co daje trzy pełne poletka na dany blok, uprawa jest losowo ograniczona do całych poletek, zaś środek będący czynnikiem jest osobno zrandomizowany do podpoletek w obrębie całych poletek (rys. 7). Przy zastosowaniu układu z dzielonymi poletkami wymagana jest nieco bardziej złożona analiza wariancji, obejmująca dwie warstwy, z których każda posiada odrębny błąd średniokwadratowy, wobec którego należy badać wpływ różnych czynników oraz ich wzajemne oddziaływanie.

1.2.4 Układy systematyczne

Nierandomizowane układy systematyczne nie są nigdy odpowiednimi układami do przeprowadzenia oceny skuteczności, z wyjątkiem kilku bardzo szczególnych przypadków (np. próby różnorodności na selektywność herbicydów). Ogólnie rzecz biorąc, są one odpowiednie jedynie w badaniach demonstracyjnych.

1.3 Skuteczność

Planując doświadczenia ważne jest, aby ustalenie koniecznej skuteczności wszystkich testów statystycznych, które mają zostać przeprowadzone. Skuteczność testu jest to prawdopodobieństwo wykrycia danej różnicy pomiędzy zabiegami, jeśli takowa różnica istnieje. Skuteczność zależy od pewnej liczby parametrów, m.in.:

- dokładności wyników (wariancja resztkowa)
- liczby powtórzeń, w tym powtórzeń w miejscu badania.

Należy wybrać układ dający możliwość wykrycia, ze statystyczną istotnością, różnicy mającej znaczenie praktyczne dla kontrastu będącego przedmiotem zainteresowania. Można mieć również związane z tym wymaganie, aby przedział ufności dotyczący ocen zabiegów nie przekraczał

uprzednio określonej szerokości. Przed rozpoczęciem badania należy dokonać wyboru pomiędzy przeprowadzeniem jednego badania lub serii badań.

Zgodnie z normą EPPO PP 1/226 *Liczba badań skuteczności*, wydajność środka ochrony roślin powinna zostać wykazana poprzez przeprowadzenie pewnej liczby badań w różnych miejscach, regionach i latach, w odmiennych warunkach środowiskowych. W związku z tym, aby zbadać wydajność środka ochrony roślin można zaplanować, przeprowadzić i dokonać oceny serii badań (aby uzyskać informacje dotyczące definicji serii badań, patrz także 3.4.1).

Ogólnie rzecz biorąc, można posługiwać się wynikami poprzednich doświadczeń, w celu wskazania możliwego zróżnicowania obserwacji. Jeżeli występują takie dane, możliwe jest dokonanie oceny układu i rozmiaru doświadczenia, aby uzyskać wymaganą skuteczność doświadczenia. Czasem możliwe jest ustalenie wymaganych ilości na podstawie rozważań teoretycznych. Dla przykładu, w przypadku danych dwumiennych, można ustalić górną granicę dla zróżnicowania proporcji. Dostępne są różne systemy komputerowe lub graficzne służące pomocą przy ustalaniu liczby koniecznych powtórzeń. Biorą one pod uwagę wielkość różnicy, którą należy oszacować, lub poziom ważności wymagany dla takiej różnicy oraz oczekiwaną dokładność. W następnej części dokumentu znajduje się kilka prostych zasad ogólnych.

1.4 Liczba zabiegów i powtórzeń w odniesieniu do stopni swobody

Aby przeprowadzić użyteczną analizę statystyczną, liczba resztkowych stopni swobody powinna być wystarczająco wysoka. W badaniu zakładającym przeprowadzenie 8 zabiegów i 4 powtórzeń w układzie losowanych bloków liczba resztkowych stopni swobody wynosi 21. Są one obliczane jako: całkowita liczba stopni swobody ($32 - 1 = 31$) minus stopnie swobody zabiegów ($8 - 1 = 7$) minus stopnie swobody bloków ($4 - 1 = 3$), tj. $31 - 7 - 3 = 21$. W badaniu zakładającym przeprowadzenie 3 zabiegów i 4 powtórzeń w 4 miejscach, liczba resztkowych stopni swobody wynosi 24. Są one obliczane jako: całkowita liczba stopni swobody ($48 - 1 = 47$) minus stopnie swobody zabiegów ($3 - 1 = 2$) minus stopnie swobody miejsc prowadzenia badania ($4 - 1 = 3$) minus stopnie swobody współoddziałujących zabiegów według miejsca prowadzenia badania ($(3 - 1) * (4 - 1) = 6$) minus stopnie swobody powtórzenia w miejscach prowadzenia badania ($((4 - 1) * 4 = 12)$), tj. $47 - 2 - 3 - 6 - 12 = 24$.

Liczba resztkowych stopni swobody powinna zwiększać się przy zwiększonej liczbie powtórzeń, zabiegów lub liczbie miejsc prowadzenia badań. Pożądana liczba resztkowych stopni swobody zależy od stopnia dokładności (skuteczności) wymaganego od badania. W razie wątpliwości należy zasięgnąć specjalistycznej porady statystycznej. Ogólnie rzecz biorąc, na podstawie doświadczeń zdobytych przy przeprowadzaniu badań/serii badań oceniających skuteczność stwierdza się, że nie należy przeprowadzać badań/serii badań o resztkowych stopniach swobody niższych niż 12. Jeżeli z jakiegokolwiek ważnego powodu zaleca się przeprowadzenie jedynie 3 powtórzeń oraz 3 zabiegów, wówczas badanie może zostać przeprowadzone w co najmniej 4 miejscach, aby uzyskać minimalną liczbę 15 resztkowych stopni swobody, wymaganą do przeprowadzenia użytecznej analizy statystycznej.

Wybór układu doświadczalnego również ma wpływ na liczbę resztkowych stopni swobody. Układ całkowicie losowy pozwala uzyskać maksymalną liczbę. Układ losowanych bloków wykorzystuje niektóre z tych stopni swobody, aby uwzględnić heterogeniczność środowiska (jak przykładowo wzdłuż jednego gradientu). Układ z dzielonymi poletkami wykorzystuje stopnie swobody, aby uzyskać możliwe źródła więcej niż jednego składnika zmienności. Osoba prowadząca doświadczenie powinna spróbować pozostawić maksymalną liczbę stopni swobody, aby oszacować wariancję resztkową podczas dokonywania wyboru optymalnego układu, w celu

zminimalizowania zmienności, uwzględniając wszelkie znane źródła heterogeniczności (patrz norma EPPO PP 1/181).

Tabela 1 umożliwia zaczerpnięcie informacji na temat związku pomiędzy liczbą powtórzeń i liczbą resztkowych stopni swobody dla różniącej się liczby zabiegów i miejsc.

1.5 Jednostki/poletka doświadczalne: rozmiar, kształt, potrzeba wytyczenia granic

Jednostką doświadczalną nazywamy część materiału badawczego, na której przeprowadzany jest pojedynczy zabieg oraz na której prowadzone są obserwacje.

Do przeprowadzenia planowanych zabiegów i powtórzeń konieczna jest wystarczająca liczba jednostek.

W praktyce materiał badawczy jest ograniczony, dlatego też nierzadko trzeba iść na kompromis. Przykłady jednostek doświadczalnych są następujące: obszar uprawy (poletko), pojemnik zawierający jedną lub więcej roślin, część rośliny (np. liść, łodyga, gałąź) oraz miejsce nęcenia agrofagów na polu. Należy w taki sposób wybrać jednostki doświadczalne, aby były one reprezentatywne dla populacji, która jest poddana badaniu oraz aby były one w jak największym stopniu jednolite. Brak jednolitości może czasami zostać złagodzony dzięki replikowaniu bloków.

Ogólnie rzecz biorąc, poletka powinny mieć prostokątny kształt oraz powinny mieć taki sam rozmiar podczas jednego badania, zaś przy serii badań ich rozmiar powinien być podobny. Dokładność zwiększa się wraz z rozmiarem poletka, jednakże tylko do pewnej granicy, jako że zmienność gleby oraz warunki zainfekowania również przejawiają tendencję wzrostową. Długie, wąskie, prostokątne poletka są odpowiednie do prowadzenia zmechanizowanego zbierania plonów. Poletka o kształcie zbliżonym do kwadratu stwarzają ryzyko wzajemnego zakłócania się poletek. Do celów obserwacji agrofagów zgrupowanych przestrzennie, takich jak niektóre chwasty i choroby odglebowe, lepsza jest większa ilość mniejszych poletek niż mniejsza ilość większych poletek.

Rozmiar poletka jest podany w określonych normach EPPO dla poszczególnych kombinacji uprawa/agrofag. W przypadkach, w których możliwe jest wystąpienie zakłóceń pomiędzy poletkami, poletka muszą być większe (poletko brutto), zaś obserwacje będą ograniczone do obszaru środkowego (poletko netto).

Różnica pomiędzy poletkiem netto a poletkiem brutto zwana jest obszarem odrzucenia. Ogólnie rzecz biorąc, normy EPPO sugerują rozmiary poletek netto, zaś decyzja o rozmiarze poletka brutto jest pozostawiona osobie przeprowadzającej badanie, która powinna określić obszary odrzucenia po rozważeniu wszelkich potencjalnych źródeł zakłóceń pomiędzy poletkami w każdym badaniu lub serii badań. Jednym z powszechnych źródeł zakłóceń jest rozsiew środka (przykładowo, oprysk lub znoszenie oparów lub boczny ruch na/w ziemi) poza obręb poletka skażając przylegające poletka. Może to mieć szczególne znaczenie w przypadku oprysków roślin wysokich. Jednakże, im większe obszary odrzucenia, tym większa możliwość minimalizacji błędów doświadczalnych.

Innym powszechnym źródłem zakłóceń jest ekspansja agrofaga (przykładowo grzyby przenoszone drogą powietrzną lub wysoce mobilne owady) z poletek nie poddanych działaniu środka lub na których zwalczanie agrofagów daje słabsze wyniki. Tego typu ekspansja może zarówno zwiększyć populację agrofagów w obrębie poletek, na których prowadzone są bardziej skuteczne zabiegi oraz zmniejszyć ją w obrębie poletek poddanych mniej skutecznym zabiegom. Podobna sytuacja ma miejsce w przypadku, gdy środek jest testowany na uprawie, wobec której

praktykowane jest zintegrowane zwalczanie, wówczas negatywny wpływ na drapieżniki i pasożyty może być maskowany przez ich migrację pomiędzy poletkami.

Innym źródłem zakłóceń jest konkurencja roślin o uzyskanie dostępu do światła i środków odżywczych. Ma to szczególne znaczenie, gdy chcemy zmierzyć wielkość plonów. Jeżeli obszary ochronne pomiędzy poletkami różnią się od samych poletek (np. ścieżki, inna roślina uprawna), należy zachować ostrożność przy dokonywaniu wyboru obszaru przeznaczonego do oceny.

W zależności od zastosowania lub używanych sprzętu do zbierania plonów, rozmiar poletka netto może zostać zwiększony powyżej rozmiaru koniecznego do przeprowadzenia obserwacji.

Poletka mogą być rozmieszczone w poprzek lub wzdłuż kierunku prac (zasiewu lub sadzenia). Układ poprzeczny (rys. 8) posiada tę przewagę, że jeśli w trakcie prac popełnione zostaną błędy (przy uprawie, zasiewie, itp.), będzie to prawdopodobnie w równym stopniu oddziaływać na wszystkie bloki. Jednakże w takim wypadku zabiegi i zbiory plonów stają się trudniejsze. Układ wzdłużny posiada przewagę praktyczną przy stosowaniu zabiegów i przy zbiorach plonów, jednakże stwarza zagrożenie większej heterogeniczności w bardzo długich blokach. Układ mieszany może stanowić kompromis.

1.6 Rola i lokalizacja poletek kontrolnych

1.6.1 Cel poletek kontrolnych

Główną cechą „poletek kontrolnych nie poddanych zabiegowi kontroli” jest to, że nie zostały one poddane działaniu jakichkolwiek badanych środków ochrony roślin. Na poletkach kontrolnych nie poddane zabiegom powinny być jednak wykonywane wszystkie czynności jednakowo wykonywane podczas badania, zwłaszcza czynności związane z uprawą oraz produkty stosowane przeciw agrofagom nie objęte badaniem. Mimo iż poletka kontrolne nie są zwykle poddawane żadnym zabiegom przeciwko agrofagom podlegającym badaniu, w niektórych przypadkach modyfikacja nie poletka kontrolnego może być przydatna dla uwzględnienia niektórych czynności wykonywanych na innych poletkach. Przykładowo, gdy na innych poletkach stosowane są preparaty w postaci oprysku przeprowadzanego za pomocą urządzeń opryskujących, poletko kontrolne może być zmodyfikowane w taki sposób, aby było poddane opryskowi samą wodą. Ideą jest powtarzanie, w miarę możliwości, czynności wykonywanych na innych poletkach, jedynie z wyjątkiem zastosowania samego preparatu.

Poletko kontrolne ma głównie na celu wykazanie odpowiedniego zainfekowania agrofagiem. Nie można na przykład wykazać skuteczności środka, a wyniki nie są znaczące, jeżeli poletko kontrolne nie potwierdza inwazji odpowiedniego agrofaga. Potwierdzenie takie może mieć charakter jakościowy (obecność gatunków dominujących, rodzaju flory, chwasty itp.) lub ilościowy (zgodność z minimalnymi i maksymalnymi wartościami progowymi, rozkład przestrzenny). W wyjątkowych okolicznościach wykorzystanie poletka kontrolnego może być niemożliwe (np. w przypadku agrofagów poddanych kwarantannie).

W zależności od celu i rodzaju doświadczenia, poletka kontrolne pełnią pozytywną rolę, a nawet wiele ról jednocześnie. Są to między innymi:

- wykazanie skuteczności nowego produktu i produktu porównawczego. Główny dowód skuteczności nowego lub porównawczego produktu jest uzyskiwany zawsze poprzez porównanie z poletkiem kontrolnym
- pomoc w obserwacjach. Wzrokowego oszacowania uszkodzeń lub skali zainfekowania można czasem dokonać w ujęciu względnym, poprzez porównanie z wynikami kontroli
- zastosowanie techniki „sąsiedniego poletka kontrolnego” w celu zmierzenia i uwzględnienia rozkładu przestrzennego w obrębie poletek

- obserwacje rozwoju agrofaga (wschód, lot, wydzielanie zarodków, itp.), w szczególności jako podstawa do określenia terminów stosowania lub obserwacji
- zapewnienie rezerwy materiału inokulacyjnego w celu upewnienia się, że poziom materiału inokulacyjnego nie jest zbyt niski lub nie jest zbyt heterogeniczny (w ekstremalnych przypadkach, może to być praktycznie jednoznaczne ze sztucznym zainfekowaniem)
- pomoc w interpretacji wyników badań. Dla przykładu, znaczna różnica pomiędzy dwoma poletkami poddawanymi zabiegom nie musi mieć jednakowego znaczenia w zależności od poziomu zainfekowania.
- zapewnienie większej dostępności wyników analizy dla użytkowników, poprzez wyrażenie ich w innej formie lub umożliwiając ich przedstawienie graficzne (np. przekształcenie wskaźnika śmiertelności na wskaźnik skuteczności)
- umożliwienie przeprowadzenia dodatkowych obserwacji, w szczególności plonów ilościowych lub jakościowych, co może być interesujące przy połączeniu z innymi wynikami badań
- wreszcie, w wyjątkowych przypadkach, tworzenie ujęcia porównawczego dla poletek poddanych zabiegom w razie braku produktu porównawczego. Może to mieć na przykład miejsce w sytuacji, gdy rodzaj produktu lub jego zastosowanie są nowe lub gdy wszelkie dostępne produkty porównawcze zostały wycofane z użytku. Rola ta jest wówczas podobna do roli produktu porównawczego, mimo iż jego interpretacja jest zupełnie inna. Poletka kontrolne można więc porównać z różnymi poletkami poddawanymi zabiegom przy zastosowaniu formalnych badań poziomu istotności, w sposób identyczny, w jaki produkt porównawczy jest porównywane z nimi w zwykłych badaniach.

1.6.2 Rodzaje układów poletek kontrolnych

Możliwe są cztery układy poletek kontrolnych.

Poletka kontrolne włączone: Poletka kontrolne rozpatrywane są w taki sam sposób jak wszystkie inne poletka poddawane zabiegom, poletka kontrolne mają ten sam kształt i rozmiar jak inne poletka, i są randomizowane podczas badań. Poletka kontrolne włączone jest najczęstszym sposobem przeprowadzania badań, natomiast inne sposoby są stosowane w wyjątkowych przypadkach (głównie przy testach herbicydów).

Poletka kontrolne imbrykowane: w badaniu poletka kontrolne są rozmieszczane systematycznie. Rozmiar i kształt poletek nie musi być taki sam jak innych poletek w badaniu. Obserwacje poczynione w obrębie tych poletek posiadają inny charakter i nie powinny być włączane do analizy statystycznej. Takie ułożenie ma na celu zapewnienie bardziej jednorodnego rozkładu oddziaływania przyległego obszaru nie poddanego zabiegom, niż jest to możliwe w przypadku randomizowanego układu włączonych poletek kontrolnych. Możliwe są różne układy, poletka mogą być umieszczone pomiędzy blokami lub pomiędzy poletkami poddanyymi zabiegom w obrębie bloków (rys. 9).

Poletka kontrolne wydzielone: poletka kontrolne są wybierane poza obszarem badania i do niego nie przylegają, natomiast obszar, na którym się znajdują musi posiadać warunki bardzo zbliżone do warunków panujących na obszarze badanym. Powielenia nie są konieczne, lecz mogą być przydatne, gdy obszar nie jest jednorodny. Obserwacje poczynione w obrębie tych poletek nie powinny być włączane do analizy statystycznej.

Poletka kontrolne przyległe: każde poletko jest podzielone na dwa podpoletka, a jedno z nich (wybrane losowo) nie jest poddawane zabiegom. Obserwacje na obu podpoletkach są prowadzone w sposób identyczny. Obserwacje poczynione na tych poletkach nie powinny być włączane do analizy statystycznej, chyba że odpowiednio uwzględniony zostanie fakt, że układ ma formę dzielonych poletek. W układzie z dzielonymi poletkami zróżnicowanie w obrębie

poetek może różnić się od tego pomiędzy poletkami, w związku z tym analiza wariancji powinna uwzględniać dwie warstwy błędu. W celu zinterpretowania wyników konieczna może okazać się specjalistyczna porada statystyczna.

1.6.3 Wybór rodzaju układu poletka kontrolnego

Wybór rodzaju układu poetek kontrolnych zależy od ich roli (ról) w badaniu. Pomimo tego, że poletko kontrolne włączone było bardzo często stosowane w przeszłości przy badaniach oceny skuteczności i nadal jest stosowane w praktyce, niekoniecznie jest metodą najbardziej stosowaną. Poniższy schemat decyzyjny zawiera wytyczne w tej kwestii.

(a) Jeżeli poletko kontrolne jest wykorzystywana w teście statystycznym, wówczas niezbędne jest „poletko kontrolne włączone”.

W innym przypadku można zastosować inny rodzaj poletka kontrolnego. W każdym przypadku należy wziąć pod uwagę heterogeniczność poetek.

(b) w przypadku dużej heterogeniczności przydatne jest „poletko kontrolne przyległe”.

Gdy heterogeniczność jest niska lub umiarkowana, należy przeanalizować wzajemne oddziaływania poetek kontrolnych z poletkami przyległymi.

(c) Jeżeli poletka kontrolne nie są podatne na zakłócenia przyległymi poletkami, wówczas przydatne są „poletka kontrolne imbrykowane”.

(d) Jeżeli poletka kontrolne są podatne na zakłócenia z przyległych poetek, wówczas należy zastosować „poletka kontrolne wyłączone”.

1.7 Wybór liczebności próby w obrębie poletka

Głównym celem pobierania kilku próbek w obrębie poletka jest zmniejszenie poziomu zróżnicowania oszacowanej średniej wielkości poletka do poziomu odpowiedniego dla szacowanej zmiennej. Liczebność próby powinna być wystarczająco duża, aby osiągnąć ten cel. Wymagana liczebność próby zależy w dużej mierze od charakteru obserwacji oraz zróżnicowania w obrębie poletka. Normy EPPO dotyczące oceny poszczególnych agrofagów, chwastów i chorób stanowią źródło informacji pomocniczych na temat liczebności prób. W praktyce liczebność próby wynosząca 10-50 składników jest zwykle wystarczająca do prawidłowego oszacowania średniej wartości w obrębie poletka, w zależności od naturalnego zróżnicowania. Należy zauważyć, że jeżeli na poletkach stosowane są zabiegi, wówczas zwiększenie liczebności próby daje jedynie ograniczony zwrot skuteczności, ponieważ porównania dokonywane pomiędzy zabiegami powinny odbywać się w skali pomiędzy poletkami.

Pobieranie próbek powinno zawsze być losowe i powinno w odpowiedni sposób obejmować obszar poletka oraz materiał doświadczalny. Z powodów praktycznych konieczne może być pobranie podpróbek. Przegląd metod pobierania próbek oraz odnośniki do odpowiedniej literatury znajdują się w Perry (1994).

2. Zasady oceny skutków stosowania środków ochrony roślin

Przy dokonywaniu oceny skutków stosowania środków w badaniu oceny skuteczności „zmiennie” oceniane są za pomocą „sposobów obserwacji”.

2.1 Zmienne

Znaczenie ma charakter zmiennej, jako że w ujęciu ogólnym ma on wpływ na wybór metody statystycznej wykorzystywanej w interpretacji wyników. Rozróżnić można kilka kategorii zmiennych.

Zmienne binarne (np. tak/nie, obecność/nieobecność): mogą one prowadzić do zmiennych dwumianowych, które odpowiadają liczbie przypadków, w których taki stan został zaobserwowany, biorąc pod uwagę znaną liczbę obserwacji (np. liczba zaatakowanych roślin na 20 losowo wybranych w obrębie poletka).

Zmienne nominalne: zmienne o ekwiwalentnej ważności, których nie da się uszeregować (nieporządkowe), np. gatunek agrofaga, różne barwy. Ogólnie rzecz biorąc, wartości tych zmiennych są wskazywane przez słowa.

Zmienne porządkowe: zmienne o wartościach będącymi klasami tworzącymi konkretny szereg, które nie są jednak mierzone. Zwykle są one zmiennymi jakościowymi, zaś klasy mogą być umieszczone względem siebie (np. zła, umiarkowana, dobra; opisowe stopnie zniszczenia liści). W innych przypadkach wartości mogą być wyrażone numerycznie (i mogą być dokładnie zmierzone, jednakże nie służy to celom praktycznym), np. kategorie pokrycia chwastami lub kategorie zakażenia mszycami.

Zmienne ilościowe: zmienne mierzalne i mierzone w praktyce, np. plon, wysokość roślin, liczba larw, procentowa powierzchnia liścia zaatakowana chorobą. Mogą być one dyskretne, jeżeli są wyrażane w liczbach całkowitych (np. wyniki obliczeń), lub mogą być ciągłe (np. waga, rozmiar). Zmienne ilościowe mogą również wynikać z odpowiednich działań matematycznych. Można uzyskać różnice lub sumy (np. różnica pomiędzy wartością przed i po zabiegu). Obliczyć można wartości względne, które mogą zostać wyrażone w formie proporcji lub stosunku. Proporcja odpowiada ilorazowi „ilości częściowej/ilości całkowitej” i zawiera się w przedziale pomiędzy 0 i 1 (np. ograniczona, ciągła zmienna ilościowa).

W praktyce, proporcja taka często jest częstotliwością względną: „częstotliwość jednej klasy/częstotliwość całkowita”, tj. zmienna jest dwumianowa lub wielomianowa. Wyrażana jest często jako % (tj. wartość pomiędzy 0 a 100). Stosunki nie posiadają górnej granicy, np. (wartość początkowa – wartość końcowa)/wartość początkowa. Mogą być również wyrażane w % (gdzie możliwe są wartości powyżej 100). Wartość wyrażona w procentach może w rzeczywistości być proporcją lub stosunkiem, lub nawet zmienną dwumianową i ważne jest, aby do celów statystycznych dokonać rozróżnienia pomiędzy tymi przypadkami.

2.2 Sposób obserwacji zmiennych

Niezbędne jest dokonanie pomiarów zmiennych w sposób możliwie jak najdokładniejszy. W praktyce należy podjąć następujące środki ostrożności przy dokonywaniu oceny badania:

- (a) ustalenie skali, klucza lub metody pomiaru przed rozpoczęciem badania. Wybrana metoda powinna być stosowana we wszystkich badaniach wchodzących w skład serii badań.
- (b) dokonywanie oceny bez uprzedniej wiedzy na temat planu zabiegów
- (c) prowadzenie prac w kolejności ułożenia bloków
- (d) stosowanie tej samej metody we wszystkich jednostkach doświadczalnych, np. obserwacje wszystkich poletek prowadzone w tym samym kierunku, w celu uniknięcia różnic w oświetleniu.

W celu dokonania oceny zmiennych, możliwe są cztery tryby obserwacji. pomiar, ocena wzrokowa, ustalanie rankingu i scoring.

2.2.1 Pomiar

Pomiar ustala wartości w sposób obiektywny. Wyniki pomiaru mogą stanowić zmienne ciągłe (waga, rozmiar) lub dyskretne (liczenia). W doświadczeniach polowych, gdy pomiar nie dotyczy

całego poletka, należy przeprowadzić go na próbie, której liczebność i sposób pobierania powinny zostać uprzednio określone, zgodnie z wymaganą dokładnością.

2.2.2 Ocena wzrokowa

Ocena wzrokowa ustala wartości w sposób subiektywny, jednakże za pomocą tej samej skali i zakresu wartości, jak w przypadku pomiarów, o których mowa powyżej. Ocena wzrokowa odnosi się zwykle do zmiennych ciągłych (np. pokrycie chwastami) lecz może także odnosić się do liczeń, jeżeli są one duże (liczba zmian patologicznych na liściu). Ogólnie rzecz biorąc, przeprowadzenie oceny wzrokowej jest łatwiejsze z odniesieniem lub poddaną działaniu kontrolą niż ma to miejsce w przypadku wartości bezwzględnych. Należy podkreślić, że uzyskiwany jest ten sam rodzaj zmiennej, co w przypadku pomiarach. Wyniki powinny być więc przedstawiane w tych samych jednostkach. W razie potrzeby, wartości uzyskane lub przekształcone mogą być uznawane za rzeczywiste oceny zmiennej ciągłej i mogą zostać przeanalizowane przy zastosowaniu normalnych procedur statystycznych.

Obserwator powinien zostać przeszkolony pod względem prowadzenia ocen, zaś jego obserwacje powinny zostać wyskalowane do normy. Jeśli spełnione zostaną te warunki, oceny mogą wykazywać bardzo dużą dokładność. Dokładność może jednak wahać się w zależności od badanej zmiennej. Przykładowo, przy poddawaniu ocenie % porażenia powierzchni liścia, wartości niskie i wysokie są oceniane z większą dokładnością niż wartości średnie. Fakt ten został wzięty pod uwagę przy opracowywaniu różnych pomocy stosowanych przy ocenie oraz skal (patrz poniżej). Fakt ten może również spowodować konieczność statystycznego przekształcenia ocenianych danych.

2.2.3 Ustalenie rankingu

Podczas ustalania rankingu każdemu pojedynczemu egzemplarzowi przyporządkowywana jest pozycja względem innych pojedynczych egzemplarzy. Wynikiem ustalania rankingu jest jakościowa zmienna porządkowa. Przy stosunkowo niewielkiej liczbie porównywanych próbek, mogą one zostać poddane ustalaniu rankingu dla określonej zmiennej w polu. W przypadku braku alternatywy dla obserwacji prowadzonych metodą ustalania rankingu, użyteczną procedurą statystyczną może być zastosowanie metod nieparametrycznych, takich jak analiza wariancji rankingu. Jednakże, skuteczność takiej metody nieparametrycznej jest zwykle mniejsza w porównaniu do procedur parametrycznych. Dlatego też ustalanie rankingu nie jest idealnym podejściem i należy go unikać w szczególności, gdy liczba powtórzeń jest niewielka.

2.2.4 Scoring

Scoring to metoda polegająca na przypisaniu badanego obiektu do jednoznacznie zdefiniowanych klas. Zestaw takich klas jest powszechnie zwany skalą, zwłaszcza kiedy, jak to się zwykle dzieje, badana zmienna jest porządkowa. Scoring jest także stosowany w przypadku zmiennych binarnych i nominalnych. Scoring z definicji jest metodą subiektywną. Może być stosowana dla badania różnego rodzaju obiektów: ordynowanych lub nie, ciągłych lub dyskretnych.

Scoring jest metodą odpowiednią do badania zmiennych jakościowych oraz zmiennych ilościowych mierzalnych z dużą dozą dokładności jedynie przy dużych nakładach. Główną zaletą tej metody jest jej szybkość i nieinwazyjność oraz fakt, iż za pomocą tej metody można opisać całe poletko za pomocą jednej wartości. Liczba kroków skali oznacza czułość metody. Nie powinna być ona zbyt niska, ponieważ uzyskane w taki sposób informacje byłyby mało użyteczne, lub zbyt wysoka, gdyż wtedy skala staje się niepraktyczna w stosowaniu.

Skale są adaptowane do konkretnych celów i nie istnieje, ogólnie rzecz biorąc, uniwersalna skala dla jednego rodzaju zmiennej. Normy EPPO zawierają wiele zalecanych przykładów skali (załącznik 1) przy dokonywaniu oceny poszczególnych kombinacji uprawa/agrofag. Ogólnie rzecz biorąc, wobec omawianych skal zastosowano pewne proste zasady, zwłaszcza przy określaniu wartości skrajnych. Najniższym punktem skali porządkowej (brak rezultatów) powinna być liczba 1 (nie 0 – co jest zarezerwowane w wielu systemach rejestracji dla obserwacji nie wykonywanych), zaś najwyższa wartość na skali powinna odpowiadać najwyższej wartości oddziaływania, z uwzględnieniem odpowiedniej kolejności kroków pośrednich. Tabela 2 przedstawia podsumowanie różnych sposobów obserwacji oraz różnych rodzajów otrzymywanych zmiennych.

2.3 Zastosowanie skali w ocenie wzrokowej i scoringu

Ocena wzrokowa i scoring to metody oceny, które są często ze sobą mylone. Na pierwszy rzut oka działania są podobne, jednakże ich wyniki są odmienne: ocena wzrokowa prowadzi do uzyskania serii ocenianych wartości ciągłej lub nieciągłej zmiennej ilościowej, zaś scoring liczby podawane są w klasach. Klasy skali scoringowej są często wyrażane w liczbach jednej sekwencji (np. 1-9), jednakże nie oznacza to, że odstęp między wartościami skali są takie same. Gdy odstęp różni się między sobą, nie zaleca się przeprowadzania analizy statystycznej bez specjalistycznej konsultacji lub dokonywania oceny parametrów statystycznych bez odpowiedniego przygotowania. Wszelkie statystyki wyprowadzane na podstawie takich obliczeń powinny być interpretowane z dużą ostrożnością. Wartości skali mogą być także przedstawiane w formie liter alfabetu, co jednocześnie kładzie nacisk na ich charakter zmiennej porządkowej oraz na niebezpieczeństwa związane ze zbyt uproszczonym podejściem.

Skale mogą, jednakże, być stosowane również jako pomoc przy ocenie wzrokowej („zmienna porządkowa z odstępami”). Jeżeli wartości na skali są rzeczywistymi wartościami zmiennej ilościowej (tak jak ma to miejsce w przypadku klucza wzrokowego % zainfekowanej powierzchni liścia), wówczas obserwator przydziela wartości skali lub dokonuje interpolacji wartości pośrednich wedle własnej oceny. Otrzymane wartości, będące w razie możliwości poddane odpowiedniemu przekształceniu, stanowią oceny zmiennych ciągłych i mogą być zgodnie z tym analizowane przy zastosowaniu zwykłych procedur statystycznych. Należy pamiętać, że w przypadku, gdy obserwator dysponuje środkami (czas, siła robocza, doświadczenie) do przeprowadzenia jeszcze bardziej dokładnej oceny lub nawet pomiarów, otrzymane dane mogą zostać przeanalizowane z jeszcze większą dokładnością i skutecznością. Jednakże nie jest celowe dokonywanie stosunkowo dokładnych ocen (przykładowo, % zainfekowania powierzchni liścia), jeżeli następnie zastąpimy je o wiele mniej dokładnymi wartościami skali. Korzyści stosowania scoringu (szybkość i prostota) występują jedynie, gdy obserwator dokonuje klasyfikacji bezpośrednio do odpowiedniej klasy scoringowej (w którym to celu obserwatorzy są szkoleni) bez podejmowania prób przeprowadzenia dokładniejszej oceny.

2.4 Jakość sposobu obserwacji

Sposoby obserwacji rozróżniane są na podstawie pewnej ilości cech:

- „dokładność” – brak odchyżeń w kontekście statystycznym
- „niezawodność” – niska zmienność (lub wariancyjność)
- „precyzja” – kombinacja dokładności i niezawodności

„czułość” – reakcja sposobu obserwacji na niewielkie zmiany wartości w jednostce doświadczalnej

„powtarzalność” – identyczna (lub bardzo zbliżona) wartość przyporządkowana przez tego samego obserwatora do identycznych jednostek doświadczalnych

„niezmiennność” – identyczna (lub bardzo zbliżona) wartość przyporządkowana przez różnych obserwatorów do tej samej jednostki doświadczalnej

Powyższe ważne cechy decydują o wyborze sposobu obserwacji do poszczególnych celów, zwłaszcza w seriach badań.

3. Analiza statystyczna wyników badań

Decyzja o konieczności przeprowadzenia analizy statystycznej wyników badań lub serii badań zależy będzie od uzyskanych wyników oraz celu badania. Analiza statystyczna nie jest konieczna we wszystkich badaniach prowadzonych do celów rejestracji. Analiza statystyczna jest szczególnie wartościowa, na przykład, przy porównywaniu wpływu zabiegów przy stosowaniu różnych dawek, skuteczności różnych formułacji tego samego środka lub wpływu na zbiory w związku z innym zabiegiem.

3.1 Zasady

Zamieszczone poniżej zapisy stanowią informacje o zarysie dobrej statystycznej praktyki przeprowadzania analizy danych. Nie jest to, i nie może być, uniwersalna recepta dla wszystkich analiz i sytuacji.

Praktycy nie powinni nigdy bagatelizować potrzeby zasięgnięcia profesjonalnej porady statystycznej. Ważne jest, aby osoby dokonujące analizy dobrze rozumieli porady, które są im udzielane. Nierzadko lepiej jest, aby przeprowadzali oni proste analizy, o której będą w stanie napisać raport i ją obronić pewną argumentacją, niż aby przyjmowali porady, na podstawie których mieliby przeprowadzić analizy zrozumiałe przez nich jedynie częściowo. Pomocna może okazać się bibliografia zamieszczona na końcu niniejszych norm. Obejmuje ona kilka wartościowych pozycji, które mają na celu zaprezentować zasady dobrej statystycznej praktyki, zamiast zestawu statystycznych zasad, których należy ślepo przestrzegać.

3.2 Analiza statystyczna pojedynczego badania

3.2.1 Podstawowa budowa i sekwencja analizy

Normy EPPO dla oceny skuteczności środków ochrony roślin zawierają zapis stanowiący, że „Analiza statystyczna powinna być przeprowadzana przy zastosowaniu odpowiednich metod, które należy wskazać”. Procedura, wedle której należy postępować, może być zilustrowana na podstawie typowego badania, w którym kilka badanych środków stosowanych jest w pojedynczej dawce i poddawanych porównaniu ze produktem porównawczym, w obecności nie poddawanej zabiegom kontroli. Skuteczność preparatu oceniana jest poprzez mierzoną zmienną ilościową. Celem badania jest porównanie środków badanych ze produktem porównawczym, a w szczególności ustalenie, które z nich są najbardziej skuteczne. Sekwencja analizy, dla badania przeprowadzonego poprawnie zgodnie z odpowiednią normą EPPO, jest następująca:

Czy badanie jest realistyczne, tj. czy możliwe jest dzięki niemu uzyskanie użytecznych danych? Będzie to miało miejsce jedynie, gdy zakażenie agrofagami w nie poddawanej zabiegom kontroli jest wystarczająco wysokie i niezbyt zmienne.

Czy wyniki są spójne? Czy środek porównawczy pozwala uzyskać oczekiwane wyniki w porównaniu do nie poddawanej zabiegom kontroli?

Jeśli spełnione są powyższe dwa warunki, wówczas uzasadnione jest porównanie badanych środków ze produktem porównawczym oraz, w miarę możliwości, przeprowadzenie porównania samych środków. Analiza powinna mieć na celu głównie oszacowanie wielkości różnic lub stosunków pomiędzy produktem testowanym a produktem porównawczym i uzyskanie oceny zmienności tych ocen przy zastosowaniu standardowej statystyki błędu, przedziału ufności lub podobnej statystyki.

Opracować można podobne schematy dla innych badań oceniających skuteczność, w szczególności dla specjalnych przypadków selektywności herbicydów oraz dla przypadków wyjątkowych, w których nie występuje odpowiedni środek porównawczy, zaś zabiegi muszą być porównywane z nie poddawaną zabiegom kontrolą (patrz część 1.6.1).

Gdy badaniem objęte są dwa (lub więcej) produkty porównawcze (patrz przykład w części 1.2.2), wówczas sposób przeprowadzania analizy powinien zostać określony przed rozpoczęciem badania. Zalecane jest oddzielne porównanie każdego środka porównawczego do nowego preparatu bez jakichkolwiek dostosowań lub poprawek. Jeżeli porównanie badanego środka z połączonymi środkami porównawczymi zostanie uznane za prawidłowe, wówczas test homogeniczności pomiędzy środkami porównawczymi może zostać przeprowadzony jako pierwszy.

3.2.2 Wybór metody analizy

Zasadniczo, rodzaj zmiennej determinuje metodę analizy. Jeżeli zmienna jest ilościowa (binarna, dwumianowa, dyskretna lub ciągła), należy zastosować parametryczną metodę statystyczną, opartą zwykle na ogólnym modelu liniowym (GLM), np. analiza wariancji, regresja liniowa, regresja logistyczna. Jeżeli zmienna jest jakościowa, wówczas odpowiednie są metody nieparametryczne.

Przy przeprowadzaniu analizy wariancji przyjmowane są trzy założenia: addytywność skutków oddziaływania, homogeniczność wariancji i normalność błędu. Zastosowanie metod nieparametrycznych jest zalecane wtedy, gdy nie zostaną spełnione powyższe założenia. Jednakże, brak addytywności i brak normalności mogą często być poprawione i nie stanowią wystarczającego powodu do analizowania danych przy zastosowaniu metod nieparametrycznych, które są ogólnie nieskuteczne.

3.2.3 Brak addytywności skutków oddziaływania

Istotną rzeczą jest rozważenie, czy skutki oddziaływania będą addytywne w skali, w której proponuje się przeprowadzenie analizy zmiennej odpowiedzi. Dla przykładu, jeśli zmienną odpowiedzi stanowi gęstość populacji owadów, wówczas prawdopodobne jest, że wyniki zabiegów, takich jak zastosowanie środka owadobójczego lub zagrzybienie będą multiplikatywne, oddziałując na część populacji. Ewentualnie, jeśli zmienną odpowiedzi jest proporcja chwastów zneutralizowanych dzięki zabiegowi z użyciem herbicydu, prawdopodobne jest, że skutki będą addytywne nie w skali naturalnej, lecz w skali probitowej lub logitowej. Powszechnie stosowane są dwie metody w celu poprawienia skali naturalnej, aby przyjęła ona bardziej realistyczny kształt:

transformacje oraz ogólne modele liniowe. Ogólne modele liniowe są formą regresji, która generalizuje analizę wariancji przy przeprowadzaniu projektowanych doświadczeń. Są one ulepszoną wersją transformacji, jako że odnoszą się one do problemu addytywności skutków oddziaływania i równości wariancji (rozkład nie-normalny) w sposób oddzielny i jednoczesny.

Umożliwiają one rozkład zmiennej odpowiedzi w celu jej bezpośredniego określenia. Dla przykładu, przy obliczeniach liczby owadów, model taki może określać logarytmiczną „funkcję łącza” (w celu odniesienia się do skutków multiplikatywnych) oraz rozkład Poissona (w celu bezpośredniego odniesienia się do problemu równości wariancji i rozkładu nie-normalnego). W innym przypadku, zmienna dwumianowa może zostać poddana analizie poprzez zastosowanie logitowej funkcji łącza (aby uzyskać addytywność) oraz poprzez określenie rozkładu dwumianowego (aby bezpośrednio dopasować dane, które mogą przybierać formę r zainfekowanych roślin spośród n roślin poddanych zabiegowi). Istnieje wiele podobieństw pomiędzy analizą dewiancji wynikającej z zastosowania ogólnego modelu liniowego, a tradycyjną analizą wariancji. W szczególności pojęcia sum kwadratów, stopni swobody, kontrastów ortogonalnych, chi-kwadrat i testów F oraz przewidywane średnie ze standardowymi błędami, wszystkie one mają swoje odpowiedniki w uogólnionych modelach liniowych i mogą być wykorzystywane w badaniach.

3.2.4 Homogeniczność wariancji

Mimo, iż transformacje stanowią rozwiązanie problemu addytywności skutków, nie zapewniają one jednak homogeniczności wariancji. Cechę tą należy sprawdzić niezależnie, mimo iż addytywność jest zwykle cechą ważniejszą. W przypadku zliczeń, transformacja logarytmiczna zapewnia zwykle zarówno addytywność, jak i równość wariancji. W przypadku danych binarnych, dwumianowych oraz danych w formie proporcji, transformacja dwumianowa, logitowa, probitowa lub logarytmiczna będzie zwykle wystarczająca w celu uzyskania addytywności, mimo iż równość wariancji może nie zostać uzyskana.

3.2.5 Normalność oraz niezależność błędów

Rozkład błędów powinien być normalny. W celu sprawdzenia tego rozkładu dostępne są testy normalne lub zobrazowania graficzne. W praktyce analiza wariancji jest często na tyle skuteczna, że powoduje odejście od normalności. W miarę możliwości należy potwierdzać, że błędy są niezależne od czynników zabiegowych.

3.3 Analiza wariancji

3.3.1 Tablice średnich

Po transformacji zalecane jest przeprowadzenie analizy wariancji, obojętnie, czy przyjmie ona formę modelu liniowego, czy uogólnionego modelu liniowego lub transformacji. Należy zaprezentować tabelę średniej każdego z zabiegów, wraz z oceną zmienności średnich, zwykle w formie błędów standardowych lub przedziału poufności. Tabela taka kładzie nacisk na wielkość oddziaływania i jej stosowanie zalecane jest dla pokonania dobrze znanego problemu polegającego na tym, że znaczenie biologiczne nie może być równe znaczeniu statystycznemu, a skutki oddziaływania mogą być znaczne tak pod względem wielkości, jak i ważności jednak nieznaczące z powodu niedostatecznej skuteczności analizy bądź testu. Analiza może także wykorzystywać uogólniony model liniowy, przy zastosowaniu którego analiza wariancji jest specjalnym przypadkiem, lub inną odpowiednią metodę.

Należy zachować staranność przy przydzielaniu jednostek doświadczalnych do poszczególnych warstw w analizie tabeli wariancji, z zabiegiem i strukturą blokującą odpowiednią do przyjętego projektu. W szczególności należy dochować wszelkich starań, aby uniknąć dobrze znanego problemu pseudo-replikacji, występującego z powodu nieuwzględnienia faktu, że zabiegi nie zostały w pełni zrandomizowane na jednostkach próbek, lecz na grupach takich jednostek.

3.3.2 Testy F oraz kontrast ortogonalny

Poza prezentacją tabel średnich i błędów standardowych oraz przeprowadzeniem formalnych testów statystycznych, dla całości danych można również przeprowadzić testy F .

Ogólny test wszystkich zabiegów nie powinien być reprezentowany jako dowód skuteczności, z wyjątkiem najprostszego z przypadków, ponieważ, generalnie rzecz biorąc, będzie on narażony na interferencje informacji z nie poddawanej zabiegowi kontroli. Zamiast tego, zaleca się, aby sumy kwadratów zabiegów zostały podzielone na elementy biologiczne poprzez zdefiniowanie (niezależnych) kontrastów ortogonalnych.

Przykładowo, w pierwszym przykładzie, w którym porównano osiem zabiegów, występowało pięć różnych preparatów badanych, dwa produkty porównawcze oraz nie poddawana zabiegowi kontrola. Osiem zabiegów przyniosło 7 df w sumie kwadratów zabiegów. Kontrastami mogą być: nie poddawana zabiegowi kontrola i średnia pozostałych siedmiu zabiegów (1 df), preparat porównawczy jeden i preparat porównawczy dwa (1 df), średnia preparatów porównawczych i średnia preparatów badanych (1 df), różnice pomiędzy średnimi samych preparatów badanych (4 df). Pierwsze dwa z powyższych kontrastów służą do likwidowania zmienności uciążliwości o względnie niewielkiej ważności biologicznej, natomiast kontrastami służącymi do ujawniania prawdziwych celów badania są dwa ostatnie kontrasty. Każdy kontrast zapewnia osobną statystykę F , która może być wykorzystywana w celu formalnego przetestowania hipotez będących przedmiotem zainteresowania. W rozpatrywanym przykładzie, hipotezami będącymi przedmiotem zainteresowania mogłyby być, średnio rzecz biorąc, założenie iż preparaty badane nie są lepsze od produktów porównawczych oraz że same produkty badane nie różnią się między sobą. Interpretacja dwóch pierwszych hipotez może być uzależniona od tego, czy kontrast pomiędzy samymi produktami porównawczymi ujawnił znaczącą różnicę. W przypadku testowania kontrastów nieortogonalnych, przykładowo oddzielnych pięciu kontrastów na 1 df pomiędzy średnią każdego badanego preparatu a określonym produktem porównawczym, testowanie powinno również być przeprowadzone za pomocą testu F (lub w razie potrzeby testu t -) przy zastosowaniu resztkowego średniego kwadratu z analizy wariancji.

Kontrasty i hipotezy będące przedmiotem zainteresowania powinny być, jeśli to możliwe, określone z wyprzedzeniem, na etapie projektu i wykorzystywane z umiarem. Testy nie powinny być wykonywane tylko dlatego, że wstępna analiza post-hoc wykazała różnice, które wydają się znaczące i mogące mieć znaczenie w przypadku testowania. Konsystencja jest z reguły lepszą wskazówką obecności rzeczywistego oddziaływania aniżeli stosowane testy znaczenia, szczególnie gdy skuteczność jest niska. Przykładowo, gdyby badany preparat okazał się skuteczniejszy od preparatu porównawczego w każdym z jedenastu odległych miejsc, jednak nie odpowiednio znaczący w każdym z nich, zdrowy rozsądek podpowiadałby, że spójność wyników jest czynnikiem ważnym (rzeczywiście możliwe jest zastosowanie dwuczłonowego testu dwumianowego celem udowodnienia, że prawdopodobieństwo uzyskania tak dużego wyniku jak ten, przypadku braku istotnych różnic pomiędzy zabiegami, jest mniejsze niż 0,001).

3.3.3 Procedury testów wielokrotnych

Do celów rejestracji, nie wszystkie pary porównawcze są odpowiednie i nie wszystkie kontrasty ortogonalne mogą być uwzględnione we wnioskach rejestracyjnych. Ze wszystkich możliwych ($k(k-1)/2$) par porównawczych, jedynie kilka jest odpowiednich do wykazania skuteczności badanego preparatu. Przykładowo, rozważmy badanie, w którym porównywane jest 7 zabiegów przy 5 różnych badanych preparatach, jednej nie poddawanej zabiegowi kontroli i jednym produkcie porównawczym. Zgodnie z zasadą opisaną w punkcie 3.2.1, należy przeprowadzić kilka odpowiednich testów. Po pierwsze, odpowiedniość badania powinna być wykazana poprzez zbadanie poziomu infekcji w nie poddawanej zabiegowi kontroli względem uprzednio określonego

poziomu zainfekowania. Po drugie, należy przetestować różnice pomiędzy produktem porównawczym a nie poddaną zabiegowi kontrolą, w celu wykazania spójności badania. Jeżeli zostanie to osiągnięte, wówczas trzecią w kolejności procedurą jest porównanie każdego testu z preparatem porównawczym, w celu wykazania przynajmniej równości oddziaływania w odniesieniu do preparatu porównawczego. Aby przeprowadzić to ostatnie badanie, należy skonsultować się z dostępną literaturą na temat wielu istniejących procedur parametrycznych i nieparametrycznych (Hothorn i Bleiholder, 2006).

W doświadczeniu czynnikowym (np. test wielokrotnego dawkowania), przeprowadzanie wszystkich porównań opierających się na parach porównawczych nie jest zwykle pomocne we kombinacjach czynnikowych (Perry, 1986). Zamiast tego bardziej odpowiednie jest przeprowadzenie analizy danych zgodnie ze strukturą badań. W zależności od wyników dwustronnej analizy wariancji, zwykle najbardziej odpowiednie jest porównywanie średnich marginalnych lub prostych z oddzielnymi poziomami innego czynnika i na odwrót.

Standardowe procedury wielokrotnego porównywania, opisane w Tukey (1953) lub szeroko stosowany test Duncana (Duncan, 1955) lub test Newman-Keulsa (Keuls, 1952), zakładają przeprowadzenie wszystkich porównań opartych na parach porównawczych, które są w naturalny sposób dwustronne. Możliwe są o wiele mniej zachowawcze procedury z użyciem odpowiednich porównań, gdy opracowane zostaną jako testy jednostronne. Testy jednostronne oraz przedziały ufności są odpowiednie pod względem biologicznym, jako że, przykładowo, przedmiotem zainteresowania jest zwykle zmniejszenie zainfekowania, nie zaś jego zwiększenie. Stosowane na szeroką skalę test wielokrotnego rozstępu Duncana oraz test wielokrotnego rozstępu Newman-Keulsa nie kontrolują globalnego zasięgu testu (poziom α), kontrolując jedynie zasięg lokalny (poziom α). W związku z tym, jeśli test jest oparty na uprzednio określonym poziomie α , wynoszącym 0,05, będzie to zgodne z prawdą jedynie przy porównaniu dwóch średnich z zabiegów, przy czym wraz ze zwiększającą się liczbą średnich porównywanych jednocześnie, poziom α wzrasta wykładniczo. Przy stosowaniu procedur testów wielokrotnego porównania, zaleca się wybieranie tylko tych procedur, o których wiadomo, że kontrolują lokalny i globalny poziom α jednocześnie.

Jako że badania polowe konieczne do rejestracji, mające na celu wykazanie skuteczności nowych badanych preparatów będą prowadzone na ostatnim etapie prac nad preparatem, oczekiwany kierunek każdej różnicy powinien być jasno widoczny z kontekstu. W związku z tym, testy jednostronne oraz jednostronne przedziały ufności są zalecane do stosowania celem uzyskania pewnego poziomu skuteczności przy normalnej liczbie replikacji stosowanej zwykle w badaniach polowych. Jednakże nie wyklucza to zastosowania innych wyżej wymienionych testów statystycznych.

3.3.4 Modele skutków losowych

Niniejsza norma skupia się na uznawaniu zabiegów jako skutki stałe. Niektórzy praktycy mogą wykazywać chęć uznawania skutków zabiegów, w niektórych doświadczeniach, szczególnie w próbach jednorodności, za losowo wybraną próbkę z większej, bliżej nieustalonej populacji. Praktyka taka zwana jest losowym modelowaniem skutków. Badania mogą obejmować również skutki trwałe i losowe, tak zwane modele mieszane. W przypadku modeli tego typu, zaleca się stosowanie nowoczesnej techniki REML (Ocena pozostałości za pomocą największego prawdopodobieństwa). Technika REML może być również wykorzystywana do przeprowadzania porównań pomiędzy kilkoma laboratoriami lub miejscami przeprowadzania badań, w celu dokonania oceny składników wariancji lub kiedy projekt nie może być zanalizowany za pomocą analizy wariancji z powodu zbyt wielu brakujących wartości powodujących niezrównoważenie.

Także i w tym przypadku, istnieje wiele podobieństw pomiędzy pojęciami i ilościami REML a pojęciami i ilościami analizy wariancji. Nie należy, jednakże, rezygnować z porady statystycznej.

3.3.5 Dane porządkowe

Nowoczesne metody analizy uporządkowanych danych kategoryjnych zostały opisane przez Agresti'ego (1984) oraz Brunnera i Munzela (2002), mimo to, w celu ich prawidłowego stosowania konieczne może okazać się zasięgnięcie specjalistycznej porady statystycznej. Dodatkowo, w niektórych przypadkach koniecznym okazać się może traktowanie zmiennych będących liczbami całkowitymi jako zmienne porządkowe, jeżeli ich zakres wariancji nie jest zbyt duży, aby uznawać je za ciągłe, a badanie jest mimo to uznawane za ważne.

3.3.6 Dane jakościowe i metody nieparametryczne

W przypadku danych prawdziwie jakościowych, przykładowo danych nominalnych oraz niektórych danych klasyfikowanych lub w przypadku danych, które nie mają dobrze znanego rozkładu parametrycznego, takich jak dane nominalne, dwumianowe, dane beta, gamma albo rozkładu Poissona, zastosowanie metod nieparametrycznych może być użyteczną procedurą statystyczną dla przeprowadzenia analizy danych. W porównaniu z metodami parametrycznymi, skuteczność metod nieparametrycznych jest mniejsza, w związku z tym należy je stosować ze szczególną ostrożnością, w przypadkach, gdy liczba replikacji jest bardzo niewielka. Jednakże ilość informacji, którą analiza taka może przekazać jest wystarczająco duża do celów niniejszych wytycznych, aby uzyskać pożyteczne wyniki przy badaniu skuteczności produktu. Spośród opisów tradycyjnych testów najlepszy nadal wydają się opisy zawarte w testach Siegel'a (1956) oraz Brunnera i Munzela (2002), które wyjaśniają wyraźnie, które testy są odpowiednie dla którego zestawu danych. Bardziej nowoczesne podejścia obejmują techniki komputerowe, takie jak testy randomizacyjne. Metody randomizacyjne mogą być bardzo użyteczne w przypadkach, w których nie można ufać sposobom parametrycznym, przykładowo, jeżeli dane są bardzo „nie normalne”, lub kiedy w danych występuje dużo zer (jeżeli mimo to badanie jest uznawane za ważne). Stosowanie innych komputerowych metod nieparametrycznych zalecane jest w celu ulepszenia oceny lub w celu dokładniejszego obliczenia zmienności oceny. Metody te obejmują „ładowanie początkowe” oraz „jackknifing”, jednakże i w tym przypadku konieczna może okazać się specjalistyczna porada.

3.4 Analiza statystyczna serii prób

Spójność oddziaływania zabiegów, np. porównania nowego preparatu względem preparatu porównawczego, dla różnych środowisk (regionów, miejsc przeprowadzania badania), jest koniecznym i ważnym kryterium w kwestii rejestracji. W związku z tym od badań pojedynczych preferowane jest przeprowadzanie serii badań.

3.4.1 Definicja

Do celów niniejszej normy, seria prób może być zdefiniowana jako zestaw zabiegów testowanych w różnych warunkach środowiskowych w jednym roku lub wielu latach. Zestaw zabiegów należących do serii prób powinien zostać poddany analizie przy wykorzystaniu tego samego modelu statystycznego.

3.4.2 Planowanie

Planując serię prób osoby prowadzące doświadczenie powinny rozważyć zdefiniowanie kluczowej kwestii próby oraz wszelkie konieczne parametry, tj. wykaz głównych zabiegów, projekt badania i replikacji, liczbę miejsc prowadzenia badania, metody testowania itp., które są wymagane przy zastosowaniu planowanego modelu biometrycznego dla przeprowadzenia analizy serii prób.

3.4.3 Cele

Cele analizy są następujące:

- ocena oddziaływania zabiegów w miejscach prowadzenia badania z upływem lat
- przetestowanie interakcji pomiędzy zabiegami, miejscami prowadzenia badania i interakcji zachodzących z upływem lat

Różnice środowiskowe i inne pomiędzy miejscami prowadzenia badania oraz zachodzące z upływem lat mogą zmienić powyższe czynniki

- w miarę możliwości, przetestowanie znaczenia różnic pomiędzy zabiegami a normami.

3.4.4 Podstawowa struktura i porządek analizy

Przed rozpoczęciem prowadzenia analizy statystycznej wyników serii prób należy zatwierdzić dane z każdego badania. Zatwierdzenie to ma zastosowanie do trzech kwestii:

- *zatwierdzenie metodologiczne*: przeprowadzenie wszystkich prób musi być zgodne z protokołem pierwotnym
- *zatwierdzenie agronomiczne i biologiczne*: na badania nie powinny mieć wpływu czynniki zewnętrzne lub szczególne. Powinny być one reprezentatywne dla regionu i roku prowadzenia. Produkty porównawcze we wszystkich próbach powinny wykazywać normalność. Presja zainfekowania powinna być odpowiednia (znacznym poziom dla badań oceniających skuteczność, niski poziom dla badań selektywności)
- *zatwierdzenie statystyczne*: próby powinny być dokładne i wykazywać typowy błąd standardowy (lub współczynnik wariancji).

Analiza serii prób ma być ukierunkowana na skuteczność oraz na interakcję pomiędzy zabiegami a środowiskiem. Celem analizy interakcji jest wykazanie braku znaczących interakcji we wszystkich lub prawie wszystkich środowiskach. Nie można wykazać tego w odpowiedni sposób jedynie poprzez obecność nieznaczącego, globalnego testu F w kontekście interakcji. Zamiast tego, w celu wykazania podobieństwa oddziaływania zabiegów we wszystkich, lub przynajmniej w większości środowisk, bardziej odpowiednim jest przeprowadzenie badań wszystkich części składowych interakcji poprzez zastosowanie kontrastów. W działaniu tym należy wykluczyć interakcje jakościowe; są one tolerowane jedynie w akceptowalnej ilości występowania w praktyce. Miejsca prowadzenia badań, które nie wykazują żadnych interakcji pomiędzy zabiegami a środowiskiem mogą następnie stanowić pulę do analizy. Miejsca prowadzenia badań, które wykazują niedopuszczalnie wysoki poziom interakcji muszą zostać poddane oddzielnej analizie i dyskusji.

3.4.5 Wybór metody statystycznej

Przy badaniach pojedynczych, metody statystyczne są determinowane przez rodzaj zmiennej, która ma zostać poddana analizie. Metody, które należy zastosować są identyczne lub podobne do tych stosowanych przy pojedynczym badaniu (np. analiza wariancji, metody nieparametryczne). Głównym celem analizy serii prób jest dokonanie pomiaru i zbadanie interakcji pomiędzy preparatami testowanymi a środowiskiem lub miejscem prowadzenia badania, tj. wykazanie, że różnice pomiędzy produktami są „równe” w każdym miejscu prowadzenia badania. Badania

mogą zostać pogrupowane przed przeprowadzeniem analizy, zgodnie z odpowiednimi kryteriami (np. typ gleby, poziom inwazji) lub po jej przeprowadzeniu, przy zastosowaniu metod analitycznych i wyników badań interakcji w celu odpowiedniego pogrupowania badań.

Podziękowania

EPPO pragnie podziękować Dr H. Bleiholder i Prof. L.A. Hothorn za szczegółowe zalecenia odnośnie do korekty niniejszej normy.

Bibliografia

- Agresti A (1984) *Analysis of Ordinal Categorical Data*. Wiley, New York (US).
- Bauer P, Röhmel J, Maurer W & Hothorn LA (1998) Testing strategies in multi-dose experiments including active control. *Statistics in Medicine* **17**, 2133–2146.
- BBA (1980) Richtlinie für Planung, Durchführung und Auswertung von Versuchen mit Pflanzenbehandlungsmitteln: 1. Versuchsplanung; 2. Versuchsdurchführung. Biologische Bundesanstalt für Land- und Forstwirtschaft, Braunschweig (DE).
- BBA (1982) Richtlinie für Planung, Durchführung und Auswertung von Versuchen mit Pflanzenbehandlungsmitteln: 3. Auswertung des Einzelversuches; 4. Sachregister, Tabellen. Biologische Bundesanstalt für Land- und Forstwirtschaft, Braunschweig (DE).
- Brunner E & Munzel U (2002) *Nichtparametrische Datenanalyse*. Springer, Berlin (DE).
- CEB (1981) Rôle et implantation des témoins sans traitement dans les essais de produits phytosanitaires. ANPP-DT4. J Arnoux.
- CEB (1983) Principes d'appréciation des effets des produits phytosanitaires dans les essais de plein-champ. ANPP-DT5. Y Ribrioux.
- CEB (1986) Utilisation des tests statistiques dans l'interprétation des essais de produits phytosanitaires. ANPP-DT6. J Arnoux – JP Gouet.
- CEB (1990) Les réseaux d'essais. ANPP-DT9. JP Gouet.
- CEB (1990) Les unités expérimentales. ANPP-DT10. JP Gouet.
- Cochran WG & Cox GM (1957) *Experimental Design*, 2nd edn. Wiley, New York (US).
- Cox DR (1958) *Planning of Experiments*. Wiley, New York (US).
- Crawley MJ (1993) *GLIM for Ecologists*. Blackwell Scientific, Oxford (GB).
- Crowder MJ & Hand DJ (1990) *Analysis of Repeated Measures*. Chapman & Hall, London (GB).
- Cullis BR & Gleeson AC (1991) Spatial analysis of field experiments – an extension to two dimensions. *Biometrics* **47**, 1449–1460.
- Dagnelie P (1969) *Théorie et Méthodes Statistiques*, 2. Duculot, Gembloux (BE).
- Denis JB (1980) Analyse de régression factorielle. *Biométrie-Praximétrie* **19**, 15–34.
- Denis JB, Gouet JP & Tranchefort J (1980) Méthodes d'étude de la structure de l'interaction génotype *milieu et de recherche d'un modèle explicatif à effets fixes: application à l'analyse des résultats d'un réseau d'essais de variété de blé tendre. *Biométrie et Génétique*, pp. 98–109. Société Française de Biométrie, Paris (FR).
- Denis JB & Vincourt P (1982) Panorama des méthodes statistiques pour l'étude de l'interaction génotype milieu. *Agronomie* **2**, 219–230.
- Dobson AJ (2002) *An Introduction to Generalized Linear Models*, 2nd edn. Chapman & Hall, CRC/ Boca Raton (US).
- Duncan DB (1955) Multiple range and multiple F tests. *Biometrics* **11**, 1–42.
- Dyke GV (1988) *Comparative Experiments with Field Crops*. Griffin, London (GB).
- Finney DJ (1971) *Probit Analysis*, 3rd edn. Cambridge University Press, Cambridge (GB).

- Finney DJ (1978) *Statistical Method in Biology Assay*, 3rd edn. Griffin, London (GB).
- Finney DJ (1980) *Statistics for Biologists*. Chapman & Hall, London (GB).
- Gouet JP (1974) *Les Comparaisons de Moyennes et de Variances. Application à l'Agronomie*. ITCF, Paris (FR).
- Gouet JP & Philippeau G (1992) *Comment Interpréter les Résultats d'une Analyse de Variance?* ITCF, Paris (FR).
- Hollander M & Wolfe DA (1973) *Non-parametric Statistical Methods*. Wiley, London (GB).
- Horn M & Vollandt R (1995) *Multiple Tests und Auswahlverfahren*. Gustav. Fischer Verlag, Stuttgart (DE).
- Hothorn LA & Bleiholder H (2006) Statistical aspects of efficacy evaluation of plant protection products in field trials – a comment to the EPPO PP1/152 (2) guideline. *Biuletyn OEPP/EPPO* **31**, 143-152.
- Hughes G & Madden LV (1992) Aggregation and incidence of disease. *Plant Pathology* **41**, 657–660.
- Hurlbert SH (1984) Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54**, 187–211.
- Keuls M (1952) The use of studentized range in connection with an analysis of variance. *Euphytica* **1**, 112–122.
- Little TM (1978) If Galileo published in HortScience. *Hortscience* **13**, 504–506.
- McCullagh P & Nelder JA (1983) *Generalized Linear Models*. Chapman & Hall, London (GB).
- Mead R (1988) *The Design of Experiments: Statistical Principles for Practical Applications*. Cambridge University Press, Cambridge (GB).
- Mead R & Curnow RN (1983) *Statistical Methods in Agriculture and Experimental Biology*. Chapman & Hall, London (GB).
- Nelder JA (1971) Contribution to the discussion of the paper by O'Neill and Wetherill. *Journal of the Royal Statistical Society Series B* **36**, 218–250.
- Parker SR, Whelan MJ & Royle DJ (1995) Reliable measurement of disease severity. *Aspects of Applied Biology* **43**, *Field experiment techniques*, pp. 205–214.
- Patterson HD & Williams ER (1976) A new class of resolvable incomplete block designs. *Biometrika* **63**, 83–92.
- Pearce SC, Clarke GM, Dyke GV & Kempson RE (1988) *Manual of Crop Experimentation*. Griffin, London (GB).
- Perry JN (1986) Multiple-comparison procedures: a dissenting view. *Journal of Economic Entomology* **79**, 1149–1155.
- Perry JN (1989) Review: population variation in entomology: 1935–50. I. Sampling. *Entomologist* **108**, 184–198.
- Perry JN (1994) Sampling and applied statistics for pests and diseases. *Aspects of Applied Biology* **37**, 1–14.
- Perry JN (1997) Statistical aspects of field experiments. W: *Methods in Ecological and Agricultural Entomology* (Ed. Dent, DR & Walton, MP), pp. 171–201. CAB International, Wallingford (GB).
- Plackett RL (1981) *The Analysis of Categorical Data*, 2nd edn. Griffin, London (GB).
- Preece DA (1982) The design and analysis of experiments: what has gone wrong? *Utilitas Mathematica* **21A**, 201–244.
- Rasch D, Herrendörfer G, Bock J, Victor N & Guiard V (1996) *Verfahrensbibliothek, Versuchsplanung und -Auswertung*. Band I. R. Oldenbourg Verlag, München (DE).

Rasch D, Herrendörfer G, Bock J, Victor N & Guiard V (1998) *Verfahrensbibliothek, Versuchsplanung und -Auswertung*. Band II. R. Oldenbourg Verlag, München (DE).

Siegel S (1956) *Non-Parametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York (US).

Tukey JW (1953) *The Problem of Multiple Comparisons*. Mimeographed monograph. Princeton University, Princeton NY (US).

Załącznik 1

Przykłady skali stosowanych w Normach EPPO odnośnie do oceny skuteczności stosowania środków ochrony roślin

(1) Nominalne

Odbarwienie liści ziemniaka (Norma EPPO PP 1/135 *Ocena fitotoksyczności*).

chloroza

żółte żyłki

żółte plamki

ogólne ciemno lub jasno zielone ubarwienie liścia

albinizm.

(2) Porządkowe bez ilościowo zdefiniowanych przedziałów

Ocena korzeni kapusty dla *Plasmodiophora brassicae* (Norma EPPO PP 1/39 *Skuteczność stosowania środków grzybobójczych wobec Plasmodiophora brassicae*):

(1) brak widocznego spęcznienia

(2) bardzo lekkie spęcznienie, ograniczone zwykle do korzeni poprzecznych

(3) spęcznienie umiarkowane na korzeniach poprzecznych i/lub palowych

(4) znaczące spęcznienie na korzeniach poprzecznych i/lub palowych.

Ocena roślin sałaty pod kątem zainfekowania *Botryotinia fuckeliana* (Norma EPPO PP 1/54 *Badanie skuteczności stosowania środków grzybobójczych wobec Botrytis spp. oraz Sclerotinia spp. na warzywach*)

(1) brak zaatakowania

(2) lekki stopień zaatakowania, infekcja tylko na podstawowych ogonkach liściowych

(3) umiarkowany stopień zaatakowania, patologiczne zmiany łodygi, brak obrączkowania łodygi

(4) znaczny stopień zaatakowania, zmiany patologiczne obrączkowania łodygi lub zainfekowanie górnych liści, sałata nienadająca się do sprzedaży (włącznie z występowaniem w czasie badania roślin całkowicie zniszczonych przez *B. fuckeliana*).

(3) Skale porządkowe ze zdefiniowanymi przedziałami w oparciu o liczby

Zaatakowanie owoców jabłek przez *Venturia inaequalis* (Norma EPPO PP 1/5 *Badanie skuteczności stosowania środków grzybobójczych wobec Venturia inaequalis oraz V. pirina*):

(1) brak zaatakowania;

(2) 1-3 plamy na owoc;

(3) > 3 plamy na owoc.

Liczba zmian patologicznych na buraku cukrowym spowodowanych przez, np. *Scutigerella immaculata* (Norma EPPO PP 1/45 *Badanie oceniające skuteczność środków owadobójczych przeciwko kompleksowi agrofagów odglebowych wśród buraków*):

(1) brak zmian patologicznych;

(2) 1-2 zmiany patologiczne;

(3) 3-5 zmiany patologiczne;

(4) > 5 zmian patologicznych.

Niektóre ze skali są częściowo oparte na liczbie, a częściowo na obszarze, np.

- (1) liść zdrowy;
- (2) 1-2 plamy na liść;
- (3) więcej niż 2 plamy na liść;
- (4) więcej niż 1/3 obszaru liścia zainfekowanego.

(4) Skale porządkowe ze zdefiniowanymi przedziałami opartymi na zmiennych ciągłych

Ocena zainfekowania łodyg pszenicy przez *Tapesia yallundae* i *Tapesia acuformis* wywołujących chorobę podsuszkową zbóż (Norma EPPO PP 1/28 Ocena skuteczności stosowania środków grzybobójczych w zwalczaniu choroby poduszki zbóż):

- (1) brak symptomów
- (2) mniej niż 50% obwodu odrośli zaatakowanego w miejscu występowania najostrożniejszej infekcji
- (3) więcej niż 50% obwodu odrośli zaatakowanego w miejscu występowania najostrożniejszej infekcji, lecz tkanka nadal nienaruszona
- (4) 100% zaatakowanego obwodu odrośli, gnijąca tkanka (zmiękczenie).

Zwykle taka skala jest, przynajmniej częściowo, logarytmiczna.

Obszar liścia oliwki zainfekowany *Spilocoea oleagina* (Norma EPPO PP 1/81 Ocena skuteczności środków grzybobójczych w zwalczaniu *Spilocoea oleagina*).

- (1) brak symptomów
- (2) zainfekowane 0-10% obszaru liścia
- (3) 10–25%
- (4) 25–50%
- (5) 50–100%.

Mimo iż skale te są pozornie logarytmiczne, praktycznie nigdy nie ma stałego kroku z centralnej wartości każdej klasy do następnej. Dlatego też, mimo iż teoretycznie liniowe wyniki odpowiadające skali logarytmicznej mogłyby być analizowane jako zmienne ciągle odpowiadające prostej transformacji wartości porządkowej, przypadek ten praktycznie nie występuje, ponieważ skale nie są prawdziwie logarytmiczne. Dodatkowo, przyporządkowanie wartości 1 do klasy zero jest niejednorodne z resztą skali. Kolejną kwestią, na którą należy zwrócić uwagę jest fakt, że klasy definiowane są przedziałami zmiennych ciągłych. W przypadku liści oliwki zainfekowanych *S. oleagina* (patrz powyżej) obserwator ogląda liść i podejmuje decyzję odnośnie do tego, czy zaklasyfikować ją do klasy 3 czy 4. Obserwator nie ogląda liścia, podejmuje decyzję, iż został on zainfekowany w 35%, następnie klasyfikuje go do klasy 4. W przypadku podjęcia takiej decyzji przez obserwatora, równie dobrze mógłby on kontynuować ocenę bezpośrednio bez używania skali, ponieważ dałoby to więcej informacji, które mogłyby zostać w pełni przeanalizowane. Z tego powodu, prezentacja przedziałów jako 0-10, 11-25, 26-50, 51-100 jest niewłaściwa i niezrozumiała. Jeżeli w takim przypadku obserwator zauważy liść, który wydaje się mu zainfekowany w około 50%, będzie on musiał zdecydować, czy zaklasyfikować go do klasy 4 czy 5, nie próbując rozróżnić procentowo pomiędzy 50 a 51 (co jest oczywiście niemożliwe).

W kilku przypadkach kategorie opisowe pomieszane są ze zdefiniowanymi przedziałami.

Ocena liści jabłoni dla *Podosphaera leucotricha* (Norma EPPO PP 1/69 Ocena skuteczności środków grzybobójczych w zwalczaniu *Podosphaera leucotricha*):

- (1) brak pylistej pleśni
- (2) lekki stopień zainfekowania (rozsiane plamy pylistej pleśni)
- (3) umiarkowany do dużego stopnia zainfekowania (maksymalnie do połowy powierzchni liścia pokrytego pylistą pleśnią)

(4) bardzo duży stopień zainfekowania (ponad połowa powierzchni liścia pokryta pylistą pleśnią, krawędzie liścia zaczynają się wyginać i wysychać).

(5) Skale porządkowe z klasami zdefiniowanymi przez ich centralne wartości

Są to skale uważane za najlepsze jako pomoc w ocenie. Najczęściej spotykane są klucze wizualne (np. dla *Cercospora beticola*, *Peronospora hyoscyami*, itp.). Klucze służą zwykle do oceny procentu zakażonej powierzchni liścia i zostały dokładnie skalibrowane. Kroki dobierane są zwykle tak, aby dogodnie obejmowały zakres spodziewanego zainfekowania, np. 1, 5, 10, 25, 50 i umożliwiały właściwą interpolację, nie zaś w regularnej, prawie logarytmicznej sekwencji (która byłaby odpowiednia gdyby skala taka była stosowana przy scoringu).

W przypadku Normy EPP0 PP 1/2 Ocena skuteczności środków grzybobójczych w zwalczaniu *Phytophthora infestans* w stosunku do ziemniaków, punkty skali ilustrowane są opisowo, nie zaś za pomocą klucza wizualnego, gdzie opisy zostały dokładnie skalibrowane na odsetek zainfekowanej powierzchni liścia. Jednakże charakter skali jest identyczny.

(0) brak zainfekowania

(1) maksymalnie 10 plam na roślinę lub maksymalnie 1 listek na 10 zaatakowanych

(5) około 50 plam na roślinę lub maksymalnie 1 listek na 10 zaatakowanych

(10) maksymalnie 4 listki na 10 zaatakowanych, rośliny nadal zachowują normalną formę

(25) niemal na każdym listku występują zmiany patologiczne, jednak rośliny zachowują normalną formę; poletko może wyglądać na zielone, mimo że każda roślina w jego obrębie będzie zainfekowana

(50) wszystkie rośliny zainfekowane i około połowa powierzchni liścia jest zniszczona przez rdzę; poletko ma kolor zielony z brązowymi plamami.

Rys. 5 Możliwe ułożenie bloków i plotek w zrandomizowanych blokach w badaniach prowadzonych w warunkach polowych. Bloki są rozrzucone po całym polu, zgodnie z uprzednio zaobserwowaną heterogenicznością.

Rys. 6 Możliwe rozmieszczenie bloków i poletek w zrandomizowanych blokach w badaniach prowadzonych w warunkach polowych. Bloki są rozrzucone po całym polu, zgodnie z kompleksową, uprzednio zaobserwowaną heterogenicznością.

Rys. 7 Przykład układu z dzielonymi poletkami. Dwa czynniki zabiegu stanowią: preparat (1,2,3,4 zrandomizowane do podpoletek w obrębie poletek) oraz metoda prowadzenia uprawy (A,B,C zrandomizowane do całych poletek w obrębie każdego z dwóch bloków).

Blok 1					Blok 2			
1A	2A	3A	4A	Całe poletko 1	2B	4B	3B	1B
3C	4C	1C	2C	Całe poletko 2	2A	3A	4A	1A
2B	3B	1B	4B	Całe poletko 3	1C	3C	2C	4C

Tabela 1 Resztkowe stopnie swobody w stosunku do liczby miejsc prowadzenia badania, zabiegów oraz powtórzeń w miejscu prowadzenia badania

Miejsca prowadzenia badań	1					4						6					
---------------------------	---	--	--	--	--	---	--	--	--	--	--	---	--	--	--	--	--

Powtórzenia Zabiegi	3	4	5	6	7	8	3	4	5	6	7	8	3	4	5	6	7	8
3	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38
4	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57
5	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64	68	72	76
6	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95
7	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102	108	114
8	14	21	28	35	42	49	56	63	70	77	84	91	98	105	112	119	126	133

Rys. 8 Podobny układ zrandomizowanych bloków, jednakże z innym rozkładem poletek w stosunku do kierunku prowadzenia prac.

4	3	2	1	Kierunek prac↓
2	2	4	3	
3	4	1	2	
1	1	3	4	
Blok 1	Blok 2	Blok 3	Blok 4	

Rozkład wzdłużny

4	4	3	1
3	1	4	2
2	2	1	3
1	3	2	4
Blok 1	Blok 2	Blok 3	Blok 4

Rozkład poprzeczny

Blok 3		Blok 4	
4	2	3	1
3	1	4	2
2	4	2	3
1	3	1	4
Blok 1		Blok 2	

Rozkład hybrydowy

Rys. 9 Przykład wykorzystania poletek poddawanych kontroli imbrykowanej dla badania wykorzystującego zrandomizowane bloki o czterech blokach i czterech zabiegach.

Tabela 2 Różne sposoby obserwacji i rodzaje zmiennych

Zmienna	Pomiar	Ocena wzrokowa	Ustalanie rankingu	Scoring
Binarna				X
Nominalna				X
Porządkowa			X	X
Dyskretna	X	X		
Ciągła skończona	X	X		

Ciąga nieskończona	X	X		
-----------------------	---	---	--	--