

O właściwą interpretację wyników doświadczeń

Na wynik każdego doświadczenia wpływa cały szereg przyczyn, które możemy podzielić na dwie grupy. Do pierwszej zaliczymy efekt badanego przez nas czynnika, lub — w doświadczeniu złożonym — badanych czynników, do drugiej wszystkie przyczyny uboczne zaciemniające i wypaczające działanie pierwszej grupy. Do tej drugiej grupy należą we wszystkich badaniach niedokładności techniczne, w doświadczeniach polowych ponadto tzw. zmienność glebowa. Poprawne pod względem technicznym wykonanie doświadczenia powinno zmniejszyć wpływ pierwszej części przyczyn ubocznych, natomiast zastosowanie odpowiedniej metody powinno pozwolić na poddanie kontroli i skutkiem tego wyodrębnić w analizie zmienności znacznej części zmienności glebowej.

Są to rzeczy powszechnie znane i uznawane dzisiaj przez przeważającą większość naszych pracowników naukowych. Znajdują się jednak jeszcze wyjątki potwierdzające regułę, spotyka się czasem zdania, że nie chodzi o to, czy różnice wywołane efektem badanego czynnika są — w znaczeniu statystycznym — istotne czy nieistotne, lecz czy są one duże czy małe. Dlatego na początek przytoczę wyniki dwu doświadczeń, w których wpływ zmienności glebowej wypaczył przeciętne plony do tego stopnia, że duże różnice między obiektami były zupełnie nieistotne i wynikały raczej z niekontrolowanych przyczyn ubocznych.

Pierwsze z tych doświadczeń, z odmianami jęczmienia, przeprowadził w roku 1951 poznański Inspektorat Doświadczalnictwa Masowego w PGR Słupia Wielka. Z pięciu wysianych w tym doświadczeniu odmian pierwsze miejsce z plonem ziarna 25,9 q/ha zajął Hanna Kleszczewski, a ostatnie Browarny PZHR z plonem 20,6 q/ha. Analiza zmienności tego doświadczenia wykazała jednak, że ta stosunkowo duża, wynosząca ponad 5 q/ha, różnica plonu ziarna była zupełnie nieistotna. Doświadczenie to założono — błędnie — w układzie systematycznym, odmiany uszeregowano w każdym powtórzeniu w tej samej kolejności. Na polu doświadczalnym panowała bardzo duża jednokierunkowa zmienność glebowa. Skutkiem tej zmienności średni plon wszystkich odmian w pierwszym powtórzeniu wynosił 31,5 q/ha, w każdym następnym o 4—5 q/ha mniej, a w piątym 13,8 q/ha; plony szóstego powtórzenia były tak niskie, że wykonawca nie wziął ich do obliczeń. Ta zmienność glebowa wypaczyła wyniki do tego stopnia, że za udowodnione z prawdopodobieństwem 0,95 można by było uznać tylko te różnice, które przekraczałyby 10 q/ha, a więc około 45% przeciętnego plonu wszystkich odmian.

Drugie doświadczenie było innego rodzaju. Założono je w Zakładzie Doświadczalnym Czechnica poprawnie, obiekty rozmieszczono losowo w siedmiu blokach. Celem doświadczenia było zbadanie wartości różnych mieszanek koniczyny z trawami jako przedplonów pod jarą pszenicę.

Najwyższy plon ziarna pszenicy, 19,4 q/ha, uzyskano po koniczynie z tymotką i kupkówką, najniższy, 13,0 q/ha, po koniczynie z tymotką szczepioną azotobakterem. Różnica między tymi plonami jest bardzo duża, wynosi około 33% plonu odmiany czołowej, jest ona większa od przedziału ufności, który — przy współczynniku ufności 0,95 — wynosi 5,39 q/ha. Wydawałoby się więc, że różnicę tę możemy uznać za istotną. Analiza zmienności wyników tego doświadczenia wykazuje nam jednak, że wniosek ten byłby niesłuszny. Jak wiadomo, analiza zmienności jakichkolwiek danych liczbowych jest to podział zmienności całkowitej, mierzonej sumą kwadratów odchyień tych liczb od ich średniej, na części w zależności od przyczyn, które wywołały daną część zmienności. Analizę zmienności omawianego doświadczenia przedstawia tabela 1.

Tabela 1

Analiza zmienności

Zmienność	Stopni swobody	Suma kwadratów	Przeciętna zmienność
Bloków	6	367,95	61,32 S ₁
Objektów	8	104,59	13,07 S ₂
Nieścistości	48	431,52	8,99 S ₃

Stosunek S₂ do S₃ wynosi 1,45; jest on znacznie mniejszy od odpowiedniego wskaźnika e^{2z} Fishera, który dla danych liczb stopni swobody wynosi 2,14. Nie możemy zatem przyjąć, że badany czynnik wywarł istotny wpływ na plony; przyczyny uboczne przez wykonawcę niekontrolowane zaciemniły i wypaczyły wyniki doświadczenia do tego stopnia, że jakiegokolwiek wnioskiowanie stało się niemożliwe.

Wydaje się, że najważniejszą z przyczyn ubocznych była w tym doświadczeniu — podobnie jak w pierwszym przykładzie — zmienność glebowa. Poprawne założenie doświadczenia pozwoliło wyeliminować — jako zmienność bloków — przeszło 40% zmienności całkowitej. Gdyby zmienność glebowa była jednokierunkowa, byłby prawdopodobnie układ losowanych bloków pozwolił zmniejszyć znacznie bardziej zmienność wewnątrzblokową. W danym jednak wypadku zmienność miała raczej charakter fluktuacyjny, wskutek czego losowanie tylko w części zapobiegło wypaczeniu przez nią wyników. Musimy przyjąć, że bardzo poważne różnice między obiektami nie obrazują wartości tych obiektów lecz wynikają przede wszystkim ze zmienności glebowej.

Z podanego przykładu można by wysnuć wniosek, że wskaźniki (testy) statystyczne nie są między sobą zgodne i że zależnie od tego, który wskaźnik zastosujemy, stwierdzimy istotność różnic lub tej istotności nie wykazemy. Wniosek taki byłby błędny. Wskaźnik t , będący podstawą do obliczania przedziału ufności, służy do stwierdzenia istotności różnic między dwoma obiektami. Maleje on ze wzrostem liczby stopni swobody i dlatego przy kilku obiektach, a zatem przy większej liczbie stopni swobody, można go stosować tylko w tym przypadku, jeśli analiza zmienności wykaże istotność badanego czynnika. W danym przykładzie przedział ufności, obliczony tylko dla dwu skrajnych obiektów, wyniósłby przy 6 stopniach swobody 7,08 q/ha. A zatem poprawne zastosowanie testu t prowadzi rów-

niez do stwierdzenia, że różnica między plonami tych obiektów wynosząca 6,4 q/ha jest nieistotna.

Znaczną zmienność glebową spotyka się często na glebach lekkich; jest ona nieraz wynikiem różnej głębokości zalegania w podłożu warstwy zwięźlejszej. Wobec konieczności wykonywania większej liczby doświadczeń na glebach lekkich należałoby opracować metody pozwalające na skuteczne wyeliminowanie wpływu tej zmienności na wyniki. Sądzę, że celowe byłoby zwiększenie liczby powtórzeń przy ewentualnym zmniejszeniu powierzchni poletek. Obok tego należałoby te doświadczenia powtarzać w czasie i przestrzeni oraz opracowywać łącznie większe serie wyników. W takiej serii zmienność związana z reagowaniem obiektów na warunki, jakie spotykają one w różnych latach i miejscowościach, jest często mniejsza niż wynikający ze zmienności glebowej błąd poszczególnych doświadczeń.

Wydaje mi się, że podane przykłady ilustrują wystarczająco — nawet z punktu widzenia praktycznego — zasadę niemożności wnioskowania z wyników nieopracowanych statystycznie, choćby różnice między obiektami były bardzo znaczne. Pozwalają one też stwierdzić, że obliczenia wykonane w sposób przeważnie u nas spotykany, w których sprawdzianem istotności różnic jest tzw. przedział ufności, nie wystarczają nieraz do prawidłowego wyciągnięcia wniosków z wyniku doświadczenia. Przeważnie dopiero przeprowadzenie pełnej analizy zmienności umożliwia poprawną interpretację wyników. Często możemy przy tym nieistotną zmienność obiektów podzielić na podgrupy i znaleźć w ten sposób czynnik, który istotnie wpłynął na wyniki danego doświadczenia. Objaśnię to na przykładzie doświadczenia wykonanego w roku 1952 w Zakładzie Doświadczalnym Mochełek pow. Bydgoszcz nad techniką stosowania nawozów fosforowych i potasowych pod żyto. Oto wyniki tego doświadczenia:

Badane objekty	Plon ziarna żyta w q/ha
(Sposób stosowania fosforu i potasu)	
PK na dno bruzdy	26,0
PK pod bronę	25,1
PK pod pług	24,7
PK 75% pod pług, 25% pod bronę	24,3
PK pod kultywator	24,2
bez fosforu i potasu	22,1
przedział ufności (przy wsp. ufn. 0,95)	3,49

Większą od przedziału ufności jest zwyczajka plonu spowodowana przez umieszczenie fosforu i potasu na dnie bruzdy; wydawałoby się więc, że tylko ten sposób stosowania nawozów należałoby — na podstawie wyniku omawianego doświadczenia — uznać za celowy. Analiza zmienności tych wyników (tabela 2) prowadzi nas jednak do innych wniosków.

Stosunek przeciętnej zmienności obiektów do przeciętnej zmienności nieściśłości wynosi 1,26; a odpowiedni wskaźnik e^{2z} — 2,77. Zmienność obiektów jest zatem zupełnie nieistotna. W ten sposób wykonana analiza zmienności tego doświadczenia nie uprawnia nas do wyciągania żadnych wniosków; nie uzasadniony jest zatem wniosek, do którego doszliśmy na podstawie przedziału ufności.

Tabela 2

Analiza zmienności

Zmienność	Stopni swobody	Suma kwadratów	Przeciętna zmienność
Całkowita	23	130,74	—
Objektów	5	33,92	6,76 S_1
Nieścistości	18	96,82	5,38 S_2

Zastanawiając się jednak głębiej nad tym doświadczeniem stwierdzimy, że zmienność wywołaną badanym czynnikiem możemy podzielić na dwie części: 1) zmienność wywołaną działaniem fosforu i potasu; 2) zmienność sposobu stosowania tych nawozów. Pełną analizę zmienności wyników doświadczenia podaje tabela 3.

Tabela 3

Analiza zmienności

Zmienność	Stopni swobody	Suma kwadratów	Przeciętna zmienność
Objektów:	5	33,92	6,764 S_1
W tym:			
przeciętnego efektu PK	1	25,39	25,39 S_2
sposobów stosowania			
PK	4	8,53	2,13 S_3
Nieścistości	18	96,82	5,38 S_4
$\frac{S_1}{S_4} = 1,26;$	$e^{2z} = 2,77;$	$\frac{S_2}{S_4} = 4,72;$	$e^{2z} = 4,35;$

Z analizy tej można wyciągnąć wniosek uzasadniony statystycznie, ale zupełnie inny od sformułowanego na podstawie przedziału ufności. Istotna w stosunku do zmienności nieścistości jest tylko zmienność przeciętnego efektu PK. Możemy więc uznać za udowodnioną zwyżkę wywołaną działaniem fosforu i potasu; zwyżka ta wynosi w danym doświadczeniu przeciętnie 2,76 q ziarna żyta z hektara. Natomiast nie mamy nawet

Tabela 4

Doświadczenia nad dawkami i porą stosowania azotu pod żyto

Obiekty	bez azotu	30 N 40 N		30 N 40 N	
		w całości wiosną	1/4 jesienia	3/4 wiosną	rzutowo
Miejscowość	p l o n z i a r n a ż y t a w q / h a				
Wielichowo	19,2	24,2	26,4	25,8	26,2
Poświętne	34,5	42,0	45,3	41,0	44,0
Sternalice	15,7	21,5	22,9	20,2	22,6
Poraj	12,1	18,8	22,8	18,7	21,6
średnio	20,4	26,6	29,3	26,4	28,6

najmniejszych danych do przypuszczania, że sposoby stosowania fosforu i potasu wywarły wpływ na wysokość plonu; musimy uznać, że różnice między tymi obiektami wywołane są raczej przyczynami przez nas niekontrolowanymi, a należącymi do ogólnej grupy błędu (nieścisłości) doświadczenia.

Wnikliwe przeprowadzenie analizy zmienności umożliwia nam często szczegółowe wyjaśnienie wyników i zbadanie wysokości wpływu każdego czynnika na wynik doświadczenia. Cztery Zakłady położone na glebach lekkich wykonały w roku 1951 doświadczenie nad dawkami i porą stosowania azotu pod żyto. Wyniki tych doświadczeń zestawiono w tabeli 4.

Pełną analizę zmienności tych wyników przedstawia tabela 5.

Tabela 5

Analiza zmienności

Zmienność	Liczba st. swobody	Suma kwadratów	Przeciętna zmienność
Miejscowości	3	1597,66	532,55 S_1
Obiektów	4	199,26	49,82 S_2
W tym:			
przeciętnego efektu działania azotu	1	174,05	174,05 S_3
dawek azotu	1	24,01	24,01 S_4
pory stosowania	1	0,90	0,90 S_5
współdziałania pory stosowania z dawkami	1	0,30	0,30 S_6
Współdziałania obiektów z miejscowościami	12	10,12	0,84 S_7
Nieścisłości	76	57,85	0,76 S_8

A oto wnioski z tej analizy: pomimo dużych różnic w przeciętnych plonach (duża zmienność miejscowości) wyniki są zgodne; badane obiekty nie reagowały istotnie inaczej na warunki, w jakich wykonano każde doświadczenie; świadczy o tym niewielka — w stosunku do nieścisłości — zmienność współdziałania obiektów z miejscowościami. Azot działał bardzo silnie: zmienność wywołana przez przeciętny efekt działania tego nawozu wynosi blisko 90% całej zmienności obiektowej. Istotny lecz znacznie mniejszy wpływ na wysokość plonów spowodowało zwiększenie dawki z 30 na 40 kg N na hektar. Natomiast wpływ pory stosowania, a więc zastosowania azotu w całości na wiosnę zamiast $\frac{1}{4}$ jesienią i $\frac{3}{4}$ wiosną, jest zupełnie nieistotny; nie można na podstawie wyników tej serii doświadczeń zalecać jednego lub drugiego sposobu stosowania azotu. Jeszcze mniejsza jest zmienność współdziałania pory stosowania z dawkami: nie możemy stwierdzić żadnej zależności między tymi dwoma badanymi czynnikami.

Gdyby zmienność współdziałania obiektów z miejscowościami była większa, gdyby zatem wyniki uzyskane w poszczególnych miejscowościach nie były zgodne, należałoby rozdzielić tę zmienność na części według tej samej zasady, którą zastosowano przy dzieleniu zmienności obiektowej.

Można by wówczas stwierdzić, który z czynników spowodował niezgodność wyników i próbować wytłumaczyć powody tej niezgodności. W danej serii, stosunkowo bardzo jednolitej, jest to oczywiście niepotrzebne.

Podaję parę przykładów wziętych z prac Sekcji Metodyki Doświadczalnictwa IUNG, dla zwrócenia uwagi na to, że tylko dokładne i wnikliwe statystyczne opracowanie materiału stanowiącego wynik doświadczenia umożliwia poprawne wnioskowanie i uzyskanie maksimum informacji. Oczywiście można nieraz pójść we wnioskach dalej i omówić wyniki statystycznie nieudowodnione; wnioski takie wskazują zwykle kierunek dalszych badań, które wykażą, czy sugestie te były słuszne czy nie. Jednak autor jak i czytelnik danej pracy muszą sobie zdawać sprawę z tego, które wnioski wynikają bezpośrednio z danego materiału, a które mają charakter przypuszczeń i wymagają sprawdzenia lub stanowią hipotezę wyjściową dla dalszych badań.