R. ELANDT (Poznań)

O PEWNYCH TESTACH INTERAKCJI W DOŚWIADCZENIACH WIELOLETNICH I WIELOKROTNYCH. ZAGADNIENIE REJONIZACJI

Wstęp

Przy opracowywaniu wyników różnych doświadczeń rolniczych przeprowadzonych w czasie i przestrzeni często odczuwa się brak łatwych i szybkich metod statystycznych, pozwalających na ugrupowanie materiału, przy którym można wyciągnąć jak najwięcej możliwie dokładnych wniosków.

Metodyka planowania i obliczania pojedynczych doświadczeń prostych i złożonych jest już dość szeroko opracowana; wystarczy przeczytać podręczniki takich autorów, jak W. G. Cochran i G. M. Cox [4], O. Kempthorne [17], N. H. Quenouille [34], C. R. Rao [35], G. W. Snedecor [42], żeby zorientować się w najnowszych osiągnięciach w tej dziedzinie.

Jednakże podstawę wnioskowania rolniczego mogą stanowić dopiero doświadczenia przeprowadzone w wielu punktach i w różnych latach, znane pod nazwą *doświadczeń wielokrotnych i wieloletnich*. Synteza statystyczna takich doświadczeń jest zagadnieniem istotnym, a jednocześnie trudnym, bo przeważnie nie mamy materiałów jednolitych, tzn. nie wszystkie obiekty doświadczenia powtarzają się w tych samych punktach przez wszystkie lata. Wynika to często z braku należytego planowania doświadczeń ogólnopolskich przez różne instytucje naukowe, czasem jednak jest koniecznością wynikającą z przesłanek logicznych. Jeżeli np. pewna odmiana daje w jakiejś miejscowości wyniki zdecydowanie ujemne, to na pewno usunie się ją w roku przyszłym zastępując odmianą odpowiedniejszą. Także ścisłość poszczególnych doświadczeń nie jest jednakowa.

Właściwe posegregowanie materiału statystycznego otrzymanego z takich doświadczeń oraz ustalenie zależności i rozbieżności między poszczególnymi obiektami odgrywa postawową rolę przy ich opracowaniu. Zaletą metod stosowanych przy takiej analizie wyników jest niewielka ilość rachunków oraz ich szybkość i łatwość. Najlepiej nadają się do tego statystyki pozycyjne i zbudowane na nich testy nieparametryczne [48].

W niniejszej pracy podaję kilka takich testów. Są one bądź pomysłem oryginalnym [11], bądź też modyfikacją testów już znanych [14], [29], zbudowanych specjalnie dla tego zagadnienia.

Po takiej analizie wstępnej, dającej nieraz dużo informacji a wymagającej stosunkowo mało pracy, można przystąpić do obliczeń bardziej szczegółowych. Jedną z metod służących do ilościowego badania pewnych zależności jest wprowadzony tu test χ^2 dla różnic, który w pewnym sensie ma w tym przypadku zastąpić analizę wariancji.

Należy zwrócić uwagę, że wszystkie powyższe testy mają zastosowanie nie tylko przy opracowywaniu doświadczeń wielokrotnych. Ponieważ jednak pomysł ich był wynikiem szukania nowych metod syntezy statystycznej doświadczeń wielokrotnych, więc też w pracy niniejszej podano je w tym tylko zastosowaniu. Zdaję sobie też sprawę, że metody tu rozważane nie wyczerpują całego zagadnienia, które jeszcze długo będzie kwestią otwartą.

Zanim przejdę do podania tych metod, omówię pokrótce doświadczenia wieloletnie i wielokrotne z punktu widzenia ich roli i znaczenia dla praktyki.

I. Doświadczenia wieloletnie i wielokrotne

1. Rola i znaczenie doświadczeń wieloletnich i wielokrotnych. Każde doświadczenie ścisłe, wielokrotne, przeprowadzone w jednej miejscowości i w jednym roku może przedstawiać pewien materiał porównawczy, nawet dość ścisły dla danych warunków klimatyczno-glebowych, ale dla praktyki rolniczej nie mający w tej formie wielkiego znaczenia. Stanowi ono jedynie pewną informację dotyczącą badanego zjawiska w określonych warunkach. Warunki te zaś zmieniają się w zależności od tego, czy rok był suchy czy wilgotny, czy wiosna była wczesna czy późna, czy dana roślina była wysiana w tej czy innej miejscowości itd. Dopiero uwzględnienie całego kompleksu warunków zmieniających się z roku na rok i zależnych od miejscowości może dać pewien przeciętny obraz przebiegu danego zjawiska i stworzyć podstawę do wniosków ogólniejszych. Stąd zrozumiałe jest żądanie, żeby każde doświadczenie — proste czy kombinowane — było powtórzone w czasie i przestrzeni, a więc żeby było doświadczeniem wieloletnim i wielokrotnym.

Na marginesie warto jeszcze poruszyć tu rolę doświadczeń jednokrotnych. Nie są one takie dokładne, jak doświadczenia wielokrotne, jednak wielka ich ilość może dać w wielu przypadkach próbę ([1]) bardziej reprezentatywną dla pewnego rejonu niż znacznie mniejsza ilość do-

([1]) Zgodnie z życzeniem Autorki zachowujemy w tym miejscu i w miejscach podobnych termin *próba* zamiast terminu *próbka*, który tu uważamy za poprawny. *(Redakcja).*

świadczeń wielokrotnych [10]. O. Kempthorne [17] twierdzi, że najekonomiczniejszą, tzn. dającą dużo informacji, a wymagającą niewielkiej pracy i kosztów, jest liczba powtórzeń równa 2.

Jak już zaznaczyliśmy, doświadczenia wieloletnie i wielokrotne przeprowadzamy dlatego, że spodziewamy się, iż w odmiennych warunkach klimatyczno-glebowych różne gatunki roślin mogą reagować odmiennie na różne terminy siewu, sposoby uprawy, rodzaje nawożenia itp., czyli — jak mówimy — mogą wykazywać *interakcję*. Badanie interakcji gra główną rolę w doświadczeniach kombinowanych, wieloletnich i wielokrotnych, i należy poświęcić temu zagadnieniu kilka słów.

2. Zagadnienie interakcji. Żadnego zjawiska w czasie i przestrzeni nie można rozpatrywać jako odosobnionego i niezależnego. W szczególności zasada ta ma istotne znaczenie w badaniu zjawisk przyrodniczych, biologicznych.

Przypuśćmy, że rozpatrujemy wartości pewnej cechy dwóch obiektów, np. średni plon dwóch odmian pszenicy A i B, w pewnych określonych warunkach środowiska, np. w pewnej miejscowości I. Jeżeli te same odmiany wysiejemy w miejscowości II o odmiennych warunkach klimatyczno-glebowych, to ich reakcja na zmianę środowiska może w ogólnym przypadku nie być jednakowa. Może się np. okazać, że plon jednej odmiany wzrośnie, a drugiej obniży się; może się też zdarzyć, że plony obu odmian wzrosną albo się obniżą, jednak zwyżka czy zniżka np. u odmiany A będzie znacznie większa niż u odmiany B. W obu przypadkach odmiany nie zachowują się jednakowo w zmienionych warunkach środowiska; mówimy wtedy, że współdziałają one ze środowiskiem, czyli wykazują istotną interakcję *obiektowo-lokalną*. Podobnie możemy określić interakcję *obiektowo-sezonową*.

Należy tu jeszcze zwrócić uwagę na właściwe używanie terminu interakcja. Przypuśćmy, że odmiany A i B wysiane w miejscowości I dały odpowiednio plony średnie $\bar{x}_A$ i $\bar{x}_B$, przy czym $\bar{x}_A$ może być różne od $\bar{x}_B$; te same odmiany wysiane w miejscowości II dały plony odpowiednio $\bar{x}_A+d+e$ oraz $\bar{x}_B+d+e'$, gdzie d jest dowolną liczbą rzeczywistą dodatnią albo ujemną, a e oraz e' błędami przypadkowymi o rozkładzie $N(0, \sigma)$. Czy wówczas powiemy, że interakcja istnieje? Nie. Oczywiście, że wraz ze zmianą środowiska nastąpiła zmiana plonu, a ponieważ była ona jednakowa dla obu odmian, więc została spowodowana głównym działaniem środowiska, tj. przede wszystkim działaniem czynników klimatyczno-glebowych, które miały jednakowy wpływ na zmianę plonu obu odmian. Rozumowanie w przypadku n obiektów opiera się na podobnych założeniach logicznych, lecz jego forma matematyczna jest bardziej skomplikowana. Dlatego też mówiąc o interakcji rozważamy zawsze co najmniej 2 obiekty i oceniamy wielkość zmian badanej cechy wraz ze zmianą wa-

runków. Jeżeli te zmiany są niewielkie, mieszczące się w granicach błędów przypadkowych, to mówimy, że reakcja tych obiektów jest podobna, i nie możemy powiedzieć, że interakcja istnieje. Interakcja jest zatem pojęciem porównawczym (a więc w pewnym sensie względnym) i wiąże się z określeniem niejednakowej, równoczesnej reakcji kilku (co najmniej dwóch) obiektów na zmianę warunków [9], nie należy go więc mieszać ze zwykłą zmianą wartości badanej cechy pojedynczego obiektu przy zmianie tychże warunków.

Badanie interakcji stanowi podstawowy problem syntezy doświadczeń wieloletnich i wielokrotnych oraz łączy się ściśle z zagadnieniem rejonizacji.

3. Zagadnienie rejonizacji. Jednym z zasadniczych zadań nauki rolniczej jest określenie dla poszczególnych gatunków czy odmian roślin właściwych im rejonów, w których dawałyby one najlepsze plony, czyli — jak mówimy — *rejonizacja* tych gatunków czy odmian. Główną rolę w rejonizacji odgrywają czynniki klimatyczno-glebowe, ale nie tylko one. Na wielkość plonu może wpłynąć wiele innych ważnych czynników. Do nich należy zaliczyć przede wszystkim termin siewu, nawożenie i uprawę. Dopiero rozpatrzenie całego kompleksu warunków może zadecydować o właściwej rejonizacji.

Jak zaznaczyliśmy, podstawę rejonizacji stanowi wielkość plonu. Lecz nie jest to jedyna cecha, którą w tym zagadnieniu kieruje się rolnik. Drugim niemniej ważnym warunkiem jest ryzyko uprawy. Na przykład niektóre pszenice są intensywne w rejonach dla nich niekorzystnych, lecz w warunkach dobrej uprawy często chorują na rdzę [37] i wobec tego, chociaż plon ich może być nawet względnie większy w porównaniu z innymi odmianami, to jednak korzystniejsze może okazać się nieraz wysianie odmian mniej plennych, ale pewniejszych. Dalej trzeba jeszcze wziąć pod uwagę, że odmiany wysiewane przez kilka lat w tej samej miejscowości wyradzają się i wobec tego gra tu rolę ziarno siewne.

Wreszcie trzecią ważną cechą mającą wpływ na rejonizację jest czynnik ekonomiczny, wiążący się bezpośrednio zarówno z wysokością plonu, jak i z ryzykiem uprawy.

Uwzględnienie wszystkich czynników wpływających na powyższe cechy wymaga doświadczeń różnego typu, gęsto rozmieszczonych w terenie i prowadzonych przez kilka lat. Doświadczenia te powinny być odpowiednio zaplanowane, tak by mogły dawać materiał wiarygodny. W tej sprawie nasze doświadczalnictwo i planowanie ma jeszcze dużo do zrobienia.

Mówiąc o rejonizacji mieliśmy dotychczas na myśli jeden obiekt (gatunek albo odmianę) i szukaliśmy kryteriów, za pomocą których można

by znaleźć dla niego najkorzystniejsze rejony uprawy. Prowadzone doświadczenia miały charakter porównawczy: rozpatrywaliśmy w nich nie jeden, lecz kilka obiektów. Można by więc zagadnienie postawić nieco inaczej, ogólniej: szukamy takich obiektów (np. odmian), które by w pewnych rejonach wykazały podobne plony, podobną reakcję na zmianę środowiska i podobną zmienność. Wówczas można by wyodrębnić grupy homogeniczne obiektów dla pewnych rejonów lub określić, jakich wyników możemy się spodziewać, jeżeli w pewnej miejscowości (czy rejonie) wysiejemy obiekty należące do grupy heterogenicznej. Przy tak postawionych zagadnieniach analizować będziemy średnie plony oraz interakcję obiektowo-lokalną i obiektowo-sezonową.

Nie jest celem mojej pracy omówienie wszystkich znanych dotychczas metod, stosowanych przy rozwiązaniu tych zagadnień; można je znaleźć w odpowiednich źródłach (zob. [3], [4], [10], [17], [19], [21], [24]-[26], [32]-[39], [42], [45]). Naszkicuję tu tylko w krótkim rzucie historycznym kilka z tych sposobów syntezy najczęściej stosowanych oraz zwrócę uwagę na ich zalety i braki, aby w ten sposób dać krótki pogląd na istotę tego zagadnienia i wskazać dalsze możliwości prac w tym kierunku.

4. Kilka podstawowych metod opracowywania doświadczeń wieloletnich i wielokrotnych. A. Metody bezpośrednie i wzorcowe. Analiza wyników doświadczeń wieloletnich i wielokrotnych polega na porównaniu obiektów bądź bezpośrednio między sobą, bądź za pomocą wzorca. Oba porównania mają swoje zalety i wady.

Porównanie bezpośrednie polega na zastosowaniu odpowiednich metod statystycznych do wartości bezwzględnych badanej cechy (najczęściej plonu) otrzymanych z doświadczeń; przy porównaniu za pomocą wzorca stosuje się te metody do liczb otrzymanych jako odchylenia od pewnej wielkości przyjętej za standard. Jeżeli np. w doświadczeniach odmianowych kilka odmian powtarza się we wszystkich punktach doświadczalnych, to jako taki standard przyjmuje się ich średnią w danej miejscowości, a plony pozostałych odmian wyraża się w odchyleniach od tego wzorca zwanego *zbiorowym*. Wzorzec ten ma właściwie przedstawiać tzw. *poziom lokalny*, tj. charakteryzować przeciętną urodzajność danej miejscowości ze względu na badane odmiany, a odchylenia dodatnie lub ujemne pozostałych odmian mają obrazować plony lepsze lub gorsze od standardu. Otóż wzorzec nie zawsze może spełnić swoją rolę. Po pierwsze, ponieważ za wzorzec bierzemy zwykle odmiany powtarzające się we wszystkich miejscowościach i latach, więc może się zdarzyć, że odmiany te będą wybrane dość przypadkowo. Jeżeli będą one zachowywać się podobnie we wszystkich miejscowościach, to np. w razie warunków niesprzyjających dla wzorca, pewne odmiany będą wykazywały

znaczne zwyżki; zostaną „przecenione" [39]. Wzorzec nie obrazuje w tym przypadku właściwie poziomu lokalnego. Za najlepszy wzorzec zbiorowy uważać można średnią ze wszystkich obiektów, powinien jednak być przy tym spełniony warunek, że wszystkie obiekty powtarzają się we wszystkich miejscowościach i latach, co rzadko zachodzi [39]. Po drugie, wzorzec nie jest wielkością dostatecznie stałą w różnych latach; stąd zmienność sezonowa obliczona na odchyleniach jest niewłaściwa i nie można obliczać średnich wieloletnich [37].

Metody wzorcowe są jednak o tyle wygodne, że przeprowadza się przy ich użyciu mniej porównań, a więc i mniej obliczeń. Jeżeli np. w doświadczeniu mamy n obiektów, to aby każdy obiekt porównać z wzorcem, trzeba wykonać n analiz, natomiast aby porównać każdy obiekt z każdym (z wyjątkiem klasycznej analizy wariancji), trzeba tych porównań wykonać aż $n(n-1)/2$.

Metody opracowywania doświadczeń wielokrotnych w pracach zachodnich opierają się przeważnie — jak mi wiadomo — na analizie wyników bezpośrednich [3], [17], [42]; w opracowaniach polskich spotykamy najczęściej metody wzorcowe [32], [36]-[39]. Zwrócimy jeszcze uwagę, że obliczenia można równie dobrze prowadzić zarówno na wartościach bezwzględnych, jak i na odchyleniach od wzorca; w każdym jednak przypadku otrzymujemy odpowiedź na nieco inne pytania.

B. Analiza wariancji. Najprostszą metodą jest klasyczna analiza wariancji, w której miejscowości i lata traktuje się jako czynniki przy klasyfikacji n-dzielnej [33]. Jeżeli ilość powtórzeń, ilość obiektów oraz dokładność we wszystkich doświadczeniach jest jednakowa, to ocenę wariancji resztowej s^2 dla wszystkich doświadczeń oblicza się jako średnią arytmetyczną wariancji resztowych poszczególnych doświadczeń s_i^2. W dalszym ciągu, zależnie od tego, co nas interesuje — odmiany, miejscowości czy lata lub interakcje tych czynników — sprawdzamy odpowiednio postawione hipotezy zerowe za pomocą testu F (zob. [17]). W praktyce najczęściej założenia te nie są spełnione, a to z dwóch powodów. Po pierwsze, poszczególne doświadczenia przeprowadzone w różnych miejscowościach dają różniące się istotnie oceny wariancji resztowej s_i^2. Ten przypadek zachodzi prawie zawsze. Dzieje się to wskutek bądź niejednakowej dokładności wykonania samych doświadczeń, bądź wskutek rzeczywiście innej zmienności, charakteryzującej zupełnie odmienne typy gleb. Wówczas na całym materiale można przeprowadzić tylko przybliżoną analizę wariancji [4] albo podzielić materiał na grupy, dla których te warunki są spełnione [10], [19]. Po drugie, ilość obiektów, a czasami nawet ilość powtórzeń w poszczególnych doświadczeniach nie jest jednakowa. Niektóre przyczyny tego wyjaśniono we wstępie. Prócz tego

te same doświadczenia nie zawsze są powtarzane w ciągu kilku lat w tych samych miejscowościach. Składa się na to wiele przyczyn. Między innymi chociażby i ta, że w ogólnym płodozmianie gospodarstwa rolnego takie doświadczenie może się czasami nie mieścić.

Do opracowania takiego nieortogonalnego układu stosuje się metody już bardziej skomplikowane. Gdy można przyjąć, że interakcja obiektowo-lokalna jest nieistotna, wówczas najczęściej porównuje się *średnie ważone*, gdzie *wagi* są obliczone na podstawie odwrotności wariancji [3], [34] Inną, bardziej skomplikowaną metodą, wymagającą długich rachunków jest metoda *dopasowywania stałych* [18], [42].

W większości przypadków przyjmujemy jednak, że średnie obiektowe wykazują istotne wahania między doświadczeniami, czyli że interakcja obiektowo-lokalna istnieje. Wówczas okazuje się lepsze porównanie średnich zwykłych, przy zastosowaniu analizy tzw. *ważonych* kwadratów dla średnich. Szczegóły tego postępowania opisane są w podręczniku G. W. Snedecora [42].

Przy badaniu interakcji obiektowo-lokalnej, zwłaszcza gdy doświadczenia nie są jednakowo dokładne i na całym materiale można przeprowadzić tylko analizę wariancji przybliżoną, może się okazać, że interakcja ta jest heterogeniczna i zbyt duża. Wówczas trzeba pójść dalej i przez rozbijanie obiektów na grupy szukać grup homogenicznych [21], [34]. Czasami jest odwrotnie — interakcja jest zbyt mała i stosujemy rozbijanie, aby otrzymać grupy heterogeniczne [4], [34].

Dłuższe studia nad możliwościami opracowywania doświadczeń wielokrotnych, nie przedstawiających materiału zgodnego zarówno co do ilości powtórzeń, jak i ilości obiektów oraz punktów doświadczalnych, nasuwają myśl, że najprostszym punktem wyjścia powinno być pojedyncze doświadczenie, do którego bez żadnych zastrzeżeń można zastosować analizę wariancji. Dalej rozważamy układ doświadczeń we wszystkich miejscowościach w każdym roku osobno. Łączenie różnych doświadczeń z okresu kilku lat często nie tylko nie daje całokształtu przebiegu zjawiska, lecz zupełnie ten obraz zamazuje. Dopiero wstępne opracowanie każdego roku osobno i porównanie wyników mogą wskazać pewne prawidłowości i kierunki, ocenić ryzyko uprawy, dać podstawy do wyprowadzania wniosków ogólniejszych.

Porównywanie ze sobą lat czy obiektów w poszczególnych latach lub miejscowościach opiera się — przy założeniu normalności — na korzystaniu ze zwykłego błędu standardowego różnicy i zbudowanej na nim odpowiedniej różnicy granicznej, gdy znane są oceny nieobciążone s_1^2 i s_2^2 wariancji dwóch obiektów. Szczegółowe omówienie tej metody, którą będziemy nazywali tu *metodą przedziałów ufności* lub *metodą różnic granicznych*, zaprojektowanej przeze mnie do syntezy doświadczeń

odmianowych z lnem włóknistym, podano w pracy [19]. Jakkolwiek nie przedstawia ona w zasadzie tego, co można by nazwać teorią zagadnienia, to jednak jest w użyciu dość prosta i dla celów praktyki rolniczej w wielu przypadkach wystarczająco dokładna.

C. Analiza regresji. Jedną z najbardziej uniwersalnych i efektownych metod stosowanych do opracowania doświadczeń wielokrotnych jest *analiza regresji* w połączeniu z metodą *krzywych ufności*, zaprojektowana i szczegółowo opracowana przez J. Neymana [25], [26]. W metodzie tej przyjmuje się, że między plonami x i y dwóch obiektów A i B rozpatrywanych parami w wielu punktach doświadczalnych, zachodzi zależność liniowa. Postawione pytanie brzmi mniej więcej następująco: Jakiej przeciętnej nadwyżki plonu obiektu B można oczekiwać, gdy plon obiektu A jest równy x?

Korzystając z metody najmniejszej sumy kwadratów wyprowadza się prostą regresji, a następnie przeprowadza się analizę tej zależności rozważając przypadki, gdy prawdziwy współczynnik regresji β jest równy, mniejszy albo większy od jedności. Przypadki $\beta \leqslant 0$ nie są tu rozpatrywane, jako mało prawdopodobne.

Oszacowanie błędu prostej regresji prowadzi do wykreślenia tzw. krzywych ufności i budowania pasów ufności dla nadwyżki Δy, w zależności od x.

Metodę tę można stosować do porównań bezpośrednich i wzorcowych, z pewnym jednak zastrzeżeniem. Należałoby każdorazowo sprawdzić, czy istnieje korelacja dodatnia między plonem jednej i drugiej odmiany. Są przykłady, zwłaszcza przy zastosowaniu wzorca, że takiej korelacji nie ma [10], [37], a czasami może ona być nawet ujemna [38]. Wówczas ten sposób analizy musi podlegać pewnej rewizji. Metody proponowane przeze mnie w następnych rozdziałach zmierzają do tego, by te zastrzeżenia uwzględnić. Do analizy regresji wrócimy jeszcze w rozdziale następnym.

Zapewne istnieje jeszcze wiele innych metod mogących posłużyć do syntezy doświadczeń wielokrotnych i wieloletnich. Właściwie każda synteza stanowi już pewną metodę, gdyż w zależności od rozważanego problemu stosujemy taki czy inny sposób postępowania.

II. Badanie kierunkowości. Tendencja

Z rozważań poprzedniego rozdziału wynika, że stosowanie znanych metod, jak np. analizy wariancji czy regresji, do bardzo różnorodnego materiału, jaki zwykle przedstawiają wyniki doświadczeń wielokrotnych, napotyka różne trudności. Dlatego też należy najpierw przeprowadzić pewną analizę wstępną materiału statystycznego, na jej podstawie

ocenić ten materiał i określić, jakie dokładniejsze metody można by do niego zastosować.

W rozdziale tym omówię kilka testów nieparametrycznych pozwalających na taką analizę wyników. Dogodność ich polega między innymi i na tym, że zostawiają one pewną dowolność założeń, a niektóre zakładają tylko bardzo proste i dość uniwersalne schematy matematyczne. To pozwala stosować je do tych zagadnień, które nie dają modeli ortogonalnych, a więc do zagadnień nastręczających doświadczalnikowi i statystykowi sporo kłopotu z opracowaniem. Poza tym są to testy przeważnie proste, nie wymagające ani zbyt trudnych, ani zbyt uciążliwych rachunków.

1. Nieparametryczny test tendencji. W wielu zagadnieniach rolniczych, zwłaszcza dotyczących rejonizacji, przy badaniu plonu lub innych cech interesuje nas zachowanie się obiektów co do kierunku ich wahań wokół pewnych statystyk, jak np. wokół średniej arytmetycznej, modalnej, mediany itp. Gdy funkcje rozkładu badanych zmiennych losowych są nieznane i co najwyżej możemy o nich założyć, że są ciągłe, wówczas do budowania testów nieparametrycznych nadają się najlepiej statystyki pozycyjne (np. mediana) i oparta na nich metoda randomizacji, zaprojektowana przez R. A. Fishera do weryfikacji hipotez statystycznych [2], [20], [41], [48], [49].

Rozważmy następujące zagadnienie: Niech $x_1, x_2, \ldots, x_N$ oznaczają wartości plonu pewnej odmiany A w N miejscowościach, a $y_1, y_2, \ldots, y_N$ odpowiednio wartości plonu odmiany B w tych samych miejscowościach. Oznaczmy przez Me_x i Me_y odpowiednie mediany i wprowadźmy nowe zmienne

$$\begin{aligned} \xi_i &= x_i - \mathrm{Me}_x, \\ \eta_i &= y_i - \mathrm{Me}_y, \end{aligned} \quad \text{dla} \quad i = 1, 2, \ldots, N. \tag{1}$$

Otrzymamy znowu dwa ciągi liczb $\xi_1, \xi_2, \ldots, \xi_N$ oraz $\eta_1, \eta_2, \ldots, \eta_N$, które przedstawiają wartości odchyleń plonów od mediany w każdej próbce w N miejscowościach. Jeżeli te odchylenia mają we wszystkich miejscowościach znaki zgodne, to mówimy, że obie odmiany wykazują *jednakową tendencję*, tzn. że w tych samych miejscowościach obie dają plony wyższe albo niższe od mediany, przy czym nic nie zakładamy o wielkości tych odchyleń. Podobnie, jeżeli znaki są niezgodne, to mówimy, że obie odmiany wykazują *niejednakową tendencję*. Można zaprojektować test, który pozwala weryfikować hipotezę o istnieniu tendencji (jednokierunkowej lub różnokierunkowej) dla każdej liczby $2r$ znaków parami zgodnych i przyjętego poziomu istotności.

Wyprowadzenie testu. Rozważmy próbę $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$ pobraną z dwuwymiarowej populacji, o której zakładamy tylko,

że jest ciągła; kształty rozkładów brzegowych i parametry tych rozkładów są nieznane.

Stawiamy hipotezę zerową H_0: *zmienne losowe X i Y są niezależne.* Hipotezy alternatywne są następujące: H_1 — *zmienne losowe X i Y mają jednakową tendencję*; oraz H_2 — *zmienne losowe mają niejednakową tendencję.* (Uwaga: obie te hipotezy alternatywne można by ująć w jedną, mianowicie: H — *obie zmienne losowe X i Y wykazują tendencje.*) Wprowadzamy statystykę U, którą określamy jako liczbę znaków parami zgodnych (lub niezgodnych); posłuży nam ona do weryfikowania hipotezy H_0.

a. Rozważmy najpierw przypadek, kiedy N jest parzyste; $N = 2n$. Wówczas w każdej próbie jest n plusów i n minusów.

Ilość możliwych układów znaków + i — w próbie (X) wynosi

$$\binom{2n}{n} = \frac{(2n)!}{n!n!};$$

podobnie w próbie (Y)

$$\binom{2n}{n} = \frac{(2n)!}{n!n!}.$$

Ilość wszystkich możliwych kombinacji znaków w próbie łącznej (X, Y) jest równa

$$\binom{2n}{n}^2 = \frac{[(2n)!]^2}{[n!]^4}.$$

Ilość sposobów otrzymywania $2r$ znaków zgodnych lub niezgodnych wynosi

$$\binom{2n}{n}\binom{n}{r}^2, \quad \text{gdzie} \quad r = 0, 1, 2, \ldots, n.$$

Zatem prawdopodobieństwo, że jest dokładnie $2r$ znaków zgodnych lub niezgodnych wynosi

$$\binom{2n}{n}\left[\binom{n}{r}^2 + \binom{n}{n-r}^2\right]\Big/\binom{2n}{n}^2 = 2\binom{n}{r}^2\Big/\binom{2n}{n}, \tag{2}$$

gdzie $r = 0, 1, \ldots, n/2$, gdy n jest parzyste, oraz $r = 0, 1, \ldots, (n-1)/2$, gdy n jest nieparzyste. Prawdopodobieństwo, że jest niemniej niż $2r$ znaków zgodnych albo niezgodnych, wynosi

$$\left[\sum_{i=r}^{n}\binom{n}{i}^2 + \sum_{i=0}^{n-r}\binom{n}{n-i}^2\right]\Big/\binom{2n}{n} = \left[\sum_{i=r}^{n}\binom{n}{i}^2 + \sum_{i=r}^{n}\binom{n}{i}^2\right]\Big/\binom{2n}{n} = \tag{3}$$

$$= 2\left[\sum_{i=r}^{n}\binom{n}{i}^2\right]\Big/\binom{2n}{n},$$

TABLICA 1

Prawdopodobieństwa określone wzorem $P(U \geqslant 2r) = \dfrac{\sum_{i=r}^{n} \binom{n}{i}^2}{\binom{2n}{n}} = \alpha$

2n \ 2r	0	2	4	6	8	10	12	14	16	18	20	22	24	26
2	1,0000	0,5000												
4	1,0000	,8333	0,1667											
6	1,0000	,9500	,5000	0,0500										
8	1,0000	,9858	,7572	,2429	0,0143									
10	1,0000	,9960	,8968	,5000	,1032	0,0040								
12	1,0000	,9990	,9600	,7165	,2836	,0401	0,0011							
14	1,0000	,9997	,9854	,8569	,5000	,1431	,0146	0,0003						
16		1,0000	,9950	,9341	,6904	,3097	,0660	,0051	0,0001					
18		1,0000	,9983	,9716	,8265	,5000	,1735	,0284	,0017					
20	1,0000	0,9999	,9994	,9884	,9105	,6718	,3281	,0894	,0115	0,0005				
22		1,0000	,9999	,9956	,9570	,8026	,5000	,1975	,0431	,0045	0,0002			
24			1,0000	,9984	,9805	,8899	,6579	,3422	,1102	,0196	,0017	0,0001		
26			1,0000	,9994	,9915	,9423	,7831	,5000	,2169	,0577	,0085	,0006		
28			1,0000	,9998	,9965	,9715	,8716	,6468	,3532	,1284	,0285	,0035	0,0002	
30			1,0000	,9999	,9986	,9966	,9285	,7670	,5000	,2330	,0715	,0134	,0014	0,0001

gdzie $r = 0, 1, \ldots, n/2$, gdy n jest parzyste, oraz $r = 0, 1, \ldots, (n-1)/2$, gdy n nieparzyste. Bardzo często, na podstawie pewnych dodatkowych informacji, można spodziewać się, że istnieje tylko jednakowa tendencja (znaki zgodne). Wówczas, rozumując analogicznie jak dla jednego tylko rodzaju znaków, otrzymujemy prawdopodobieństwo, że jest $2r$ znaków zgodnych, równe

$$P(U = 2r) = \binom{n}{r}^2 \Big/ \binom{2n}{n}, \quad \text{gdzie} \quad r = 0, 1, \ldots, n.$$

Prawdopodobieństwo, że jest niemniej niż $2r$ znaków zgodnych, wynosi

$$P(U \geqslant 2r) = \Big[\sum_{i=r}^{n} \binom{n}{i}^2\Big] \Big/ \binom{2n}{n}, \quad \text{gdzie} \quad r = 0, 1, \ldots, n. \tag{4}$$

Tablica 1 przedstawia rozkład prawdopodobieństw określonych wzorem (4) dla $2n = 2, 4, 6, \ldots, 30$.

Ustalając dany poziom istotności α, np. $\alpha = 0{,}05$ albo $\alpha = 0{,}01$, można obliczyć, ile co najmniej powinno być znaków zgodnych albo niezgodnych, aby hipotezę o braku tendencji można było odrzucić, czyli przyjąć hipotezę przeciwną — że tendencja istnieje.

Oznaczmy tę najmniejszą ilość znaków przez $2R$. Obliczymy ją jako najmniejszą wartość $2r$ spełniającą nierówność

$$\binom{2n}{n}^{-1} \sum_{i=r}^{n} \binom{n}{i}^2 \leqslant \frac{1}{2}\alpha. \tag{5}$$

W tablicy 2 wypisane są najmniejsze liczby znaków parami zgodnych albo niezgodnych wymagane, by hipotezę o istnieniu tendencji można było przyjąć na poziomach istotności $\alpha \leqslant 0{,}05$ i $\alpha \leqslant 0{,}01$ dla prób liczących $2n = 8, 10, \ldots, 30$ obserwacji.

TABLICA 2

Najmniejsza ilość znaków parami zgodnych lub niezgodnych przy spełnieniu hipotezy o istnieniu tendencji dwóch populacji (test dwustronny)

	$2n$	6	8	10	12	14	16	18	20	22	24	26	28	30
$\alpha \leqslant 0{,}05$	$2R$	(6)	8	10	12 (10)	12	14	16 (14)	16	18 (16)	18	20	20	22
$\alpha \leqslant 0{,}01$	$2R$			10	12	14	16 (14)	16	18	18	20	22 (20)	22	24

Najmniejszą liczbę znaków zgodnych przy spełnieniu hipotezy o jednakowej tendencji otrzymamy jako najmniejszą wartość $2r$ spełniającą nierówność

$$\binom{2n}{n}^{-1} \sum_{i=r}^{n} \binom{n}{i}^2 \leqslant \alpha. \tag{6}$$

Chcąc korzystać z tablicy 2 należałoby poziom istotności podzielić przez 2. Na przykład w próbie liczącej $2n = 20$ elementów powinno być co najmniej 16 znaków zgodnych albo niezgodnych, by można było hipotezę o istnieniu tendencji przyjąć na poziomie istotności $\alpha \leqslant 0{,}05$. W tej samej próbie przy 16 znakach zgodnych przyjmujemy hipotezę o jednakowej tendencji na poziomie istotności $\alpha \leqslant 0{,}025$.

Ponieważ zmienna U jest dyskretna, więc wartości $2R$ dla hipotezy dwustronnej i jednostronnej nie będą się znacznie różniły dla tego samego α. W tablicy 2 w nawiasach podano wartości $2R$ ulegające zmianie przy teście jednostronnym na poziomach $\alpha \leqslant 0{,}05$ i $\alpha = 0{,}01$.

b. Rozważmy teraz przypadek, gdy N jest nieparzyste, tj. gdy $N = 2n+1$. Wówczas mogą być następujące sytuacje:

b1. Mediany Me_x i Me_y przypadają w tej samej miejscowości, np. w j-tej; wtedy $\xi_j = \eta_j = 0$ i zagadnienie sprowadza się do przypadku a, tj. do liczby obserwacji $N = 2n$.

b2. Mediany Me_x i Me_y przypadają w różnych miejscowościach, czyli $\xi_j = 0$, $\eta_l = 0$, gdzie $j \neq l$. Wówczas mogą zachodzić następujące kombinacje znaków:

	ξ_j	η_l			ξ_j	η_l
(I)	+	+		(III)	+	−
(II)	−	−		(IV)	−	+

Wtedy w przypadkach (I) oraz (II) będziemy przyjmowali, że co najmniej dwa znaki są zgodne, a w przypadkach (III) oraz (IV), że co najmniej dwa znaki są niezgodne. I znowu przypadek b2 możemy sprowadzić do przypadku a, to jest gdy $N = 2n$, przy czym dla (I) oraz (II) ilość znaków zgodnych może być równa $2r$, gdzie $r = 0, 1, \ldots, n$, a dla (III) oraz (IV) ilość znaków zgodnych może być równa $2r$, gdzie $r = 1, 2, \ldots, n$, i odwrotnie dla znaków niezgodnych

Wartość oczekiwana i wariancja. Wykażemy najpierw, że wyrażenie

$$P(U = 2r) = \binom{n}{r}^2 \Big/ \binom{2n}{n}$$

jest funkcją rozkładu. Jak wiadomo (z teorii rozkładu hipergeometrycznego),

$$\sum_{x=0}^{m} \binom{b}{x}\binom{c}{m-x} = \binom{b+c}{m}. \tag{7}$$

Sumę $\sum_{i=0}^{n}\binom{n}{i}^2$ można napisać w postaci

$$\sum_{i=0}^{n}\binom{n}{i}^2 = \sum_{i=0}^{n}\binom{n}{i}\binom{n}{n-i} = \binom{2n}{n}, \tag{8}$$

skąd

$$\Big[\sum_{i=0}^{n}\binom{n}{i}^2\Big]\Big/\binom{2n}{n} = 1.$$

Z kolei obliczamy wartość oczekiwaną

$$\mathrm{E}(U) = \Big[\sum_{i=0}^{n} 2i\binom{n}{i}^2\Big]\Big/\binom{2n}{n} = 2\Big[\sum_{i=1}^{n} i\binom{n}{i}\binom{n}{i}\Big]\Big/\binom{2n}{n}.$$

Łatwo zauważyć, że

$$i\binom{n}{i} = n\binom{n-1}{i-1},$$

skąd

$$\mathrm{E}(U) = 2\Big[\sum_{i=1}^{n} n\binom{n-1}{i-1}\binom{n}{i}\Big]\Big/\binom{2n}{n} = 2n\Big[\sum_{j=0}^{n-1}\binom{n-1}{j}\binom{n}{j+1}\Big]\Big/\binom{2n}{n} =$$

$$= 2n\Big[\sum_{j=0}^{n-1}\binom{n-1}{j}\binom{n}{n-1-j}\Big]\Big/\binom{2n}{n}.$$

Stosując odpowiednio wzór (7) otrzymujemy

$$\mathrm{E}(U) = 2n\binom{2n-1}{n-1}\Big/\binom{2n}{n} = \frac{2n\cdot n!\,n!}{2n(2n-1)!}\cdot\frac{(2n-1)!}{(n-1)!\,n!} = n. \tag{9}$$

Wreszcie wyznaczamy wariancję. Najpierw obliczamy wartość oczekiwaną

$$\mathrm{E}(U^2) = \Big[\sum_{i=0}^{n}(2i)^2\binom{n}{i}^2\Big]\Big/\binom{2n}{n} = 4\Big[\sum_{i=1}^{n} i^2\Big(\frac{n!}{i!\,(n-i)!}\Big)^2\Big]\Big/\binom{2n}{n} =$$

$$= 4n^2\Big[\sum_{i=1}^{n}\Big(\frac{(n-1)!}{(i-1)!\,(n-i)!}\Big)^2\Big]\Big/\binom{2n}{n} = 4n^2\Big[\sum_{j=0}^{n-1}\Big(\frac{(n-1)!}{j!\,(n-1-j)!}\Big)^2\Big]\Big/\binom{2n}{n} =$$

$$= 4n^2\Big[\sum_{j=0}^{n-1}\binom{n-1}{j}^2\Big]\Big/\binom{2n}{n}$$

i opierając się na wzorze (8) otrzymujemy

$$\mathrm{E}(U^2) = 4n^2\binom{2n}{n}^{-1}\binom{2(n-1)}{n-1} = \frac{4n^2 nn}{2n(2n-1)} = \frac{2n^3}{2n-1};$$

stąd wariancja

$$D^2(U) = \mathrm{E}(U^2)-\big(\mathrm{E}(U)\big)^2 = 2n^3/(2n-1)-n^2 = n^2/(2n-1). \tag{10}$$

Rozkład graniczny. Zbieżność do rozkładu normalnego.

TWIERDZENIE 1. *Niech funkcja rozkładu zmiennej losowej $U = 2r$ będzie określona wzorem*

$$P_n(2r) = \binom{n}{r}^2\binom{2n}{n}^{-1}. \tag{11}$$

Przyjmijmy

$$x = (2r-n)/\sqrt{\tfrac{1}{2}n}. \tag{12}$$

Jeżeli $n \to \infty$, to

$$\lim_{n\to\infty} \tfrac{1}{2}\sqrt{\tfrac{1}{2}n}\, P_n(2r) = e^{-x^2/2}/\sqrt{2\pi}, \tag{13}$$

jednostajnie dla wszystkich r, dla których x leży w dowolnym, skończonym przedziale $\langle a, b\rangle$.

Dowód. Postępowanie jest analogiczne jak przy dowodzie twierdzenia lokalnego Moivre'a-Laplace'a.

Przy dowodzie skorzystamy ze wzoru Stirlinga

$$k! = \sqrt{2\pi k}\, k^k e^{-k} e^{\Theta_k}, \quad \text{gdzie} \quad |\Theta_k| < 1/12k, \tag{14}$$

oraz z następujących uwag: Z wzoru (12) mamy

$$2r = n+x\sqrt{\tfrac{1}{2}n}, \qquad 2(n-r) = n-x\sqrt{\tfrac{1}{2}n}. \tag{15}$$

Ponieważ x jest ograniczone (z założenia), więc jeżeli $n \to \infty$, to także $r \to \infty$. Rozważmy zatem rozkład

$$P_n(2r) = \binom{n}{r}^2\binom{2n}{n}^{-1} = \frac{(n!)^4}{(r!)^2\big((n-r)!\big)^2(2n)!}$$

i zastosujmy wzór Stirlinga:

$$P_n(2r) = \frac{4\pi^2 n^2 n^{4n} e^{-4n} e^{\Theta_n}}{2\pi r r^{2r} e^{-2r} e^{2\Theta_r} 2\pi(n-r)(n-r)^{2(n-r)} e^{-2(n-r)} e^{2\Theta_{n-r}} \sqrt{4\pi n}\,(2n)^{2n} e^{-2n} e^{\Theta_{2n}}} =$$

$$= \frac{n^{2(2n+1)}}{\sqrt{2\pi}\,(2n)^{2n+1/2}\, r(n-r) r^{2r}(n-r)^{2(n-r)}}\, e^{\Theta},$$

gdzie $\Theta = 4\Theta_n - 2\Theta_r - 2\Theta_{n-r} - \Theta_{2n}$.

Dalej otrzymujemy

$$(16)\qquad P_n(2r) = \frac{n^{3/2}}{\sqrt{2\pi}\sqrt{2r(n-r)}} \cdot \frac{n^{2n}}{2^{2n} r^{2r} (n-r)^{2(n-r)}} \cdot e^{\Theta} =$$

$$= \frac{1}{\sqrt{2\pi}} \left(\frac{n}{2}\right)^{1/2} \cdot \frac{n}{r(n-r)} \left(\frac{n}{2r}\right)^{2r} \left(\frac{n}{2(n-r)}\right)^{2(n-r)} e^{\Theta}.$$

Jak łatwo zauważyć, $\Theta \to 0$, gdy $n \to \infty$, zatem $e^{\Theta} \to 1$.

Oszacujmy

$$\frac{r(n-r)}{n} = \frac{\frac{1}{2}(n+x\sqrt{\frac{1}{2}n})\frac{1}{2}(n-x\sqrt{\frac{1}{2}n})}{n} = \frac{n^2-\frac{1}{2}nx^2}{4n} =$$

$$= \frac{n^2}{4n}\left(1-\frac{x^2}{2n}\right) = \frac{n}{4}\left(1-\frac{x^2}{2n}\right);$$

jeżeli $n \to \infty$, to $(1-x^2/2n) \to 1$ i stąd

$$\left(\frac{n}{2}\right)^{1/2} \frac{n}{r(n-r)} \to \left(\frac{n}{2}\right)^{1/2} \frac{4}{n} = 2\left(\frac{2}{n}\right)^{1/2}.$$

Oszacujmy teraz wielkość

$$\log A = \log\left(\frac{n}{2r}\right)^{2r}\left(\frac{n}{2(n-r)}\right)^{2(n-r)} = -2r\log\frac{2r}{n} - 2(n-r)\log\frac{2(n-r)}{n} =$$

$$= \left(-n-x\sqrt{\frac{n}{2}}\right)\log\left(1+x\sqrt{\frac{1}{2n}}\right) - \left(n-x\sqrt{\frac{n}{2}}\right)\log\left(1-x\sqrt{\frac{1}{2n}}\right).$$

Przy założeniach powyższego twierdzenia, dla dostatecznie wielkich n, wartość $x\sqrt{1/2n}$ może być dowolnie mała, zatem funkcję $\log(1+x\sqrt{1/2n})$ i podobnie $\log(1-x\sqrt{1/2n})$ można rozwinąć w szereg potęgowy. Ograniczając się do pierwszych dwóch wyrazów, po wykonaniu bardzo prostych rachunków otrzymujemy

$$\log A = -\frac{1}{2}x^2 + O\left(\frac{1}{\sqrt{n}}\right);$$

jeżeli $n \to \infty$, to $O(1/\sqrt{n}) \to 0$ i mamy $\log A = -\frac{1}{2}x^2$, skąd $A = e^{-x^2/2}$.

Wstawiając odpowiednio do wzoru (16) otrzymujemy

$$\lim_{n\to\infty} \tfrac{1}{2}(\tfrac{1}{2}n)^{1/2} P_n(2r) = e^{-x^2/2}/\sqrt{2\pi}\,.$$

Stąd już łatwo przejść do twierdzenia granicznego całkowego.

TWIERDZENIE 2. *Jeżeli λ jest dowolną liczbą dodatnią, to prawdopodobieństwo $P\left(U < n+\lambda\sqrt{\tfrac{1}{2}n}\right)$ jest zbieżne jednostajnie ze względu na λ do całki*

$$\frac{1}{\sqrt{2\pi}}\int_0^{\lambda} e^{-x^2/2}\,dx\,.$$

Zagadnienie mocy i efektywności testu. Teoria funkcji mocy dla testów parametrycznych jest dość dokładnie opracowana, a moce testów najczęściej stosowanych, jak np. testu t, F czy testu opartego na współczynniku korelacji, są znane [12].

W przypadku wnioskowania nieparametrycznego sprawa ta jest nieco trudniejsza, bo trudno jest sprecyzować i wyspecyfikować hipotezy alternatywne. Zagadnieniem tym zajmowali się R. A. Fisher, W. Hoeffding, E. L. Lehmann, J. Neyman, C. Stein, A. Wald. Testy oparte na wzorach kombinatorycznych można praktycznie stosować do dużych prób, gdyż na ogół moc ich nie jest wielka [47].

A. Wald w pracach swoich definiuje pojęcie testu zgodnego [46] jako minimum wymagań, żeby test był „optymalny". Sprecyzowaniem warunków na to, żeby test był optymalny, zajmowali się E. L. Lehmann i C. Stein [20], określając testy najmocniejsze i najostrzejsze ze względu na bardzo proste alternatywy, oraz W. Hoeffding [16], rozwiązując to zagadnienie w sposób ogólny w oparciu na funkcji decyzji Walda. Cytowane przykłady dotyczą głównie zagadnienia dwóch prób.

Ciekawe badania i pomysły tzw. asymptotycznej mocy i efektywności testów nieparametrycznych podane są w pracach W. Hoeffdinga [15], A. Stuarta [43] i D. Teichroewa [44].

Na ogół można powiedzieć, że mimo licznych prac o wnioskowaniu nieparametrycznym [40], teoria tego zagadnienia nie jest dokładnie opracowana i dla większości znanych i stosowanych testów nieparametrycznych nie obliczono funkcji mocy i efektywności. Zagadnienie mocy zostało opracowane bardzo dokładnie tylko dla testu znaku Dixona i Mooda ([5], [6], [7]), wprowadza się tam jednak parametr p i sprecyzowanie alternatywy nie nastręcza trudności.

Mocy testu tendencji dotąd nie obliczyłam; jest to celem moich dalszych badań.

PRZYKŁAD 1. Dane są plony w q/ha dwóch gatunków roślin oleistych: gorczycy i kapusty abisyńskiej, w 27 miejscowościach. Zbadać ich tendencję.

TABLICA 3

Gatunek \ Miejscowość	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Gorczyca	1,88	0,67	2,32	1,24	2,85	3,00	3,87	1,84	1,55	0,31	0,65	1,80	**1,85**	1,57
	+	−	+	−	+	+	+	−	−	−	−	−	0	−
Kapusta abisyńska	8,60	1,99	6,50	6,32	4,60	5,03	4,00	0,51	**4,01**	1,24	1,79	2,29	5,90	2,19
	+	−	+	+	+	+	−	−	0	−	−	−	+	−

Gatunek \ Miejscowość	15	16	17	18	19	20	21	22	23	24	25	26	27
Gorczyca	0,71	3,83	0,83	1,06	2,55	2,02	2,80	4,24	0,78	2,89	1,52	2,26	2,48
	−	+	−	−	+	+	+	+	−	+	−	+	+
Kapusta abisyńska	3,88	6,32	3,70	2,66	4,36	3,71	4,26	4,93	2,11	4,49	2,26	3,64	4,70
	−	+	−	−	+	−	+	+	−	+	−	+	+

Otrzymaliśmy $\mathrm{Me}_x = 1{,}85$ i $\mathrm{Me}_y = 4{,}01$.

Ilość znaków zgodnych jest równa $2r = 22$. Ponieważ na podstawie tablicy 2 dla $\alpha \leqslant 0{,}01$ (przy teście jednostronnym) otrzymujemy $2r > 2R$, bo $22 > 20$, więc przyjmujemy hipotezę, że istnieje jednakowa tendencja.

2. Nieparametryczny test niezależności χ^2 Fishera. Przypuśćmy, że dana jest pewna próba losowa $z_1, z_2, \ldots, z_N$ i że jej elementy sklasyfikowano według dwóch lub więcej cech. W szczególności, gdy każda klasyfikacja zawiera dwie klasy, mamy tablicę 2×2 zwaną także tablicą *czteropolową*.

TABLICA 4

Czteropolowa tablica kontyngencji. Klasyfikacja według populacji i według znaków odchyleń od median

B \ A	+	−	Razem
+	a	c	$a+c$
−	b	d	$b+d$
Razem	$a+b$	$c+d$	$a+b+c+d$

Za $z_1, z_2, \ldots, z_N$ można przyjąć próbę dwuwymiarową (ξ_1, η_1), $(\xi_2, \eta_2), \ldots, (\xi_N, \eta_N)$ o liczności $N = 2n$ par, gdzie ξ_i i η_i określone są wzorami (1). Klasyfikację możemy przeprowadzić 1° według populacji (np. niezależnie od odmiany A albo B) i 2° według znaków; każda wartość ξ_i należąca do populacji A może być dodatnia lub ujemna i podobnie każda wartość η_i należąca do populacji B. Innych możliwości

nie ma; w przypadku $N = 2n+1$ zero traktujemy podobnie jak w paragrafie poprzednim i zagadnienie rozważa się dla $N = 2n$. Wówczas otrzymujemy tablicę czeropolową dla próby złożonej z $2n$ elementów (tablica 4, gdzie a, b, c, d oznaczają odpowiednio ilości elementów należących równocześnie do dwóch przecinających się klas, przy czym $a+b+c+d = 2n$).

Interesuje nas zagadnienie, czy ta klasyfikacja jest niezależna, tzn. czy kombinacje znaków $+\,+$, $+\,-$, $-\,+$, $-\,-$ występują jednakowo często, czy też pewne kombinacje występują częściej niż inne. Gdyby się okazało, że występowanie znaków zgodnych jest częstsze niż znaków przeciwnych, to można by przyjąć hipotezę o jednakowej tendencji; odwrotnie, wielka ilość znaków przeciwnych zadecydowałaby o przyjęciu hipotezy o tendencji przeciwnej. W jednym i drugim przypadku otrzymalibyśmy, że klasyfikacja jest zależna. Gdyby jednak występowanie wszystkich czterech kombinacji par znaków było jednakowo częste, wówczas moglibyśmy mówić o tym, że klasyfikacja jest niezależna, czyli że kwestii jednakowej czy różnej tendencji nie można w tym przypadku rozstrzygnąć.

Test Fishera niezależności klasyfikacji tablicy czteropolowej [13], [14], który tu proponujemy do badania tendencji, opiera się na zasadzie proporcjonalności częstości poszczególnych klas. Mianowicie po prostu oblicza się wartość χ^2 z jednym stopniem swobody, stosując wzór

$$\chi^2 = (ad-bc)^2(a+b+c+d)/(a+b)(c+d)(a+c)(b+d) \tag{17}$$

lub

$$\chi^2 = 2(ad-bc)^2/n^3, \tag{17'}$$

przy czym pożądane jest (ale niekonieczne), żeby były zachowane warunki $(a+b)(a+c) \geqslant 10n$ oraz $(c+d)(b+d) \geqslant 10n$ (porównaj pracę [28]). Jeżeli zachodzi nierówność $\chi^2_{\text{obl}} > \chi^2_{\text{tab}}$, gdzie χ^2_{tab} oznacza wartość odczytaną w tablicy rozkładu χ^2 dla wybranego poziomu istotności, np. $\alpha = 0{,}05$, to mówimy, że klasyfikacja nie jest niezależna i w zależności od tego, których znaków, zgodnych czy niezgodnych, jest więcej, mówimy o jednakowej lub niejednakowej tendencji. Jeżeli jest $\chi^2_{\text{obl}} \leqslant \chi^2_{\text{tab}}$, to nie można zaprzeczyć hipotezie o niezależności i wobec tego nie można przyjąć, że tendencja istnieje.

Należy zaznaczyć, że test Fishera jest tylko przybliżony. Istnieje test dokładny dla tablicy czteropolowej [13], wymagający jednak nieco więcej rachunków i tu zajmować się nim nie będziemy.

Zastosujmy test Fishera do zbadania tendencji gorczycy i kapusty abisyńskiej, co już robiliśmy w poprzednim paragrafie.

TABLICA 5

Gorczyca \ Kapusta abis.	+	−	Razem
+	11	2	13
−	2	11	13
Razem	13	13	26

Otrzymujemy

$$\chi^2_{\text{obl}} = 2(121-4)^2/13^3 = 2\cdot 117^2/13^3 = 12{,}462;$$
$$\chi^2_{\text{tab}} = 3{,}841$$

dla liczby stopni swobody $\nu = 1$ i $\alpha = 0{,}05$.

Ponieważ $12{,}462 > 3{,}841$, więc klasyfikacja jest zależna, a ponieważ jest więcej znaków zgodnych niż przeciwnych, więc możemy przyjąć hipotezę o jednakowej tendencji.

Test ten ma jeszcze tę właściwość, że nie trzeba zakładać ciągłości rozkładu populacji generalnej (cechy jakościowe).

III. Badanie interakcji. Podobieństwo obiektów

Za pomocą testów tendencji, stwierdzając zależność, można badać kierunek wahań w obu populacjach; testy te nie określają jednak formy tej zależności, a więc w przypadku tendencji jednakowej nie zawsze nadają się do badania interakcji.

Jednym z takich nieparametrycznych testów, na podstawie których można ocenić nie tylko tendencję, ale także stwierdzić, czy zależność jest liniowa (lub w przybliżeniu liniowa), jest *test asocjacji*, skonstruowany przez P. S. Olmsteada i J. W. Tukey'a [22], [29], zwany także *testem rogu*. Jest on o tyle wygodny, że nie zakłada nic o funkcjach rozkładu populacji generalnych, poza tym, że są ciągłe. Drugą jego zaletą jest to, że należy do tzw. testów *graficznych*, tzn. takich, które dadzą się przedstawić za pomocą wykresu. Pozwala to od razu z rysunku zorientować się, czy zależność liniowa istnieje i niekiedy ocenić „na oko" stopień tej zależności.

Przy korzystaniu z tego testu nie wykonuje się żadnych rachunków, trzeba tylko policzyć pewną ilość punktów.

1. Nieparametryczny test asocjacji. Rozważmy próbę (x_1, y_1), $(x_2, y_2), \ldots, (x_N, y_N)$ reprezentującą N punktów na płaszczyźnie xy. Przyjmujemy, że obie zmienne są ciągłe, tzn. że prawdopodobieństwo dwóch jednakowych obserwacji w każdej próbie jest równe zeru. Niech Me_x oznacza wartość mediany dla x-ów, Me_y zaś dla y-ów. Wyprowadźmy

proste $x = \mathrm{Me}_x$ oraz $y = \mathrm{Me}_y$. Dzielą one płaszczyznę xy na cztery kwadranty albo cztery *rogi*. Oznaczmy te kwadranty znakami +, —, +, — tak, że prawy górny i lewy dolny są dodatnie (rys. 1). Liczbę obserwacji zaopatrzymy w znak + albo — w zależności od tego, w którym kwadrancie te obserwacje liczymy. Gdy N jest parzyste ($N = 2n$), wówczas linie mediany nie przecinają żadnego punktu. Gdy N jest nieparzyste ($N = 2n+1$), wówczas prosta $x = \mathrm{Me}_x$ przechodzi przez punkt (x^*, y_j), a prosta $y = \mathrm{Me}_y$ przez punkt (x_i, y^*), gdzie x^* i y^* są odpowiednimi wartościami mediany. Zastępując te dwa punkty jednym o współrzędnych (x_i, y_j) (rys. 1) otrzymujemy $2n$ punktów, do których można stosować metodę poniżej omówioną. Należy jeszcze wyznaczyć linie przerywane na rysunku 1. Aby wyznaczyć linię przerywaną równoległą do prostej $x = \mathrm{Me}_x$ po stronie lewej, przesuwamy się równolegle do tej prostej, począwszy od lewego brzegu kartki na prawo dopóty, dopóki jakiś punkt spotkany nie znajdzie się w kwadrancie o znaku przeciwnym niż pierwszy punkt spotkany. Podobnie wyznaczamy prostą po prawej stronie posuwając się w lewo oraz dolną posuwając się równolegle do $y = \mathrm{Me}_y$ w górę i górną posuwając się w dół (rys. 1).

Niech r_1 oznacza ilość punktów (z uwzględnieniem znaku) po stronie lewej od linii lewej, r_2 — ilość punktów po stronie prawej od linii prawej, r_3 — poniżej linii dolnej i wreszcie r_4 — powyżej linii górnej.

Niech $r = r_1+r_2+r_3+r_4$ oznacza algebraiczną sumę tych punktów. Jeżeli X i Y są zmiennymi losowymi niezależnymi, to r zależy tylko od uporządkowania i jego rozkład da się wyrazić za pomocą pewnych wzorów kombinatorycznych. P. S. Olmstead i J. W. Tukey [29] wykazali, że dla wielkich N jest

$$\lim_{N\to\infty} P(|r| \geqslant k) = \frac{9k^3+9k^2+168k+208}{216\cdot 2^k}$$

TABLICA 6

Krytyczne wartości $|r|$ dla testu rogu

α	$\lvert r\rvert$
0,10	9
,05	11
,02	13
,01	14-15
,005	15-17
,002	17-19
,001	18-21

Prawdopodobieństwo $P(|r| \geqslant k)$ jest stablicowane dla $2n = 2,4,6,8,10,14$ i ∞ oraz dla $k = 1,2,\ldots$ $\ldots,30$ (por. [29]). Na uwagę zasługuje fakt, że $P(|r| \geqslant k)$ jest bardzo szybko zbieżne i w praktyce okazuje się, że granice na k wyznaczone dla $\alpha = 0{,}05$ i $\alpha = 0{,}01$ przy $N = \infty$ można stosować do prób o liczności większej od 10.

Jeżeli $|r| \geqslant k$ przy dostatecznie małym α, to hipotezę o niezależności obu zmiennych losowych odrzucamy i przyjmujemy hipotezę przeciwną, tj. uważamy, że te zmienne są ze sobą związane. W tablicy 6 podano wartości r dla różnych poziomów istotności, których przekroczenie po-

woduje odrzucenie hipotezy o niezależności. Liczby mniejsze odnoszą się do prób wielkich, większe — do małych. Z tablicy tej widać, że dla prób wielkich (praktycznie biorąc większych od 10) wystarczy, żeby było $|r| \geqslant 11$ dla $\alpha = 0{,}05$, a $|r| \geqslant 14$ dla $\alpha = 0{,}01$, by odrzucić hipotezę o niezależności.

Należy jeszcze zauważyć, że nie wyznaczono funkcji mocy dla tego testu i wobec tego nie jest znane prawdopodobieństwo, jak często popełniamy błąd drugiego rodzaju, tzn. jak często przyjmujemy hipotezę fałszywą.

W zastosowaniu do naszego zagadnienia interakcji niezaprzeczenie hipotezy o niezależności (w praktyce przyjęcie tej hipotezy) może mieć dwojaką interpretację.

Jeżeli obiekty nie wykazują istotnego zróżnicowania wartości badanej cechy w poszczególnych miejscowościach, a wahania na zbiorze wszystkich miejscowości są małe i mają charakter losowy, a więc środowisko jest mało zróżnicowane i nie obserwuje się jego działania głównego na wartość badanej cechy, to trudno jest mówić o jakimś działaniu wzajemnym. Ten przypadek jednak w doświadczeniach wielokrotnych praktycznie biorąc nie zachodzi, zmienność klimatyczno-glebowa jest bowiem zwykle taka wielka, że wpływ ten zawsze występuje.

Pozostaje więc zwykle przypadek drugi, kiedy wpływ środowiska istnieje, jest jednak taki duży i powoduje takie zmiany, że nie można stwierdzić żadnej prawidłowości w jego oddziaływaniu na 2 różne obiekty A i B. Zdanie: „przy dużym zróżnicowaniu cechy w dwóch próbach, są one niezależne" znaczy, że wysokie wartości cechy w jednej próbie mogą odpowiadać bądź niskim, bądź wysokim wartościom tej samej cechy w drugiej próbie. A zatem zwiększenie np. plonu odmiany A nie zawsze pociąga za sobą zwiększenie plonu odmiany B. Nasuwa się tu jednak zagadnienie *ryzyka* uprawy. Gdy np. zachowanie się odmiany A w pewnym rejonie jest znane, wówczas możemy przypuszczać, że wprowadzenie odmiany B jest połączone z pewnym ryzykiem, gdyż nie można przewidzieć, jak się ona zachowa.

Jeżeli jednak za pomocą testu rogu stwierdzimy, że zmienne są zależne, to aby określić, czy interakcja istnieje, należy rozpatrzyć następujące możliwości:

1° Istnieje zależność liniowa, a punkty gromadzą się przeważnie w kwadrancie II i IV (—, —). Oznacza to, że wysokie wartości x mają tendencję do łączenia się z niskimi wartościami y (kwadrant II) i odwrotnie w kwadrancie IV.

Wobec tego można przyjąć w tym przypadku, że interakcja istnieje. Obliczony współczynnik regresji liniowej będzie ujemny.

2° Istnieje zależność liniowa, a punkty gromadzą się przeważnie w kwadrancie I i III (+, +). Wówczas wysokie wartości x mają tendencję do łączenia się z wysokimi wartościami y (kwadrant I) i podobnie w kwadrancie III. Można więc przyjąć, że oba obiekty wykazują tę samą tendencję, nie można jednak powiedzieć, czy zmiany jednego obiektu są podobne do zmian badanej cechy drugiego obiektu. W takim przypadku należy jeszcze zbadać dodatkowo współczynnik regresji liniowej. Tu jednak trzeba już założyć normalność obu populacji generalnych.

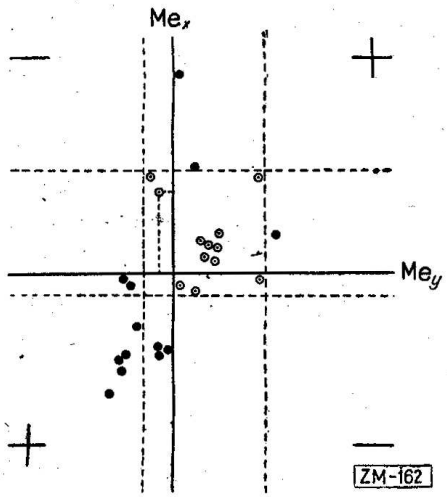
Dla współczynnika regresji liniowej $\beta < 1$ albo $\beta > 1$ zmiany obu zmiennych nie są jednakowe, zatem interakcja istnieje. Dla współczynnika regresji liniowej $\beta = 1$ zmiany są jednakowe i nie ma interakcji. Streszczając możemy sporządzić tablicę 7.

TABLICA 7

Wnioski odpowiadające krytycznym wartościom $|r|$ i β

Poziom istotności	$\|r\|$	Zależność	Współczynnik regresji β	Interakcja	Tendencja
0,05	<11	nie istnieje	0	nie istnieje	nie istnieje
0,01	<14				
0,05	$\geqslant 11$	istnieje	$\leqslant 0$	istnieje	niejednakowa
			< 1 albo > 1	istnieje	jednakowa
0,01	$\geqslant 14$		1	nie istnieje	jednakowa

Test rogu nie odpowiada więc od razu na pytanie, czy interakcja istnieje, dostarcza jednak ważnych informacji dotyczących kierunku przebiegających zmian i przy wstępnych badaniach właściwości dwóch populacji może dać duże usługi, zwłaszcza, że jest prosty w użyciu.



Rys. 1

Zastosujmy go do naszego poprzedniego przykładu i zbadajmy, czy gorczyca i kapusta abisyńska, które wykazały tendencję jednakową, wykazują także reakcję podobną czy też różną w tych miejscowościach. Mamy tu następujące dane: $N = 2n - 1 = 27$. $Me_x = 1{,}85$, $Me_y = 4{,}01$. Rysunek 1 przedstawia diagram punktowy potrzebny do zastosowania testu rogu. Jest $r_1 = 6$, $r_2 = 1$, $r_3 = 9$, $r_4 = 2$, zatem $r = 18$. Dla $\alpha = 0{,}05$, a nawet dla $\alpha = 0{,}01$ widzimy, że zmienne są zależne, ponieważ $18 > 14$. Nie można więc stwierdzić od razu istnienia

interakcji. Obliczmy współczynnik regresji liniowej $b = 0{,}83$. Błąd współczynnika regresji wynosi $s_b = 0{,}31$, skąd $t = |0{,}83-1|/0{,}31 = 0{,}55$. Ponieważ dla liczby stopni swobody $\nu = 25$ i poziomu ufności $\alpha = 0{,}05$ jest $t_{tab} = 2{,}060$, więc możemy przyjąć, że $\beta = 1$ i interakcji nie obserwujemy; możemy uważać, że na ogół reakcja obu gatunków w tych miejscowościach jest podobna. Dokładniejszą analizę otrzymamy stosując metodę regresji Neymana [25], [26].

2. Test χ^2 dla różnic. Dotychczas rozpatrywane testy, za pomocą których wykrywaliśmy bądź tendencję przebiegu zjawiska, bądź kierunek reakcji obiektów na środowisko, należą do typu testów nieparametrycznych i nie wymagają żadnych założeń o populacji generalnej poza tym, że badana cecha jest ciągła; nie dają one jednak dokładnej, ilościowej oceny interakcji, co przecież zarówno w zagadnieniach naukowych, jak i gospodarczych ma często duże znaczenie.

Ponieważ wielkość interakcji jest właściwie uzależniona od wielkości zmian, a więc różnic w poszczególnych miejscowościach między dwoma obiektami, można by zastosować jakiś test oparty na tej różnicy. Ponieważ zaś prawie zawsze można przyjąć, że zjawiska tego typu podlegają prawu rozkładu normalnego, więc proponuję tu pewną modyfikację testu χ^2 w zastosowaniu do badania różnic.

Wyprowadzenie testu. Niech $x_1, x_2, \ldots, x_N$ oznaczają próbę wylosowaną z populacji, w których badana cecha x ma rozkład odpowiednio $N(\mu_1, \sigma_{x_i})$ i podobnie $y_1, y_2, \ldots, y_N$ oznaczają próbę wylosowaną z populacji, w których cecha Y ma rozkład odpowiednio $N(\mu_2, \sigma_{y_i})$.

Przyjmijmy

$$\begin{aligned} \Delta x_i &= x_i - \mu_1, \\ \Delta y_i &= y_i - \mu_2, \end{aligned} \qquad i = 1, 2, \ldots, N. \tag{18}$$

Otrzymujemy odpowiednio dwa ciągi obserwacji

$$\begin{aligned} \Delta x&: \quad \Delta x_1, \Delta x_2, \ldots, \Delta x_N, \\ \Delta y&: \quad \Delta y_1, \Delta y_2, \ldots, \Delta y_N, \end{aligned} \tag{19}$$

przy czym przyjmujemy, że zmienne Δx i Δy są niezależne. Rozważmy zmienną losową

$$d_i = \Delta x_i - \Delta y_i, \qquad i = 1, 2, \ldots, N. \tag{20}$$

a następnie

$$z_i = d_i/\sigma_{d_i}, \qquad i = 1, 2, \ldots, N, \tag{21}$$

gdzie

$$\sigma_{d_i} = \sqrt{\sigma_{x_i}^2 + \sigma_{y_i}^2}. \tag{22}$$

Zmienne losowe z_i $(i = 1, 2, \ldots, N)$ mają jednakowy rozkład $N(0, 1)$, zatem suma

$$\chi^2 = \sum_{i=1}^{N} z_i^2 = \sum_{i=1}^{N} (d_i/\sigma_{d_i})^2$$

ma rozkład χ^2 o N stopniach swobody.

Jeżeli dla danego poziomu istotności α jest $\chi^2_{\text{obl}} > \chi^2_{\text{tab}}$, to znaczy, że nie wszystkie z_i są z tej samej populacji o rozkładzie $N(0, 1)$, czyli że nie wszystkie d_i są przypadkowe, a więc interakcja istnieje. Jeżeli zaś $\chi^2_{\text{obl}} \leqslant \chi^2_{\text{tab}}$, to nie możemy o istnieniu interakcji niczego konkretnego powiedzieć.

W praktyce zwykle wartości $\mu_1, \mu_2, \sigma^2_{x_i}$ i $\sigma^2_{y_i}$ są nieznane, w wielu jednak przypadkach można wyrachować ich oceny nieobciążone i stosować χ^2 jako test przybliżony.

Ocena średnich μ_1, μ_2 oraz wariancji $\sigma^2_{d_i}$. Jak zaznaczyliśmy, test χ^2 do badania interakcji znajduje przede wszystkim zastosowanie w syntezie doświadczeń wieloletnich i wielokrotnych, co pozwala dość dokładnie ocenić μ_1, μ_2 i $\sigma^2_{d_i}$. Pokażemy to najlepiej na przykładzie doświadczeń wielokrotnych z 5 odmianami lnu włóknistego przeprowadzonych w 10 miejscowościach w 1948 r. Każde doświadczenie zostało założone metodą losowanych bloków i zbadane za pomocą analizy wariancji. W tablicy 6 wypisano średnie plony słomy odziarnionej w q/ha oraz średnie kwadraty odchyleń losowych (wariancje resztowe) s_i^2 dla każdego doświadczenia.

Niech wartości w pierwszej kolumnie (średnie dla odmiany A poprzez wszystkie miejscowości) odpowiadają wartościom zmiennej losowej x_A, w drugiej kolumnie — wartościom zmiennej losowej x_B, w trzeciej x_C itd.

Niech $\bar{x}_A$ oznacza *średnią* arytmetyczną dla odmiany A, obliczoną na podstawie doświadczeń ze wszystkich miejscowości i podobnie $\bar{x}_B$, $\bar{x}_C$ itd. Ponieważ x_{Ai} jest średnią dla odmiany A, która w i-tej miejscowości była powtórzona r_i razy, więc średnią $\bar{x}_A$ oblicza się z wzoru

$$\bar{x}_A = \Big(\sum_{i=1}^{N} x_{Ai} r_i\Big) \Big/ \Big(\sum_{i=1}^{N} r_i\Big), \tag{24}$$

gdzie N oznacza ilość miejscowości. Ponieważ w naszym przykładzie jest $r_i = r = 6$ dla $i = 1, 2, \ldots, 10$, więc średnia określona wzorem (24) została obliczona na podstawie 60 obserwacji i można ją uważać za dostatecznie dobre przybliżenie średniej μ_A, tj.

$$\bar{x}_A \approx \mu_A. \tag{25}$$

Podobnie jest dla pozostałych odmian.

Wariancją zmiennej x_{Ai} jest

$$s^2_{x_{Ai}} = \frac{s_i^2}{r_i}, \quad i = 1, 2, \ldots, N. \tag{26}$$

Taka sama jest ocena wariancji s_{xBi}, $s^2_{x_{Ci}}$ itp. Mamy zatem

$$s^2_{x_{Ai}} = s^2_{x_{Bi}} = \ldots = s^2_{x_i} = \frac{s_i^2}{r_i}. \tag{27}$$

Wartości $s^2_{x_i}$ podano w ostatniej kolumnie tablicy 8.

TABLICA 8

Średni plon słomy odziarniowej w q/ha pięciu odmian lnu włóknistego w poszczególnych punktach doświadczalnych w 1948 r.

Rejon	Odmiany / Miejscowości	*A* *LCSD* 210	*B* *Daros* II	*C* *Mathis Edel*	*D* *Liral Crown*	*N* *Con-current*	$\bar{x}_{\text{miejsc.}}$	s_i^2	$s^2_{x_i}$
Poznańskie	1. Swadzim	63,3	53,0	53,8	47,4	49,8	53,5	13,23	2,20
	2. Pętkowo	66,4	56,3	59,6	54,7	67,0	60,8	5,15	0,86
	3. Michorzewo	56,4	40,1	46,1	36,5	45,2	44,9	12,48	2,08
Śląsk Dolny	4. Sternalice	49,2	45,8	48,0	39,3	52,0	46,9	8,67	1,44
	5. Wronów	49,9	45,0	43,4	45,4	48,7	46,5	7,87	1,31
	6. Kotowiecko	41,0	31,4	35,4	21,4	34,5	32,7	3,43	0,57
	7. Swojec	31,0	26,9	27,4	26,1	32,7	28,8	8,48	1,41
Pomorze	8. Radostowo	23,6	22,2	32,3	23,2	30,2	26,3	2,59	0,43
Lubelskie	9. Felin	38,9	29,8	36,1	19,3	36,7	32,2	17,99	3,00
Łódzkie	10. Puczniew	55,7	50,1	54,3	50,1	56,3	53,3	6,93	1,16
	$\bar{x}_{\text{odm.}}$	47,5	40,1	43,6	36,3	45,3	42,6		0,1446 $= s^2_{\bar{x}_{\text{odm.}}}$

W naszym przykładzie jest $r_i = r = 6$ dla wszystkich $i = 1, 2, \ldots, N$. Wariancję s_i^2 obliczono na podstawie doświadczenia o ilości elementów $rk = 6 \cdot 5 = 30$ (gdzie $k = 5$ oznacza ilość odmian). Ocena $s^2_{x_i}$ określona wzorem (27) jest już możliwym do przyjęcia przybliżeniem wariancji $\sigma^2_{x_i}$.

Przyjmijmy

$$\Delta x_{Ai} = \Delta x_{Ai} - \bar{x}_A, \quad \Delta x_{Bi} = \Delta x_{Bi} - \bar{x}_B, \quad i = 1, 2, \ldots, N,$$

i tak dalej.

TABLICA 9

Odchylenia średnich plonów słomy oddziarnionej w q/ha od średnich odmianowych obliczonych na podstawie wszystkich punktów doświadczalnych (wartości Δx_i)

Odmiana / Miejscowość	A	B	C	D	E	Znaki
1	+15,8	+12,9	+10,2	+11,1	+ 4,5	+ + + + +
2	+18,9	+16,2	+16,0	+18,4	+21,7	+ + + + +
3	+ 8,9	+ 0,0	+ 2,5	+ 0,2	− 0,1	+ + + + −
4	+ 1,7	+ 5,7	+ 4,4	+ 3,0	+ 6,7	+ + + + +
5	+ 2,4	+ 4,9	− 0,2	+ 9,1	+ 3,4	+ + − + +
6	− 6,5	− 8,7	− 8,2	−14,9	−10,8	− − − − −
7	−16,5	−13,2	−16,2	−10,2	−12,6	− − − − −
8	−23,9	−17,9	−11,3	−13,1	−15,1	− − − − −
9	− 8,6	−10,3	− 7,5	−17,0	− 8,6	− − − − −
10	+ 8,2	+10,0	+10,7	+13,8	+11,0	+ + + + +

Tablica 9 przedstawia wartości zmiennych losowych Δx_{Ai}, Δx_{Bi}, ... dla $i = 1, 2, \ldots, N$. Łatwo tu zauważyć (stosując test tendencji), że kierunek zmian plonu jest jednakowy dla wszystkich możliwych kombinacji par odmian.

Zgodnie ze wzorem (20) obliczamy wartości d_i dla wszystkich odmian; przedstawiono je w tablicy 10. Ponieważ $s^2_{x_i} = s^2_{y_i}$, więc wariancję $s^2_{d_i}$, a zatem i standardowy błąd s_{d_i} obliczamy łatwo ze wzoru

$$\sigma_{d_i} \approx s_{d_i} = s_{x_i}\sqrt{2}. \tag{28}$$

Także wielkość tę podano w tablicy 10.

TABLICA 10

Wartości d_i

Odmiana / Miejsc.	$A-B$	$A-C$	$A-D$	$A-E$	$B-C$	$B-D$	$B-E$	$C-D$	$C-E$	$D-E$	s_{d_i}
1	+2,9	+ 5,6	+ 4,7	+11,3	+2,7	+1,8	+8,4	−0,9	+5,7	+6,6	2,10
2	+2,7	+ 2,9	+ 0,5	− 2,8	+0,2	−2,2	−5,5	−2,4	−5,7	−3,3	1,31
3	+8,9	+ 6,4	+ 8,7	+ 9,0	−2,5	−0,2	+0,1	+2,3	+2,6	+0,3	2,04
4	−4,0	− 2,7	− 1,3	− 5,0	+1,3	+2,7	−1,0	+1,4	−2,3	−3,7	1,70
5	−2,5	+ 2,6	− 6,7	− 1,0	+5,1	−4,2	+1,5	−9,3	−3,6	+5,7	1,62
6	+2,2	+ 1,7	+ 8,4	+ 4,3	−0,5	+6,2	+2,1	+6,7	+2,6	−4,1	1,07
7	−3,3	− 0,3	− 6,3	− 3,9	+3,0	−3,0	−0,6	−6,0	−3,6	+2,4	1,68
8	−6,0	−12,6	−10,8	− 8,8	−6,6	−4,8	−2,8	+1,8	+3,8	+2,0	0,93
9	+1,7	− 1,1	+ 8,4	+ 0,0	−2,8	+6,7	−1,7	+9,5	+1,1	−8,4	2,45
10	−1,8	− 2,5	− 5,6	− 2,8	−0,7	−3,8	−1,0	−3,1	−0,3	+2,8	1,52

W tablicy 11 podano wartości $z_i = d_i/s_{d_i}$ i wartości χ^2 obliczone dla każdej kombinacji. Ponieważ średnie μ_1 i μ_2 były szacowane, więc zmienna określona wzorem (23) będzie miała rozkład χ^2 o $\nu = N-2$ stopniach swobody.

TABLICA 11

Znormalizowane wartości $z_i = d_i/s_{d_i}$

Odmiana / Miejsc.	$A-B$	$A-C$	$A-D$	$A-E$	$B-C$	$B-D$	$B-E$	$C-D$	$C-E$	$D-E$
1	+1,38	+ 2,67	+ 2,24	+5,38	+1,29	+0,86	+4,00	−0,43	+2,71	+3,14
2	+2,06	+ 2,21	+ 0,38	−2,14	+0,15	−1,68	−4,20	−1,83	−4,35	−2,52
3	+4,36	+ 3,14	+ 4,26	+4,41	−1,23	−0,10	+0,05	+1,13	+1,27	+0,15
4	−2,35	− 1,59	− 0,76	−2,94	+0,76	+1,59	−0,59	+0,82	−1,35	−2,18
5	−1,54	+ 1,60	− 4,14	−0,62	+3,15	−2,59	+0,93	−5,74	−2,22	+3,52
6	+2,06	+ 1,59	+ 7,85	+4,02	−0,47	+5,79	+1,96	+6,26	+2,43	−3,83
7	−1,96	− 0,18	− 3,75	−2,32	+1,79	−1,79	−0,36	−3,57	−2,14	+1,43
8	−6,45	−13,55	−11,61	−9,46	−7,10	−5,16	−3,01	+1,93	+4,09	+2,15
9	+0,69	− 0,45	+ 3,43	+0,00	−1,14	+2,73	−0,69	+3,88	+0,45	−3,43
10	−1,18	− 1,64	− 3,68	−1,84	−0,46	−2,50	−0,66	−2,04	−0,20	+1,84
	84,61	211,13	276,81	176,42	69,04	89,86	48,80	113,31	62,09	69,86

Z tablicy χ^2 odczytujemy, że dla ilości stopni swobody $N-2 = 8$ i $\alpha = 0{,}01$ jest $\chi^2 = 20{,}090$. Łatwo zauważyć, że wszystkie pary wykazują istotną interakcję, a więc reakcja różnych odmian w różnych warunkach środowiska nie jest jednakowa.

W tym zagadnieniu rozpatrywaliśmy odmiany na tle miejscowości; rejonizacja w tym ujęciu miałaby głównie za zadanie znalezienie takich odmian, które zachowują się podobnie we wszystkich N miejscowościach. Można by rozpatrywać także miejscowości na tle odmian, tj. szukać takich miejscowości, w których wszystkie odmiany zachowują się podobnie.

Obliczenie $\sigma^2_{x_i}$ oraz z_i będzie oparte na podobnym rozumowaniu; normalizacja odbywa się nie według wierszy, lecz według kolumn [19].

W zasadzie należałoby w syntezie doświadczeń wielokrotnych przeprowadzić oba typy badań i dopiero wtedy mielibyśmy rejonizację właściwą; należy znaleźć takie odmiany, które reagują podobnie w pewnych określonych miejscowościach.

Gdybyśmy chcieli stosować test χ^2 do doświadczeń jednokrotnych, wówczas wariancja σ^2_i byłaby rzeczywiście nieznana i na podstawie materiału obserwacyjnego nie można byłoby jej ocenić. Wtedy należy skorzystać — o ile to możliwe — z danych doświadczeń ścisłych

przeprowadzonych w tych samych miejscowościach albo w miejscowościach bardzo blisko położonych, o podobnych warunkach klimatyczno-glebowych.

Można by tu pójść jeszcze dalej i uogólnić zagadnienie w następujący sposób.

Na podstawie dużego materiału obserwacyjnego (liczącego setki, a nawet tysiące pomiarów), wziętego np. po prostu z roczników statystycznych, ocenić dokładnie wariancję σ_i^2 dla danej odmiany czy gatunku. Ta właściwa wariancja σ_i^2 może być potem używana w wielu zagadnieniach przy opracowywaniu doświadczeń jednokrotnych w całej Polsce lub w pewnych rejonach. Warto przy okazji zaznaczyć, że prace te rozwiązałyby równocześnie w pewnym sensie także zagadnienie liczności próby reprezentatywnej dla poszczególnych gatunków, co jest bardzo ważne dla całokształtu badań związanych z planowaniem doświadczeń.

3. Podobieństwo obiektów. Test *u*. Test χ^2 dla różnic, określony wzorem (23) albo (24), pozwala stwierdzić istnienie interakcji między dwoma obiektami A i B, przy czym nic nie zakłada o kierunku tych różnic. Weźmy np. odmiany A i B w miejscowościach 5 i 6. Z tablicy 9 widzimy, że w miejscowości 5 obie odmiany wykazują zwyżkę plonu, a w miejscowości 6 obie odmiany zniżkę w stosunku do odpowiednich średnich. Jednak zwyżka odmiany B jest większa i podobnie zniżka.

Gdybyśmy np. w miejscowości 5 przestawili liczby, tzn. w kolumnie A napisali +4,9, a w kolumnie B napisali +2,4, nie wpłynęłoby to na wartość χ^2, natomiast zadecydowałoby o tym, że teraz zwyżka dla odmiany A byłaby większa niż dla odmiany B. Tak więc test χ^2 wykrywa istnienie interakcji, ale nie mierzy jej dostatecznie, to znaczy nie informuje nas o jej kierunku. Dlatego też potrzebne są dodatkowe badania, które by tej informacji udzieliły. Wprowadźmy najpierw pewną definicję *podobieństwa obiektów*.

Definicja. Dwa obiekty biologiczne nazywać będziemy *podobnymi* co do tej samej cechy, jeżeli średnie wartości tej cechy oraz ich reakcja (*zmiana* wartości badanej cechy) na zmianę środowiska są jednakowe.

Definicja ta ma charakter bardziej biologiczny niż matematyczny, chodzi tu bowiem o organizmy żyjące w pewnych warunkach, a nie o twory abstrakcyjne.

W zależności od tego, którą cechę badamy i za pomocą jakich testów, można by tę definicję sformułować w postaci pewnych warunków czy funkcji matematycznych. W podanej formie jednak jest ona dostatecznie jasna i w zależności od stosowanego aparatu matematycznego można z łatwością każdorazowo te warunki określić.

Dlatego oprócz testu χ^2 dla różnic należy jeszcze zastosować zwykły test u oparty na założeniu normalności zmiennych. Ponieważ suma niezależnych zmiennych normalnych o rozkładach $N(\mu_A, \sigma_{x_i})$ ma rozkład normalny $N\left(N\mu_1, \sqrt{\sum_{i=1}^{N} \sigma_{x_i}^2}\right)$, więc rozkład średniej $\bar{x}_A$ jest postaci $N\left(\mu_A, \sqrt{\sum_{i=1}^{N} \sigma_{x_i}^2}/N\right)$. Podobnie rozkład średniej $\bar{x}_B$ jest $N\left(\mu_B, \sqrt{\sum_{i=1}^{N} \sigma_{x_i}^2}/N\right)$. Rozkład różnicy $\bar{x}_A - \bar{x}_B = D$ jest też normalny: $N\left(\mu_A - \mu_B, \sqrt{2\sum_{i=1}^{N} \sigma_{x_i}^2}/N\right)$.

Przyjmijmy

$$\sqrt{2\sum_{i=1}^{N} \sigma_{x_i}^2}/N = \sigma_D.$$

Stąd $N(\mu_A - \mu_B, \sigma_D)$ jest rozkładem różnicy średnich odmianowych D. Zamiast σ_D bierzemy jej przybliżenie s_D obliczone z wzoru

$$\sigma_D = s_D = \sqrt{2\sum_{i=1}^{N} s_{x_i}^2}/N$$

oraz stosujemy asymptotyczny test u (dla dużych prób). Przyjmijmy $u = |D|/s_D$. Jeżeli $|D| > u_0 s_D$, to różnica średnich odmianowych jest istotna; jeżeli $|D| \leqslant u_0 s_D$, to jest ona nieudowodniona. W przypadku badania interakcyj miejscowości na tle odmian, uzyskuje się wzory analogiczne [19]. W naszym przykładzie jest

$$s_D = \tfrac{1}{10}\sqrt{2 \cdot 14{,}46} = \sqrt{0{,}2892} = 0{,}52 \text{ q/ha}.$$

Dla $\alpha = 0{,}05$ jest $u_0 = 1{,}96$, zatem $u_0 s_D = 1{,}02$ q/ha. Każda para odmian wykazuje istotną różnicę średnich.

Wnioskowanie na podstawie testu u i testu χ^2 może pogłębić nasze wiadomości o zachowaniu się dwóch obiektów, mianowicie:

1. Jeżeli u jest nieistotne i χ^2 nieistotne, to obydwa obiekty nie różnią się ani plonem średnim, ani reakcją i można przyjąć, że obiekty są podobne.

2. Jeśli u jest nieistotne, a χ^2 istotne — efekt przeciętny obu odmian jest jednakowy, jednak reakcja poszczególnych obiektów jest różna; gdyby więc zależało nam na tym, aby efekt przeciętny plonów dla pewnego rejonu był jednakowy, a nic nie zakładalibyśmy o reakcji, można by te obiekty zaklasyfikować jako równoważne dla całego rejonu.

3. Jeśli u jest istotne oraz χ^2 istotne — obiekty różnią się zarówno przeciętnym efektem, jak i reakcją, nie są więc podobne albo — jak powiemy — są biologicznie odległe; kierunek wahań jest na ogół jednakowy.

4. Wreszcie jeśli u jest istotne, a χ^2 nieistotne, to reakcja obu obiektów jest podobna, tzn. dają one w tych samych miejscowościach mniej więcej jednakowe zwyżki lub zniżki plonu, jednak pod względem średniego efektu jeden obiekt jest lepszy w porównaniu z drugim.

W tablicy 12 opisano oba testy i wynikającą z nich ocenę podobieństwa dwóch obiektów.

W naszym przykładzie wszystkie pary obiektów wykazują zarówno różne efekty przeciętne, jak i reakcje; można by wobec tego starać się wyodrębnić rejony mniejsze (np. geograficzne) i tam badać zachowanie się odmian. Ponieważ ilość miejscowości jest tu jednak zbyt mała, więc te badania nie dałyby w tym przypadku ciekawych wyników.

Poza tym należałoby zbadać jeszcze w podobny sposób miejscowości na tle odmian i dopiero próbować przeprowadzić rejonizację.

Oczywiście w przypadku tak małej liczby punktów sama rejonizacja jest dość problematyczna; wybrany przykład odgrywa raczej rolę pewnego modelu.

TABLICA 12

Ocena podobieństwa obiektów na podstawie testów

L. p.	u	χ^2	Obiekty
1	$\leqslant u_0$	$\leqslant \chi_0^2$	podobne
2	$\leqslant u_0$	$> \chi_0^2$	efekt przeciętny jednakowy, reakcja różna
3	$> u_0$	$> \chi_0^2$	odległe
4	$> u_0$	$\leqslant \chi_0^2$	efekt przeciętny różny, reakcja jednakowa

TABLICA 13

Spis wyników dotyczących istnienia tendencji, interakcji i zróżnicowania średnich

	A	*B*	*C*	*D*	*E*
A		* * *	* * *	* * *	* * *
B			* * *	* * *	* * *
C				* * *	* * *
D					* * *
E					

IV. Wnioski

Po przeprowadzeniu analizy całego materiału statystycznego przez zastosowanie odpowiednich testów, należy zestawić wyniki i sformułować pewne wnioski dotyczące badanego zagadnienia.

Oznaczamy przez * wyniki: że istnieje jednakowa tendencja, że istnieje interakcja, że istnieje zróżnicowanie średnich obiektowych, a przez 0, że te zjawiska nie zachodzą. Ustawienie znaków * i 0 w sposób podobny do tablicy 13 obrazuje od razu otrzymane wyniki.

W naszym przykładzie tablica jest o tyle nieciekawa, że zawiera same znaki *, co świadczy o dużym zróżnicowaniu odmian zarówno pod względem średniego plonu, jak i ich reakcji na środowisko. Należy jednak

zauważyć, że tendencja jest jednakowa — na ogół w lepszych warunkach wszystkie odmiany podnoszą swój plon, a w gorszych obniżają, jednak te zwyżki i zniżki nie są jednakowe; jedne odmiany są bardziej stateczne w swej reakcji, inne wrażliwsze na zmianę warunków.

Należy zatem szukać grup, które by przedstawiały materiał bardziej jednorodny i do którego można by zastosować postulat rejonizacji.

W naszym przykładzie wszystkie różnice średnich odmianowych wziętych parami były istotne. Różnica graniczna wynosiła 1,02 q/ha. Gdybyśmy jednak rozpatrywali nie pary, lecz pewne grupy po 3, 4 i więcej odmian, wówczas wartość u_0 obliczona z rozkładu normalnego nie nadawałaby się, gdyż tablice wartości granicznych są ułożone dla pojedynczych zdarzeń. Stosując bardzo proste rozumowanie (zob. [27]) można obliczyć prawdopodobieństwo, że k wartości pada w przedział wahań dopuszczalnych. Odpowiednie tablice zostały obliczone przez M. Olekiewicza [27]. Jest jednak pewna trudność przy ich stosowaniu, ponieważ kombinacje par dwóch średnich nie są niezależne, a tablice obliczono dla prób niezależnych. Bardziej przydatne okazały się tablice opracowane przez D. B. Duncana [8] i w nieco zmodyfikowanej postaci podane przeze mnie w tablicy 14. Zastosowanie tej tablicy ma szczególne znaczenie przy określaniu granicznej różnicy zaliczenia kilku obiektów do tej samej grupy jednorodnej. Niech s_D oznacza standardowy błąd różnicy. Najkrótszy przedział, który określa, że k obiektów można zaliczyć do tej samej grupy, tzn. że nie obserwujemy istotnego zróżnicowania średnich, obliczamy mnożąc s_D przez odpowiednią wartość t z tablicy 14 dla danej ilości stopni swobody ν i dla danego k.

W naszym przykładzie różnica graniczna dla pojedynczych zdarzeń wynosi 1,02 q/ha. Gdybyśmy chcieli stwierdzić, czy wszystkie 5 odmian różnią się między sobą istotnie *en bloc*, wówczas należałoby wziąć nie $t = 1{,}96$, lecz $t = 2{,}18$ (dla $\nu = \infty$ i $k = 5$). W tym przykładzie zróżnicowanie *en bloc* istnieje. Można dalej próbować zaliczyć 4, 3 itd. odmiany do jednej grupy, posługując się tablicą 14.

W syntezach doświadczeń, zwłaszcza w zagadnieniach hodowlanych zachodzi często pytanie, która z odmian jest najlepsza albo która z odmian jest lepsza od odmiany kontrolnej, np. pewnej odmiany miejscowej czy też odmiany o znanych własnościach biologicznych. Bardzo ciekawe uwagi o grupowaniu i porównywania obiektów można znaleźć w pracach F. Mostellera [23], E. Paulsona [30], [31] i w ostatnich pracach R. E. Bechhofera [1].

W pracy niniejszej rozpatrywaliśmy przeważnie przykłady doświadczeń przeprowadzonych w kilku miejscowościach w jednym roku, aby nie zaciemniać rozważań. Podobną analizę można by przeprowadzić w jednej miejscowości w ciągu kilku lat.

TABLICA 14

Wielokrotny test t

ν \ k	2	3	4	5	6	7	8	9	10	12	14	16	18	20	50	100
1	12,71	12,71	12,71	12,71	12,71	12,71	12,71	12,71	12,71	12,71	12,71	12,71	12,71	12,71	12,71	12,71
2	4,30	4,30	4,30	4,30	4,30	4,30	4,30	4,30	4,30	4,30	4,30	4,30	4,30	4,30	4,30	4,30
3	3,18	3,18	3,18	3,18	3,18	3,18	3,18	3,18	3,18	3,18	3,18	3,18	3,18	3,18	3,18	3,18
4	2,78	2,83	2,84	2,84	2,84	2,84	2,84	2,84	2,84	2,84	2,84	2,84	2,84	2,84	2,84	2,84
5	2,57	2,64	2,68	2,71	2,71	2,71	2,71	2,71	2,71	2,71	2,71	2,71	2,71	2,71	2,71	2,71
6	2,45	2,53	2,57	2,60	2,60	2,60	2,60	2,60	2,60	2,60	2,60	2,60	2,60	2,60	2,60	2,60
7	2,37	2,45	2,50	2,53	2,54	2,55	2,55	2,55	2,55	2,55	2,55	2,55	2,55	2,55	2,55	2,55
8	2,31	2,40	2,45	2,49	2,51	2,52	2,52	2,52	2,52	2,52	2,52	2,52	2,52	2,52	2,52	2,52
9	2,26	2,36	2,41	2,45	2,47	2,49	2,49	2,49	2,49	2,49	2,49	2,49	2,49	2,49	2,49	2,49
10	2,23	2,35	2,38	2,42	2,45	2,45	2,45	2,45	2,45	2,45	2,45	2,45	2,45	2,46	2,46	2,46
11	2,20	2,31	2,37	2,40	2,42	2,43	2,44	2,45	2,45	2,45	2,45	2,45	2,45	2,46	2,46	2,46
12	2,18	2,28	2,35	2,38	2,40	2,42	2,43	2,43	2,45	2,45	2,45	2,45	2,45	2,46	2,46	2,46
13	2,16	2,27	2,33	2,37	2,39	2,41	2,42	2,43	2,44	2,44	2,45	2,45	2,45	2,45	2,45	2,45
14	2,14	2,25	2,31	2,35	2,38	2,40	2,41	2,42	2,43	2,44	2,45	2,45	2,45	2,45	2,45	2,45
15	2,13	2,23	2,30	2,34	2,38	2,39	2,40	2,42	2,43	2,43	2,44	2,45	2,45	2,45	2,45	2,45
16	2,12	2,23	2,28	2,33	2,36	2,38	2,40	2,41	2,43	2,43	2,44	2,45	2,45	2,45	2,45	2,45
17	2,11	2,21	2,28	2,32	2,35	2,38	2,39	2,40	2,42	2,43	2,44	2,45	2,45	2,45	2,45	2,45
18	2,10	2,21	2,27	2,31	2,35	2,37	2,38	2,40	2,41	2,43	2,44	2,45	2,45	2,45	2,45	2,45
19	2,09	2,20	2,26	2,31	2,34	2,37	2,38	2,40	2,41	2,43	2,43	2,45	2,45	2,45	2,45	2,45
20	2,09	2,19	2,25	2,30	2,34	2,36	2,38	2,39	2,40	2,43	2,43	2,45	2,45	2,45	2,45	2,45
22	2,07	2,18	2,24	2,29	2,33	2,35	2,37	2,38	2,40	2,42	2,43	2,44	2,45	2,45	2,45	2,45
24	2,06	2,17	2,23	2,28	2,32	2,34	2,36	2,38	2,39	2,41	2,43	2,44	2,45	2,45	2,45	2,45
26	2,06	2,16	2,22	2,27	2.31	2,33	2,36	2,38	2,39	2,41	2,43	2,44	2,45	2,45	2,45	2,45
28	2,05	2,15	2,21	2,26	2,31	2,33	2,35	2,37	2,38	2,40	2,43	2,44	2,45	2,45	2,45	2,45
30	2,04	2,15	2,21	2,26	2,30	2,33	2,35	2,37	2,38	2,40	2,43	2,43	2,45	2,45	2,45	2,45
40	2,02	2,13	2,19	2,24	2,28	2,31	2,33	2,35	2,37	2,40	2,42	2,43	2,45	2,45	2,45	2,45
60	2,00	2,11	2,18	2,22	2,26	2,29	2,32	2,34	2,35	2,38	2,40	2,43	2,44	2,45	2,46	2,46
100	1,98	2,09	2,16	2,21	2,25	2,28	2,31	2,33	2,35	2,37	2,40	2,42	2,44	2,45	2,50	2,50
∞	1,96	2,06	2,14	2,18	2,23	2,26	2,31	2,31	2,33	2,36	2,39	2,41	2,43	2,45	2,55	2,60

Uwzględnienie wszystkich tendencji, zróżnicowania plonu i reakcji na środowisko w przeciągu kilku czy kilkunastu lat na zbiorze tych samych miejscowości, może stanowić dopiero podstawę do właściwej rejonizacji.

Dziękuję serdecznie prof. dr M. Olekiewiczowi za wiele cennych uwag i wskazówek.

Prace cytowane

[1] R. E. Bechhofer, *A single-sample multiple decision procedure for ranking means of normal populations withs known variances*, Ann. Math. Stat. 25 (1954), str. 273-289.

[2] B. H. Camp, *Some recent advances in mathematical statistics*, I, Ann. Math. Stat. 13 (1942), str. 62-73.

[3] W. G. Cochran, *Problems arising in the analysis of series of similar experiments*, Jour. Roy. Stat. Soc. 4 (1937), str. 102-118.

[4] W. G. Cochran and G. M. Cox, *Experimental designs*, New York, London, 1950.

[5] W. J. Dixon and A. M. Mood, *The statistical sign test*, Amer. Stat. Ass. 41 (1946), str. 557-566.

[6] W. J. Dixon, *Power functions of the sing test and power efficiency for normal alternatives*, Ann. Math. Stat. 24 (1953), str. 467-473.

[7] W. J. Dixon, *Power under normality of several nonparametric tests*, Ann. Math. Stat. 25 (1954), str. 610-614.

[8] D. B. Duncan, *Multiple range and multiple F tests*, Biometrics 11 (1955), str. 1-42.

[9] R. Elandt, *Zastosowania analizy wariancyjnej do złożonych zagadnień sadowniczych*, Rocz. Nauk Roln. 66-A-2 (1953), str. 107-129.

[10] — *Metody syntezy statystycznej doświadczeń jednopowtórzeniowych*, Roczn. Nauk. Roln. 70-A-3 (1955), str. 443-462.

[11] — *On some non-parametric test of the tendency* (nieopublikowane).

[12] Ch. D. Ferris, C. F. Grubbs and Ch. L. Weaver, *Operating characteristics for the common statistical test of significance*, Ann. Math. Stat. 17 (1946), str. 178-197.

[13] R. A. Fisher, *Statistical methods for research workers*, London 1947.

[14] — *On the interpretation of Chi-Square from contingency tables, and the calculation of P*, Journal of the Royal Statistical Society 85 (1922), str. 87-94.

[15] W. Hoeffding, *The large-sample power of tests based on permutations of observations*, Ann. Math. Stat. 23 (1952), str. 169-192.

[16] — "*Optimum*" *nonparametric tests*, Proceedings of the second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley and Los Angelos 1951.

[17] O. Kempthorne, *The design and analysis of experiments*, New York, London 1952.

[18] M. G. Kendall, *The advanced theory of statistics*, London 1947.

[19] M. Kurhański, *Charakterystyka odmian lnu pod względem technologicznym*, Prace Inst. Przem. Wł. Łyk. I.1 (1953), str. 3-20.

[20] E. L. Lehmann and C. Stein, *On the theory of some non-parametric hypotheses*, Ann. Math. Stat. 20 (1949), str. 28-45.

[21] Z. Łubkowski, *Jedna z metod syntetycznego opracowania wyników doświadczeń*, Roczn. Nauk. Roln. 69-A-2 (1954), str. 267-277.

[22] A. Mood, *The introduction to the theory of statistics*, New York, Toronto, London 1950.

[23] F. Mosteller, *A k-sample slippage test for an extreme population*, Ann. Math. Stat. 19 (1949), str. 58-65.

[24] A. Mudra, *Einführung in die Methodik der Feldversuche*, Leipzig 1952.

[25] J. Neyman, *O metodach opracowywania doświadczeń wielokrotnych*, Roczn. Nauk. Roln. i Leśn. 28 (1932), str. 155-210.

[26] — *O pewnych twierdzeniach z rachunku prawdopodobieństwa, które służą za podstawę do rozwiązywania szeregu zagadnień doświadczalnictwa rolniczego*, Rocz. Nauk Roln. i Leśn. 31 (1933), str. 223-276.

[27] M. Olekiewicz, *Tables of significance limits for the largest critical ratio out of k ratios*, Ann. Univ. M. Curie-Skłodowska, sectio A, IV. 11 (1950), str. 115-121

[28] — *Badania prawidłowości w biologii i agrobiologii* (w druku).

[29] P. S. Olmstead and J. W. Tukey, *A corner test for association*, Ann. Math. Stat. 18 (1947), str. 495-513.

[30] E. Paulson, *A multiple decision procedure for certain problems in the analysis of variance*, Ann. Math. Stat. 20 (1949), str. 95-98.

[31] — *On the comparison of several experimental categories with a control*, Ann. Math. Stat. 23 (1952), str. 239-246.

[32] J. Przyborowski i Z. Nawrocki, *Doświadczenia z odmianami pszenicy jarej przeprowadzone w Polsce w latach 1936-38*, Prace Naukowe Rolnicze, Dodatek 8 do Przeglądu Doświadczalnictwa Rolniczego, Warszawa 1939.

[33] J. Przyborowski i H. Wileński, *Analiza zmienności wyników doświadczeń wielokrotnych*, Dodatek 20 do XLIX tomu Roczn. Nauk Roln. i Leśn., Kraków 1938.

[34] N. H. Quenouille, *The design and analysis of experiments*, London 1953.

[35] C. R. Rao, *Advanced statistical methods in biometric research*, New York, London 1952.

[36] T. Ruebenbauer, *Doświadczenia z odmianami pszenicy ozimej przeprowadzone w Polsce w latach 1923-1936*, Kraków 1937.

[37] — *Wyniki doświadczeń z odmianami pszenicy ozimej przeprowadzonych w Polsce w latach 1946-1949*, Warszawa 1952.

[38] — *Wyniki doświadczeń z odmianami pszenicy jarej wykonanych w Polsce w latach 1946-1949*, Warszawa 1952.

[39] — *Wyniki doświadczeń z odmianami owsa wykonanych w Polsce w latach 1946-1949*, Warszawa 1952.

[40] I. R. Savage, *Bibliography on nonparametric statistics and related topics*, Jour. Amer. Stat. Ass. 47 (1953), str. 844-906.

[41] H. Scheffé, *Statistical inference in the non-parametric case*, 14 (1943), str. 305-332.

[42] G. W. Snedecor, *Statistical methods*, Ames, Iowa 1946.

[43] A. Stuart, *The asymptotic relative efficiences of distribution-free tests of randomness against normal alternatives*, Journ. Amer. Stat. Ass. 49 (1954), str. 147-157.

[44] D. Teichroew, *Empirical power functions for non-parametric two-sample tests for small samples*, Ann. Math. Stat. 26 (1955), str. 340-344.

[45] A. Wald, *On the analysis of variance in case of multiple classifications with unequal class frequencies*, Ann. Math. Stat. 12 (1940), str. 96-100.

[46] — and J. Wolfowitz, *On the test, whether two samples are from the same population*, Ann. Math. Stat. 11 (1940), str. 147-162.

[47] — — *Statistical tests based on permutations of the observations*, Ann. Math. Stat. 15 (1944), str. 358-372.

[48] S. S. Wilks, *Order statistics*, Bull. Amer. Math. Soc. 54 (1948), str. 6-50.

[49] J. Wolfowitz, *Non-parametric statistical inference*, Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, I, Berkeley and Los Angelos 1951.

ZAKŁAD DOŚWIADCZALNICTWA ROLNICZEGO I BIOMETRII przy KATEDRZE HODOWLI ROŚLIN I NASIENNICTWA WYŻSZEJ SZKOŁY ROLNICZEJ W POZNANIU

Praca wpłynęła 25. 1. 1956

Р. ЭЛЯНДТ (Познань)

О НЕКОТОРЫХ КРИТЕРИЯХ ИНТЕРАКЦИИ В МНОГОЛЕТНИХ И МНОГОКРАТНЫХ ОПЫТАХ. ВОПРОС РАЙОНИЗАЦИИ

РЕЗЮМЕ

В статье рассмотрен метод статистического синтеза многолетних и многократных опытов, в особенности определение и оценка объектно-локальной и объектно-периодической интеракции двух объектов. Прилагаемые критерия являются или видоизменением некоторых непараметрических критериев, как напр. критерия χ^2 и критерия ассоциации Ольмстеада и Тюки [29], или идеей автора.

Автору принадлежат два критерия: непараметрический критерий тенденции и критерий χ^2 для разностей.

1. Непараметрический критерий тенденции. Пусть $(x_1, y_1), \ldots, (x_N, y_N)$ представляет N-элементную выборку из двумерной совокупности. Положим, что $N = 2n$. Пусть

$$\xi_i = x_i - \mathrm{Me}_x, \quad \eta_i = y_i - \mathrm{Me}_y \quad (i = 1, 2, \ldots, N),$$

причем Me_x и Me_y обозначают медианы.

Если пары (ξ_i, η_i) для $i = 1, 2, \ldots, N$ одного знака (или разных знаков), тогда скажем, что у случайных величин X и Y равная (или разная) тенденция.

Проверяем нулевую гипотезу H_0: случайные величины X и Y независимы, при альтернативной гипотезе H: случайные величины X и Y имеют тенденцию.

Чтобы сконструировать критерий, авторка вводит статистику U, определенную как число знаков сходных (или несходных) парами, и вычисляет вероятность $\mathrm{P}(U \geqslant 2r)$ (при предположении, что имеет место гипотеза H_0). Эту вероятность определяет по формуле

$$\mathrm{P}(U \geqslant 2r) = \left(\sum_{i=r}^{n} \binom{n}{i}^2\right) \Big/ \binom{2n}{n}.$$

Таблица 2 содержит критическое количество знаков, сходных или несходных парами, для критерия тенденции (двустороннего или одностороннего) на уровне значимости $\alpha \leqslant 0{,}05$ и $\alpha \leqslant 0{,}01$.

Случай $N = 2n+1$ посредством простого преобразования сводится к случаю $N = 2n$.

Вычислено тоже математическое ожидание и дисперсию случайной величины U, определенных по формулам

$$\mathrm{E}(U) = n, \quad D^2(U) = n^2/(2n-1).$$

Доказано также локальную предельную теорему, т. е. что

$$\lim_{n\to\infty} \tfrac{1}{2}\sqrt{\tfrac{1}{2}n}P_n(2r) = e^{-x^2/2}/\sqrt{2\pi} \quad \text{при} \quad x = (2r-n)/\sqrt{\tfrac{1}{2}n}.$$

2. Критерий χ^2 для разностей. Пусть $x_1, \ldots, x_N$ представляют N-элементную выборку из совокупностей, исследуемый признак X которых имеет распределение $N(\mu_1, \sigma_{x_i})$ и, подобным образом, $y_1, \ldots, y_N$ представляют выборку из совокупностей, признак Y которых имеет распределение $N(\mu_2, \sigma_{y_i})$.

Статистика

$$\chi^2 = \sum_{i=1}^{N} \frac{[(x_i-y_i)-(\mu_1-\mu_2)]}{\sigma_{x_i}^2+\sigma_{y_i}^2}$$

имеет распределение χ^2, у которого N степени свободы, когда известны параметры μ_1 и μ_2, и $N-2$ степени свободы, когда эти параметры оцениваются.

Критерий χ^2 для разностей употребляют при определении интеракции и вместе с критерием u (основанным на нормальном распределении) составляет основание биологической оценки сходности объектов.

Только совместное толкование интеракции, тенденции и разностей средних приводит к выводам. Многократный критерий Ф. Дункана [8] находит здесь применение.

R. ELANDT (Poznań)

ON CERTAIN INTERACTION TESTS IN SERIAL EXPERIMENTS THE PROBLEM OF STRATIFICATION

SUMMARY

The authoress deals with the problem of methods of the statistical synthesis of serial experiments, and in particular with the definition and evaluation of treatment-places and treatment-seasons interaction of two treatments. The tests given are either modifications of certain (non-parametric) tests, such as the χ^2-test for a four-field contingency table [14] and the test for association of Olmstead and Tukey [29], or the authorress' original ideas.

The authoress gives two tests of her own: a non-parametric test of tendency and the χ^2-test for differences.

1. The non-parametric test of tendency. Let $(x_1, y_1), \ldots, (x_N, y_N)$ be an N-element sample drawn out at random from a two-dimensional population. Suppose $N = 2n$. Let

$$\xi_i = x_i - \mathrm{Me}_x, \quad \eta_i = y_i - \mathrm{Me}_y \quad \text{for} \quad i = 1, 2, \ldots, N$$

where Me_x and Me_y denote the respective medians.

If the signs of the pairs (ξ_i, η_i) for $i = 1, 2, \ldots, N$ are the same (or different), then it is asserted that the random variables X and Y show the same (or different) tendency.

We verify the null-hypothesis H_0: the random variables X and Y are independent, against the alternative hypothesis H: the random variables X and Y show a tendency.

In order to construct the test the authoress introduces a statistic U, which is defined as the number of signs which are the same (or different) in pairs, and she calculates the probability $\mathrm{P}(U \geqslant 2r)$ (under the assumption that H_0 is satisfied). This probability is defined by the formula

$$\mathrm{P}(U \geqslant 2r) = \Big(\sum_{i=r}^{n} \binom{n}{i}^2\Big) \Big/ \binom{2n}{n}.$$

Table 2 gives the critical values of the signs which are the same (or different) in pairs for the (one-tail and two-tail) test of tendency on the level of significance $\alpha = 0{,}05$ and $\alpha = 0{,}01$.

The case $N = 2n+1$ is reducible by a simple modification to the case $N = 2n$.

The expected value and the variance of the random variable U have also been calculated. They are defined, respectively, by the formulas

$$\mathrm{E}(U) = n, \qquad D^2(U) = n^2/(2n-1).$$

The authoress proves also the local limit theorem, showing that

$$\lim_{n\to\infty} \tfrac{1}{2}\sqrt{\tfrac{1}{2}n}\,P_n(2r) = e^{-x^2/2}/\sqrt{2\pi} \quad \text{where} \quad x = (2r-n)/\sqrt{\tfrac{1}{2}n}.$$

The χ^2-test for differences. Let $x_1, \ldots, x_N$ be an N-element sample drawn at random from normal populations $N(\mu_1, \sigma_{x_i})$, and similarly let $y_1, \ldots, y_N$ be a sample drawn out at random from a populations $N(\mu_2, \sigma_{y_i})$.

The statistic

$$\chi^2 = \sum_{i=1}^{N} \frac{[(x_i - y_i) - (\mu_1 - \mu_2)]^2}{\sigma_{x_i}^2 + \sigma_{y_i}^2}$$

is distributed as χ^2 with N degrees of freedom if the parameters μ_1, μ_2 are known, and with $N - 2$ degrees of freedom if they are estimated.

The χ^2-test for differences serves to define the interaction and, compared with the test u (based on the normal distribution), provides the basis of the estimation of the biological similarity of treatments.

It is only the joint interpretation of interaction, tendency and differentiation of the means that makes it possible to draw conclusions. The „multiple" test of F. Duncan [8] finds application here.