

ZASTOSOWANIE METOD WIELOWYMIAROWYCH DO OCENY WSPÓLZALEŻNOŚCI CECH UŻYTKOWYCH ZIARNA I MĄKI WYBRANYCH ODMIAN PSZENICY OZIMEJ

**Andrzej Gregorczyk¹⁾, Janusz Smagacz²⁾, Grzegorz Hury¹⁾,
Stanisława Ułasik¹⁾**

¹⁾ Katedra Uprawy Roli, Roślin i Doświadczalnictwa Akademii Rolniczej
w Szczecinie, Słowackiego 17, 71-434 Szczecin

e-mail: agregor@agro.ar.szczecin.pl

²⁾ Zakład Systemów i Ekonomiki Produkcji Roślinnej Instytutu Uprawy
Nawożenia i Gleboznawstwa – Państwowego Instytutu Badawczego

Czartoryskich 8, 24-100 Puławy

e-mail: smagacz@iung.pulawy.pl

Streszczenie

Badano dziesięć odmian pszenicy ozimej, rosnących przy zróżnicowanym poziomie nawożenia azotem. Oznaczono jedenaście cech ziarna i mąki.

Do opracowania danych zastosowano następujące metody statystyczne: regresję wieloraką, procedurę aglomeracji Warda w analizie skupień oraz analizę czynnikową.

Wykazano istotną zależność wartości walorymetrycznej pszenicy od kombinacji liniowej pięciu innych cech. Wyodrębniono trzy skupienia odmian pszenicy. Zredukowano wyjściowe dane do trzech wymiarów, łącznie wyjaśniających ok. 80% zmienności ogólnej.

Słowa kluczowe: analiza czynnikowa, analiza skupień, cechy technologiczne pszenicy, regresja wieloraka

Klasyfikacja AMS 1993: 97U70

1. Wstęp

W eksperymentach rolniczych zazwyczaj bada się wpływ różnych czynników doświadczalnych na kształtowanie w obiektach więcej niż jednej cechy. Badane cechy są często ze sobą skorelowane, dlatego interesujące może być wykrycie ogólnych prawidłowości w związkach między tymi zmiennymi. Wówczas należy stosować metody wielowymiarowej analizy danych, na przykład analizę regresji wielorakiej, analizę skupień, analizę czynnikową (Seber, 1984; Morrison, 1990). Metody te pozwalają między innymi na zbadanie zależności między jedną wybraną cechą a pozostałymi zmiennymi, grupowanie obiektów według hierarchii ich podobieństwa, uzyskiwanie informacji o teoretycznie ukrytej strukturze badanych cech, a więc umożliwiają kompleksową ocenę wyników doświadczenia.

Nawożenie azotem jest jednym z ważniejszych czynników plonotwórczych, mającym również duży wpływ na jakość ziarna i mąki pszenicy (Stankowski i Mortensen, 1997; Mazurek i Sułek, 1999; Wróbel i Szempliński, 1999). Najczęściej w badaniach agrotechnicznych nad nawożeniem oceniana jest jedna, najwyżej kilka odmian, co uniemożliwia jednoczesną ocenę wielu odmian o zróżnicowanych właściwościach dużej liczby cech. Celem przeprowadzonych badań była ocena współzależności jedenastu cech użytkowych ziarna i mąki dziesięciu wybranych odmian pszenicy ozimej.

2. Materiał i metody

Materiał do badań stanowiły próby ziarna pszenicy ozimej pochodzące z doświadczeń polowych przeprowadzonych w latach 2003/2004-2005/2006 przez Zakład Systemów i Ekonomiki Produkcji Roślinnej IUNG-PIB Puławy w RZD Grabowo.

Badano następujące odmiany pszenicy ozimej: Rywalka, Finezja, Turnia, Kobiera, Nutka, Tonacja, Soraja, Sława, Symfonia, Kaja. Zróżnicowano poziom nawożenia azotem: 90, 120, 150, 180 kg N ha⁻¹. Ilość wysiewu wynosiła 450 ziaren na 1m². Siew wykonano w terminie optymalnym. Zastosowano też optymalne nawożenie fosforowe i potasowe.

Oznaczono masę 1000 ziaren, gęstość ziarna, zawartość białka w ziarnie, liczbę opadania, zawartość i rozpuszczalność glutenu, sedimentację oraz wyko-

nano charakterystykę farinograficzną mąki obejmującą wodochłonność, rezy-stencję, rozmiękczenie i wartość walorymetryczną. Badania ziarna i mąki wy-konano stosując standardowe metody (Jakubczyk i Haber 1983).

Do analiz statystycznych wybrano cechy, które były zmiennymi losowymi ciągłymi. Wykonano – metodą krokową postępującą – analizę regresji wielora-kiej, mającą na celu określenie istotnych liniowych zależności między wartością walorymetryczną a pozostałymi cechami. Ogólnie model matematyczny regresji wielorakiej można przedstawić równaniem

$$Y = \alpha_0 + \sum_{i=1}^n \alpha_i X_i$$

gdzie: Y – zmienna objaśniana, X_1, X_2, \dots, X_n – zmienne objaśniające, $\alpha_0, \alpha_1, \dots, \alpha_n$ – parametry modelu (współczynniki regresji).

Estymowane równanie regresji ma postać

$$y = a_0 + \sum_{i=1}^n a_i x_i$$

Dla każdej zmiennej niezależnej X_i wyznacza się współczynniki a_0 i a_i , mi-nimalizujące sumę kwadratów odchyłeń wartości doświadczalnych od wartości regresyjnych. Następnie sprawdzana jest istotność tej regresji testem F. Do podzbioru wybiera się zmienną, dla której wartość statystyki Snedecora jest największa i przekracza wartość progową $F = 1,00$. Procedurę powtarza się dla pozostałych zmiennych, z uwzględnieniem zmiennych włączonych do modelu. Po ustaleniu wszystkich zmiennych niezależnych mających istotny wpływ na zmienną objaśnianą, dokonuje się końcowej estymacji parametrów regresji me-todą najmniejszych kwadratów, w której minimalizuje się sumę kwadratów odchyłeń obserwacji od wartości teoretycznych. Estymowane wartości współ-czynników regresji weryfikuje się testem t-Studenta. Adekwatność wybranego modelu przeprowadzono za pomocą analizy wariancji w regresji, posługując się testem F Snedecora.

W celu wyodrębnienia podobnych grup badanych odmian, ze względu na wszystkie cechy, zastosowano metodę analizy skupień. Celem uniezależnienia wyników grupowania obiektów od jednostek miar, dokonano standaryzacji zmiennych. Jako miarę odległości dwóch obiektów przyjęto odległość euklide-

sową oraz wykorzystano procedurę aglomeracji Warda. W procedurze tej spośród wszystkich możliwych do łączenia par skupień wybiera się tą parę, która daje skupienie o minimalnym zróżnicowaniu.

Natomiast celem redukcji liczby zmiennych i wykrycia struktury w powiązaniach między badanymi cechami zastosowano analizę czynnikową. Obliczenia analizy czynnikowej pozwoliły na zastąpienie zmiennych początkowych mniejszą liczbą nowych zmiennych, czyli tzw. czynników, od których pierwotne zmienne są liniowo zależne i które najlepiej wyjaśniają korelacje zachodzące między nimi. Liczbę czynników ustalono według kryterium Kaisera na podstawie wartości własnych macierzy korelacji oraz wykresu osypiska Cattella. Aby obliczyć wartości ładunków zastosowano procedurę składowych głównych z rotacją varimax znormalizowaną. Wszystkie obliczenia wykonano korzystając z pakietu Statistica 8.1.

3. Wyniki i podsumowanie

Bardzo ważną oceną pszenicy pod względem technologicznym jest jej wartość walorymetryczną, która określa wpływ, jaki wywiera dana pszenica na mieszankę z inną pszenicą lub innymi pszenicami.

Do opisu zależności między wartością walorymetryczną a pozostałymi cechami badanych odmian pszenicy posłużyły równania regresji wielorakiej, które wyznaczono osobno dla każdej dawki azotu i roku badań oraz całego okresu eksperymentu (Tab.1). W tej tabeli zamieszczono też wartości współczynnika determinacji (R^2), testu F, standardowego błędu estymacji (SE) i współczynników determinacji cząstkowej (r^2).

Wszystkie modele charakteryzują się znaczną zdolnością objaśniającą; świadczą o tym duże wartości współczynnika determinacji (powyżej 95%).

W modelach dla dawek azotu 120, 150, 180 kg N·ha⁻¹ stwierdzono łączny liniowy wpływ rezystencji, rozmiękczenia, rozplýwalności glutenu i zawartości białka na zmienną objaśnianą. W przypadku dawki 90 kg N·ha⁻¹ niewielki (na pograniczu istotności) udział miała jeszcze gęstość ziarna.

Dla danych zestawionych kolejno w latach zauważono większą różnorodność cech objaśniających na wartość walorymetryczną pszenic. Najbardziej rozbudowany model (zawierający sześć zmiennych objaśniających) uzyskano w roku 2006.

Tabela 1. Zależność wartości walorymetrycznej od cech użytkowych ziarna i mąki odmian pszenicy ozimej

Dawka azotu (kg·ha ⁻¹) lub rok (lata)	Równanie regresji wielorakiej	R ²	F	SE	r ²
90	$y = 2,172x_1 - 0,171x_2 + 1,245x_3 - 0,518x_4 + 0,628x_5 - 7,508$ (11,4 ^{***}) (-5,2 ^{***}) (2,6 [*]) (-1,7 [†]) (1,6 [†])	97,9	149,77 ^{***}	2,52	84,4/x ₁ ; 52,8/x ₂ ; 22,3/x ₃ ; 11,1/x ₄ ; 10,2/x ₅
120	$y = 2,200x_1 + 0,128x_2 + 1,648x_3 + 0,915x_4 + 5,831$ (0,04 ^{***}) (3,0 ^{**}) (2,5 [*]) (2,3 [*]) (4,7 ^{***})	95,7	140,30 ^{***}	3,08	80,1/x ₁ ; 26,4/x ₂ ; 20,3/x ₃ ; 17,1/x ₄
150	$y = 1,599x_1 + 0,188x_2 + 1,408x_3 + 1,068x_4 + 4,968$ (1,8 ^{**}) (4,9 [*]) (2,7 [*]) (3,1 ^{**}) (7,0 [*])	96,9	200,68 ^{***}	2,93	84,6/x ₁ ; 49,1/x ₂ ; 22,8/x ₃ ; 28,5/x ₄
180	$y = 1,642x_1 + 0,141x_2 + 1,839x_3 + 1,503x_4 + 8,837$ (3,0 ^{***}) (3,5 ^{**}) (3,2 ^{**}) (3,9 ^{**}) (5,5 ^{**})	97,2	215,62 ^{***}	2,83	87,2/x ₁ ; 33,4/x ₂ ; 29,3/x ₃ ; 37,7/x ₄
2004	$y = 2,709x_1 + 0,148x_2 + 0,392x_3 + 0,299x_4 + 4,179$ (18,8 ^{***}) (11,8 ^{***}) (2,2 [*]) (3,8 [*])	96,5	242,16 ^{***}	1,33	91,0/x ₁ ; 79,9/x ₂ ; 12,0/x ₃ ; 28,9/x ₄
2005	$y = 1,854x_1 + 0,198x_2 - 0,0118x_3 + 8,965$ (20,2 ^{***}) (11,1 ^{**}) (-2,0 [*]) (46,4 ^{***})	97,3	437,90 ^{***}	2,40	91,8/x ₁ ; 77,3/x ₂ ; 10,1/x ₃
2006	$y = 0,151x_1 + 1,267x_2 + 3,347x_3 + 0,674x_4 + 0,397x_5 + 1,314x_{10} + 8,321$ (3,8 ^{***}) (4,8 ^{**}) (4,1 ^{**}) (5,6 ^{***}) (2,6 [*]) (7,1 ^{***})	98,8	470,93 ^{***}	1,81	29,8/x ₁ ; 86,9/x ₂ ; 33,2/x ₃ ; 48,7/x ₄ ; 17,3/x ₅ ; 60,4/x ₁₀
2004-2006	$y = 1,769x_1 + 0,163x_2 + 1,664x_3 + 0,927x_4 + 0,384x_5 + 9,294$ (3,2 ^{***}) (8,4 ^{***}) (6,0 ^{***}) (5,2 ^{***}) (2,1 [*])	96,3	601,41 ^{***}	2,88	82,4/x ₁ ; 38,3/x ₂ ; 24,2/x ₃ ; 19,0/x ₄ ; 3,6/x ₅

y – wartość walorymetryczna

x₁ – rezystencja (min)x₂ – rozmięczenie (j.B.)x₃ – zawartość białka (%)x₄ – rozpywalność glutenu (min)x₅ – gęstość ziarna (kg hl⁻¹)x₆ – sedymentacjax₇ – liczba opadania (s)x₈ – wodochłonność (%)x₉ – masa 1000 ziarn (g)x₁₀ – zawartość glutenu (%)

(…) – wartość statystyki testowej t-Studenta

F – wartość statystyki testowej F-Snedecora

R² – współczynnik determinacji (%)

SE – standardowy błąd estymacji

r² – współczynnik determinacji cząstkowej (%)

Poziom isiościoci: □ - 0,1

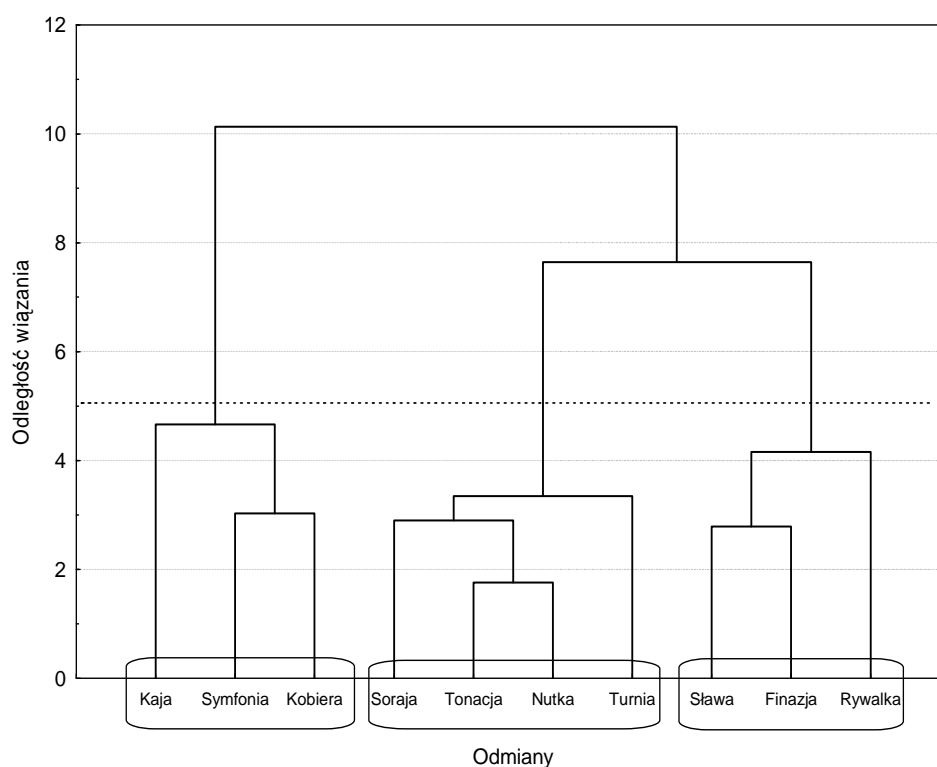
* - 0,05

** - 0,01

*** - 0,001

Ogólnie, wykorzystując łączne dane dla czterech poziomów nawożenia azotem i trzech lat badań można zauważyć istotne skorelowanie wartości walo-rymetrycznej mąki z rezystencją i rozmiękczeniem ciasta, rozplywalnością glutenu, zawartością białka w ziarnie oraz gęstością ziarna.

Knapowski i Balcewicz (2004) zastosowali do opisu zależności między cechami technologicznymi ziarna i mąki pszenicy ozimej odmiany Mikon równania regresji liniowej. Analiza regresji wielorakiej jest procedurą często wykorzystywaną w badaniach rolniczych, natomiast nie spotyka się tej metody w odniesieniu do cech technologicznych odmian pszenicy ozimej.



Rys. 1. Dendrogram odmian pszenicy (odległości euklidesowe)

Metody analizy skupień pozwalają na praktyczne grupowanie badanych obiektów (np. odmian) według hierarchii ich podobieństwa i przedstawienia rezultatów w postaci tzw. dendrogramów. Analizując diagram drzewa aglomeracji (rys. 1) można wyodrębnić trzy skupienia odmian ze względu na ich podobieństwo.

bieństwo w zmianach wartości wszystkich cech technologicznych. Pierwsze skupienie zawiera odmiany Tonacja, Nutka, Soraja i Turnia – euklidesowa odległość wiązania 3,345. Drugie grupuje odmiany Sława, Finezja i Rywalka (odległość wiązania 4,158). Przy odległości wiązania 4,662 powstaje trzecie skupienie odmian pszenicy: Symfonia, Kobiera, i Kaja.

Na podstawie macierzy korelacji wszystkich cech jakościowych pszenicy obliczono wartości własne, które odpowiadają jedenastu hipotetycznym czynnikom (tab. 2). Tabela 2 zawiera także część wariancji przypadającej na poszczególne czynniki, a także skumulowane wartości własne i skumulowaną zmienność.

Według kryterium Kaisera (wartości własne większe od jedności) i kryterium ospyska Cattella (punkt łagodnego spadku na wykresie wartości własnych), do dalszej analizy wyodrębniono pierwsze trzy czynniki.

Tabela 2. Wartości własne i procent całkowitej wariancji

Czynnik	Wartość własna	Wariancja (%)	Wartość własna skumulowana	Wariancja skumulowana (%)
1	4,616	41,96	4,616	41,96
2	2,533	32,03	7,149	64,99
3	1,601	14,56	8,750	79,55
4	0,715	6,50	9,465	86,05
5	0,496	4,51	9,962	90,56
6	0,407	3,70	10,369	94,27
7	0,338	3,08	10,708	97,35
8	0,158	1,44	10,867	98,79
9	0,079	0,72	10,947	99,52
10	0,028	0,26	10,976	99,78
11	0,023	0,21	11,000	100,00

Pierwszy czynnik wyjaśnia ok. 42% ogólnej zmienności danych wejściowych, czynnik drugi wykorzystuje 32% wariancji wspólnej, a trzecia składowa – blisko 15%.

Z tabeli 2 wynika również, że trzy pierwsze składowe objaśniają prawie 80 % zmienności całkowitej, czyli utrata informacji spowodowana redukcją wymiaru jest stosunkowo niewielka (ok. 20 %).

W celu wyodrębnienia ładunków czynnikowych zastosowano metodę składowych głównych, a celem uzyskania tzw. prostej struktury czynników, macierz ładunków czynnikowych została poddana rotacji varimax znormalizowanej.

Tabela 3 zawiera wartości ładunków czynnikowych, które można interpretować jako współczynniki korelacji między jedenastoma cechami wyjściowymi a wyodrębnionymi czynnikami wspólnymi.

Tabela 3. Ładunki czynnikowe po rotacji varimax znormalizowanej
(wartości pogrubione są większe od 0,7)

Cecha	Czynnik 1	Czynnik 2	Czynnik 3
Wartość walorymetryczna	0,9491	0,1605	0,0847
Zawartość białka	0,5330	0,7972	-0,1034
Masa 1000 ziaren	-0,4471	-0,6729	0,2381
Gęstość ziarna	0,2241	-0,2620	0,8139
Liczba opadania	-0,0953	0,2084	0,7971
Zawartość glutenu	0,1503	0,9636	0,1087
Rozpływalność glutenu	-0,6408	0,6652	-0,0820
Wodochłonność	0,0267	0,6047	0,4476
Rozmięczenie	-0,9183	-0,0154	-0,0935
Sedymentacja	0,8281	0,2096	-0,0930
Rezystencja	0,8214	0,2191	0,0607

Z analizy kolumny czynnika 1. wynika, że skorelowane są z nim w najbardziej cechy jakościowe mąki – wartość walorymetryczna, rozmięczenie, sedymentacja i rezystancja; skrótowo ten czynnik dałoby określić się jako „jakościowy”. Zmienne zawartość białka i zawartość glutenu mają duże ładunki czynnikowe z drugim czynnikiem, który można nazwać czynnikiem „białkowym”. Z kolei w największym stopniu czynnik 3. jest zdeterminowany przez gęstość ziarna (i liczbę opadania), a więc może to być składowa „fizyczna”.

Na podstawie powyższego układu korelacyjnego można przyjąć, że najwięcej zbiorczych informacji o własnościach technologicznych analizowanych odmian pszenicy wnosi osiem wyżej wyróżnionych cech. Kwadraty ładunków czynnikowych interpretuje się jako udziały wyjaśnionej wariancji. Przykładowo dla zmiennej wartość walorymetryczna, 90,0% jej wariancji jest wyjaśniona przez czynnik 1., a pozostałe dwa czynniki dodają kolejne 3,3% wariancji tej cechy pszenicy.

Analiza czynnikowa jest cenną, choć dość trudną do interpretacji metodą wielowymiarowej analizy danych (Morisson, 1990; Gregorczyk i in., 1998, 2004). Procedurę tą do zbadania wzajemnych zależności między zawartością składników mineralnych w ziarnie a parametrami jakościowymi mąki pszenicy zastosował Schnug i in. (1993)

Literatura cytowana

- Gregorczyk A., Chełpiński P., Ostrowska K., Gawska A. (2004). Application of multidimensional data analysis for describing the growth of new apple tree cultivars. *Folia Univ. Agric. Stetin.* 240(96), 75-80.
- Gregorczyk A., Ostrowska K., Chełpiński P. (1998). Comparison of multidimensional data analysis methods describing the growth of young apple trees. *Folia Hort.* 10(1), 75-84.
- Jakubczyk T., Haber T. (1983). *Analiza zbóż i przetworów zbożowych*. Wyd. SGGW Warszawa.
- Knapowski T., Balcewicz M. (2004). Evaluation of quantitative features of Mikon cultivar winter wheat grain and flour depending on selected agronomic factors. *Elect. J. Pol. Agric. Univ.* 7(1), 1-17.
- Morrison D.F. (1990). *Wielowymiarowa analiza statystyczna*. PWN Warszawa.
- Schnug E., Evans E., Haneklaus S. (1993). Multivariate analysis of relations between mineral element concentrations in grain and baking quality of wheat. *J. Agronomy and Crop. Sci.* 170, 1-11.
- Seber G.A.F. (1984). *Multivariate observations*. J. Wiley and Sons, New York.
- Stankowski S., Mortensen (1997). Wpływ zwiększonego stężenia CO₂ i nawożenia azotem na jakość ziarna pszenicy. *Biul. IHAR.* 204, 191-196.
- Wróbel E., Szempliński W. (1999). Plonowanie i wartość technologiczna ziarna pszenicy ozimej nawożonej zróżnicowanymi dawkami azotu. *Pam. Puł.* 118, 463-469.

APPLICATION OF MULTIVARIATE METHODS FOR EVALUATION OF RELATIONSHIPS BETWEEN QUALITATIVE TRAITS OF GRAIN AND FLOUR OF WINTER WHEAT VARIETIES

Summary

The ten varieties of winter wheat growing at different level of nitrogen fertilization have been investigated. The grain and flour eleven traits have been estimated. The data have been evaluated by means of the following statistical methods: multiple regression, Ward's procedure of agglomeration in cluster analysis and factor analysis. Significant dependence of calorimeter value of wheat on linear combination of five other traits has been obtained. Three clusters of wheat varieties have been separated. Primary variables have been reduced to three dimensions, which explained about 80% of general variation.

Key words and phrases: cluster analysis, factor analysis, multiple regression, wheat technological parameters

Classification AMS 1993: 97U70