

# Uwagi o zastosowaniach wykresu probabilistycznego gamma w wielozmiennej analizie wariancji

Wiesław Wagner

Katedra Metod Matematycznych i Statystycznych Akademii Rolniczej w Poznaniu

## Streszczenie

W pracy podano różne zastosowania wykresu probabilistycznego gamma typu kwantyl-kwantyl w wielozmiennej analizie wariancji. Wykres ten umożliwia badanie wektorowych kontrastów efektów głównych i interakcyjnych z jednym stopniem swobody w doświadczeniach czynnikowych typu  $2^k$ , efektów głównych i interakcyjnych w doświadczeniach czynnikowych oraz badanie jednorodności wielu macierzy kowariancji o jednakowej liczbie stopni swobody.

## 1. Wstęp

Wielozmianą analizę wariancji stosuje się do wnioskowania o wpływie czynników doświadczalnych, wyróżnionych w układzie doświadczalnym, na kształtowanie się  $p$  cech badanych jednocześnie na jednostkach doświadczalnych. Obejmuje ona testowanie hipotez głównych i szczegółowych o efektach głównych i interakcyjnych w modelach stałych (np. Caliński i Kaczmarek, 1973; Ahrens i Läuter, 1979). Do sprawdzenia tych hipotez stosowane są testy wymagające spełnienia założeń o wielowymiarowym rozkładzie normalnym i jednorodności macierzy kowariancji. Sprawdzenie pierwszego z tych założeń zostało omówione przez Wagnera (1990b).

Badanie wspomnianych założeń można także przeprowadzić metodami graficznymi. Metody te traktuje się jako uzupełnienie wobec stosowanych testów statystycznych. Użycie metod graficznych wymaga przekształcenia wektorów losowych w jednowymiarowe zmienne losowe, których rozkład jest znany przy prawdziwości hipotezy zerowej  $H_0$ . Przez hipotezę  $H_0$  będziemy rozumieli założenie, że nieobserwowalne wektory błędów losowych w wielozmianym modelu liniowym mają wielowymiarowy rozkład normalny z zerowym wektorem wartości oczekiwanych i jednakową, nieznaną, lecz dodatnio określoną macierzą kowariancji  $\Sigma$ .

W pracy podajemy zastosowania jednej z metod graficznych – wykresu probabilistycznego gamma typu kwantyl-kwantyl – w wielozmiennej analizie wariancji. Obejmuje ona badanie kontrastów efektów głównych i interakcyjnych w doświadczeniach czynnikowych oraz układu macierzy kowariancji o jednakowej liczbie stopni swobody.

Pojęcie wykresu probabilistycznego typu kwantyl-kwantyl podali Wilk i Gnanadesikan (1968), a także Wagner (1990a). Zainteresowanie skupiamy na wykresie probabilistycznym gamma (WPG), gdyż rozmaite wielkości jakie rozważamy w pracy są przybliżone zmiennymi o rozkładzie gamma.

## 2. Rozkład gamma i estymacja parametrów

Niech  $Y$  będzie zmienną losową o rozkładzie gamma określonym gęstością

$$f(y; \lambda, \eta, \rho) = \frac{\lambda^\eta}{\Gamma(\eta)} (y-\rho)^{\eta-1} \exp\{-\lambda(y-\rho)\} \quad ,$$

dla  $y > \rho$ ,  $\lambda > 0$ ,  $\eta > 0$  i  $-\infty < \rho < \infty$ , gdzie  $\rho$  jest parametrem przesunięcia,  $\lambda$  – skali, a  $\eta$  – kształtu. W szczególności, gdy  $\rho = 0$ , to mamy gęstość  $f(y; \lambda, \eta, 0) \equiv f(y; \lambda, \eta)$ , natomiast jeżeli  $\rho = 0$  i  $\lambda = 1$ , to mamy standardowy rozkład gamma o gęstości  $f(y; 1, \eta, 0) \equiv f(y; \eta)$ . Stosujemy oznaczenia  $Y \sim G(\lambda, \eta, \rho)$ ,  $Y \sim G(\lambda, \eta)$  i  $Y \sim G(\eta)$ , gdy zmienna losowa  $G$  ma rozkład o gęstości  $f(y; \lambda, \eta, \rho)$ ,  $f(y; \lambda, \eta)$  i  $f(y; \eta)$ .

*Lemat 1.* Funkcje gęstości rozkładów  $G(\lambda, \eta)$  i  $G(\lambda, \eta, \rho)$  wyznaczamy z gęstości  $G(\eta)$  z przekształceń liniowych  $y \rightarrow \lambda y$  i  $y \rightarrow \lambda(y-\rho)$ .

*Dowód.* Wynika z bezpośrednich przeliczeń.  $\square$

W naszych rozważaniach zainteresowani jesteśmy głównie rozkładami  $G(\eta)$  i  $G(\lambda, \eta)$ . Dla nich podamy estymację parametrów metodą największej wiarygodności.

Niech  $y_1, \dots, y_n$  stanowi próbę  $n$  niezależnych realizacji zmiennej losowej  $Y$ , o której zakładamy, że przyjmuje tylko wartości dodatnie. Oznaczamy przez  $\mathbf{y} = (y_1, \dots, y_n)'$ ,

$$P = \prod_{i=1}^n y_i \quad \text{oraz} \quad S = \sum_{i=1}^n y_i \quad .$$

Zajmujemy się estymacją parametru  $\eta$  w rozkładzie  $G(\eta)$ .

Funkcja wiarygodności przyjmuje tutaj postać

$$L(\eta; \mathbf{y}) = \prod_{i=1}^n f(y_i; \eta) = \left( \frac{1}{\Gamma(\eta)} \right)^n P^{\eta-1} e^{-S} \quad .$$

Po zlogarytmowaniu, obliczeniu pochodnej względem  $\eta$  i przyrównaniu jej do zera, otrzymujemy równanie

$$\psi(\eta) = \frac{1}{n} \ln P \quad ,$$

gdzie  $\psi(\eta) = \Gamma'(\eta)/\Gamma(\eta)$  jest funkcją di-gamma (pochodną logarymiczną funkcji gamma). Numeryczny algorytm jej wyznaczenia podaje Gnanadesikan (1977).

Z kolei dla rozkładu  $G(\lambda, \eta)$  funkcja wiarygodności przyjmuje postać

$$L(\lambda, \eta; y) = \prod_{i=1}^n f(y_i; \lambda, \eta) = \left(\frac{\lambda^n}{\Gamma(\eta)}\right)^n P^{\eta-1} e^{-\lambda S} .$$

Znowu jak wyżej po zlogarytmowaniu, obliczeniu pochodnych cząstkowych względem  $\eta$  i  $\lambda$  i przyrównaniu ich do zera, otrzymujemy układ równań

$$\begin{aligned} \psi(\eta) - \ln \lambda &= \frac{1}{n} \ln P , \\ \frac{\eta}{\lambda} &= \frac{1}{n} S . \end{aligned}$$

W obu wypadkach znajdujemy parametry metodą iteracji. Jako pierwsze przybliżenia można wziąć rozwiązania z układu (z estymacji metodą momentów)

$$\bar{y} \approx E(Y) = \frac{\eta}{\lambda} \quad \text{i} \quad \frac{1}{n} \sum_{i=1}^n y_i^2 \approx E(Y^2) = \frac{\eta(\eta+1)}{\lambda^2} .$$

Inną metodę estymacji proponowali Wilk i in. (1962). Korzystają oni z próbkowych statystyk pozycyjnych  $y_{(1)} \leq \dots \leq y_{(n)}$ , które są uporządkowanymi w ciąg niemalejący wielkościami z próby. Biorą oni pod uwagę  $K$  statystyk pozycyjnych  $y_{(1)} \leq \dots \leq y_{(M)} \leq \dots \leq y_{(K)}$ , ( $K \leq n$ ), spośród nich wykorzystują  $M$  statystyk pozycyjnych ( $M \leq K$ ). Podali oni odpowiednie tablice dla wyznaczenia estymatorów parametrów  $\lambda$  i  $\eta$  zależnie od ilorazu  $K/M = 1.0, 1.1, 1.2, 1.3, 1.4, 1.6, 1.8, 2.0, 2.3, 2.6$  i  $3.0$ .

### 3. Kwantyle rozkładu gamma i wpg

Niech  $q^*(\alpha)$ ,  $\tilde{q}(\alpha)$  i  $q(\alpha)$  oznaczają kwantyle rzędu  $0 < \alpha < 1$  odpowiednio rozkładów  $G(\lambda, \eta, \rho)$ ,  $G(\lambda, \eta)$  i  $G(\eta)$ . Określa się je jako rozwiązania równań:

$$\int_0^{q^*(\alpha)} f(y; \lambda, \eta, \rho) dy = \alpha , \quad \int_0^{\tilde{q}(\alpha)} f(y; \lambda, \eta) dy = \alpha , \quad \text{lub} \quad \int_0^{q(\alpha)} f(y; \eta) dy = \alpha$$

Związek między tymi kwantylami wyraża lemat.

*Lemat 2.* Kwantyle  $\tilde{q}(\alpha)$  i  $q^*(\alpha)$  rozkładów  $G(\lambda, \eta)$  i  $G(\lambda, \eta, \rho)$  wyrażają się przez kwantyle  $q(\alpha)$  rozkładu  $G(\eta)$  relacją  $\tilde{q}(\alpha) = q(\alpha)/\lambda$  i  $q^*(\alpha) = q(\alpha)/\lambda + \rho$ .

*Dowód.* Dla wyznaczenia kwantyli  $\tilde{q}(\alpha)$  od  $q(\alpha)$ , stosujemy przekształcenie  $y = \lambda \tilde{y}$ , czyli  $\tilde{y} = y / \lambda$ , a stąd granice całkowania przechodzą na 0 oraz  $q(\alpha) / \lambda = \tilde{q}(\alpha)$ . Z kolei przy wyznaczaniu  $q^*(\alpha)$ , stosujemy przekształcenia  $\tilde{y} = y^* - \rho$ , czyli  $y^* = \tilde{y} + \rho = y / \lambda + \rho$ . Dokonując zamiany granic całkowania, mamy nowe granice całkowania  $\rho$  oraz  $q(\alpha) / \lambda + \rho = q^*(\alpha)$ .

Wyznaczenie kwantyli  $q(\alpha)$  można przeprowadzić różnymi metodami. Jedna z nich sprowadza się do zastosowania metody iteracyjnej Newtona-Raphsona wyznaczenia pierwiastka równania

$$\sum_{k=0}^{\infty} \frac{(-1)^k (q(\alpha))^{k+\eta}}{k! (k+\eta)} - \alpha \Gamma(\eta) = 0.$$

Inne podejście zasugerował Gnanadesikan (1977), przyjmując dolne ograniczenie  $q(\alpha) = [\alpha \Gamma(\eta)]^{1/\eta}$ . Wilk i in. (1962) wyznaczają kwantyle z równania

$$[q(\alpha)]^\eta e^{-q(\alpha)} T_m(q(\alpha), \eta) = \alpha \Gamma(\eta),$$

gdzie

$$T_m(q(\alpha), \eta) = \sum_{j=0}^{m-1} d_j,$$

$$d_0 = 1/\eta,$$

$$d_{j+1} = q(\alpha) d_j / (\eta + j + 1),$$

a  $m (> q(\alpha) - \eta)$  jest tak dobrane, aby

$$\frac{q(\alpha)}{\eta - q(\alpha) + m} d_{m-1} \leq 10^{-7} T_m(q(\alpha), \eta).$$

Łączne prawdopodobieństwa wyznaczamy ze wzoru

$$\alpha_j = \frac{j-a}{n-2a+1}, \quad j = 1, \dots, n,$$

gdzie  $a \in <0, 0.5>$  (Blom, 1958). Wstawiając do podanych wzorów różne  $\alpha_j$ , otrzymujemy kwantyle rozkładu  $G(\eta)$  równe  $q_1(\alpha_1)$ ,  $q_2(\alpha_2)$ , ...,  $q_n(\alpha_n)$ .

*Definicja 1.* Wykresem probabilistycznym rozkładu gamma  $G(\eta)$  nazywamy zbiór par  $\{(q_j(\alpha_j), y_{(j)})\}$ ,  $j = 1, \dots, n$ , gdzie  $q_j(\alpha_j)$  są kwantylami rozkładu gamma  $G(\eta)$  rzędu  $0 < \alpha_j < 1$ , natomiast  $y_{(j)}$  są próbkowymi statystykami pozycyjnymi.

Skoro pokazaliśmy jak można otrzymać kwantyle  $\tilde{q}_j(\alpha_j)$  i  $q_j^*(\alpha_j)$ , więc analogiczne definicje formułuje się dla wykresów probabilistycznych rozkładów  $G(\lambda, \eta)$  i  $G(\lambda, \eta, \rho)$ .

#### 4. Zastosowania

##### 4.1. Badanie kontrastów efektów głównych i interakcyjnych w doświadczeniach typu $2^k$ .

Niech  $A, B, \dots, K$  będą  $k$  czynnikami oraz niech  $A_0, A_1, B_0, B_1, \dots, K_0, K_1$  stanowią ich poziomy. Liczba kombinacji między poziomami wynosi  $2^k$ . Kombinacje te zapisujemy według porządku Yatesa:  $1' = A_0 B_0 \dots K_0$ ,  $a = A_1 B_0 \dots K_0$ ,  $b = A_0 B_1 \dots K_0$ ,  $ab = A_1 B_1 \dots K_0$ ,  $\dots$ ,  $ab \dots k = A_1 B_1 \dots K_1$ . Kombinacje numerujemy kolejno liczbami naturalnymi  $1, \dots, n$ , a  $i$ -tą kombinację wyrażamy ciągiem  $i = (i_1, \dots, i_k)$ ;  $i_1, \dots, i_k = 0, 1$ , z kolejnym numerem  $i = 1 + \sum_{l=1}^{k-1} i_l 2^l$ . Kombinacje traktowane jako obiekty doświadczalne mogą być rozlosowane w jednym z podstawowych układów doświadczalnych, w tym także w układzie split-plot. Przyjmujemy, że każda kombinacja (obiekt) jest replikowana  $r$  razy, czyli łącznie mamy  $n \cdot r$  obserwowanych wektorów  $p$ -wymiarowych. Dalej rozważamy sumy tych obserwacji dla poszczególnych cech, które wyraża wektor  $y_i' = (y_{i1}, \dots, y_{ip})$ ,  $i = 1, \dots, n$ . Układ  $n$  wektorów wierszowych  $p$ -wymiarowych zaobserwowanych sum tworzy  $(n \times p)$ -wymiarową macierz

$$Y' = \begin{bmatrix} y_1' \\ \vdots \\ y_n' \end{bmatrix} = [Y_1, \dots, Y_p] \quad ,$$

gdzie  $y_i \in \mathbb{R}^p$  oraz  $Y_j \in \mathbb{R}^n$ . Zastosowanie wykresów probabilistycznych do każdej cechy oddzielnie, czyli do wektorów  $Y_1, \dots, Y_p$ , jest omówione w pracy Wagnera (1990a).

O wektorach  $y_1', \dots, y_n'$  zakładamy, że są wzajemnie nieskorelowane i mają jednakowe macierze kowariancji  $\Sigma$ , czyli

$$E[(y_i - E(y_i)) (y_{i'} - E(y_{i'}))'] = \begin{cases} \Sigma, & i=i' \\ 0, & i \neq i' \end{cases}$$

przy  $i, i' = 1, \dots, n$ , gdzie  $\Sigma$  jest macierzą dodatnio półokreśloną. Postać wektora wartości oczekiwanych  $E(y_i')$  wyraża  $E(Y') = X B$ , gdzie  $X : n \times q$  jest znaną macierzą układu, a  $B : q \times p$  – macierzą  $q$  nieznanych parametrów dla  $p$  cech. Struktura układu doświadczalnego typu  $2^k$  pokazuje, że  $q = 3^k$ .

*Przykład 1.* Niech  $A$  i  $B$  będą dwoma czynnikami o poziomach  $A_0, A_1, B_0, B_1$ . Przyjmujemy, że na każdej jednostce doświadczalnej obserwujemy jednocześnie  $p = 3$  cechy. Kombinacje  $A_i B_j$ :  $i_1 = 0, 1$ ,  $i_2 = 0, 1$ , są rozlosowane w układzie kompletnej randomizacji o  $r$  replikacjach. Model liniowy dla sum obserwacji cech dla kombinacji  $(i_1, i_2)$  jest postaci

$$y_{i_1 i_2 j} = \mu_j + \alpha_{i_1 j} + \beta_{i_2 j} + (\alpha\beta)_{i_1 i_2 j} + e_{i_1 i_2 j}$$

przy  $i_1, i_2 = 0, 1; j = 1, 2, 3$ . Zakładamy, że  $E(e_{i_1 i_2 j}) = 0$ ,  $D^2(e_{i_1 i_2 j}) = \sigma_{jj}$  oraz  $\text{Cov}(e_{i_1 i_2 j}, e_{i_1 i_2 j'}) = \sigma_{jj}$ . Tutaj mamy  $n=4$ ,  $q=3$ ,

$$\begin{aligned} \mathbf{X} &= [\mathbf{1}_2 \otimes \mathbf{1}_2, \mathbf{1}_2 \otimes \mathbf{I}_2, \mathbf{I}_2 \otimes \mathbf{1}_2, \mathbf{I}_2 \otimes \mathbf{I}_2] \\ \mathbf{B} &= [\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\alpha\beta}]' \end{aligned}$$

gdzie

$$\boldsymbol{\mu}: 1 \times 3, \quad \boldsymbol{\alpha}: 2 \times 3, \quad \boldsymbol{\beta}: 2 \times 3, \quad \boldsymbol{\alpha\beta}: 4 \times 3.$$

Celem naszych rozważań jest zastosowanie WPG do analizy kontrastów każdy z 1 stopniem swobody. Uzyskujemy je z macierzy  $\mathbf{Y}$  przez przekształcenie ortogonalne

$$x_{ij} = \mathbf{R}'_i \mathbf{Y}_j, \quad i = 1, \dots, n; \quad j = 1, \dots, p,$$

gdzie  $\mathbf{R}'_i$  jest  $i$ -tym wierszem macierzy  $\mathbf{R}$  podanej przez Wagnera (1990a). Mamy wówczas  $E(x_{ij}) = \mathbf{R}'_i E(\mathbf{Y}_j) = \mathbf{R}'_i \mathbf{X} \mathbf{B}_j$ , gdzie  $\mathbf{B}_j$  jest wektorem parametrów efektów głównych i interakcyjnych odniesionych do  $j$ -tej cechy. Wielkości  $x_{ij}$  wyrażają kontrasty dla efektów głównych i interakcyjnych kolejnych kombinacji  $i = (i_1, \dots, i_k)$  dla  $j$ -tej cechy. Wektory tych kontrastów oznaczamy przez  $\mathbf{x}'_i = (x_{i1}, \dots, x_{ip})$  przy  $i = 2, \dots, n$ , gdyż pierwszy wiersz macierzy  $\mathbf{R}$  zawiera identyczne elementy równe  $1/\sqrt{n}$ . Z własności macierzy  $\mathbf{R}$  wynika, że  $\mathbf{x}'_1$  jest wektorem postaci  $\sqrt{n}(\bar{y}_1, \dots, \bar{y}_p)'$ , a ponadto

$$\mathbf{R}'_i \mathbf{R}'_{i'} = \begin{cases} 1, & i=i' \\ 0, & i \neq i' \end{cases}$$

Korzystając z ortogonalności macierzy  $\mathbf{R}$ , mamy  $D^2(x_{ij}) = \mathbf{R}'_i D(\mathbf{Y}_j) \mathbf{R}_i = \mathbf{R}'_i (\sigma_{jj} \mathbf{I}) \mathbf{R}_i = \sigma_{jj}$  oraz  $\text{Cov}(x_{ij}, x_{i'j'}) = \sigma_{jj}$ . A zatem wektory  $\mathbf{x}'_i$  mają wspólną macierz kowariancji  $\boldsymbol{\Sigma}$ .

Jeżeli kontrasty są nieistotne, to  $E(x_{ij}) = 0$  dla wszystkich  $i, j$ . Takie założenie będziemy oznaczać przez  $Z$ . Przy tym założeniu stwierdzamy, że jeżeli  $\mathbf{y}'_i$  mają wielowymiarowy rozkład normalny, to także taki rozkład mają wektory  $\mathbf{x}'_i$ , ściślej  $\mathbf{x}'_i \sim N(0, \boldsymbol{\Sigma})$ . Kontrasty  $\mathbf{x}'_i$ ,  $i = 2, \dots, n$  można także traktować jako punkty w przestrzeni  $R^p$ , a ich miarę odległości od początku układu współrzędnych wyrażają formy kwadratowe dodatnio półokreślone  $d_i = \mathbf{x}'_i \mathbf{A} \mathbf{x}_i$ , gdzie macierz  $\mathbf{A}$  dobieramy według jednego z wariantów:

$$\mathbf{A} = \begin{cases} \mathbf{I}, \\ \text{diag}(1/s_1^2, \dots, 1/s_p^2), \\ \mathbf{S}^{-1}, \text{ gdy } |\mathbf{S}| \neq 0, \\ \mathbf{S}^-, \text{ gdy } |\mathbf{S}| = 0, \end{cases}$$

gdzie  $s_j^2$  są próbkowymi wariancjami cech.

Dla wybranej macierzy  $A$  każda odległość  $d_i, i = 2, \dots, n$  ma rozkład wyrażony liniową kombinacją  $m$  wzajemnie niezależnych zmiennych losowych  $\chi_1^2$  (lub  $\chi_1^2$  przy założeniu  $Z$ ), gdzie  $m = r(A) \leq p$  (Wilk i Gnanadesikan, 1964). W szczególności, gdy  $A = b\Sigma$ , to każde  $d_i \sim b \chi_p^2$ , gdzie  $b \neq 0$  jest dowolną stałą. Jak pokazali Patnaik (1949) i Box (1954), mimo że  $\Sigma$  jest nieznaną, to przybliżenie kombinacji liniowej niezależnych zmiennych losowych może być nadal stosowane.

Tak więc przy założeniu  $Z$  odległości  $d_2, \dots, d_n$  (choć w przybliżeniu) traktujemy jako próbę losową pobraną z populacji o rozkładzie gamma  $G(\lambda, \eta)$ .

Badanie odstępstwa od założenia  $Z$  rozważanych kontrastów badamy poprzez ich odchylenie od liniowości na WPG  $\{(\tilde{q}_j(\alpha_j), d_{(j)}), j = 2, \dots, n\}$ , gdzie  $\tilde{q}_j(\alpha_j)$  są kwantylami rzędu  $\alpha_j$  rozkładu  $G(\lambda, \eta)$ , a  $d_{(j)}$  są próbkowymi statystykami pozycyjnymi takimi, że  $d_{(2)} \leq d_{(3)} \leq \dots \leq d_{(n)}$ . Prosta na WPG jest nachylona pod kątem  $\text{tg} \varphi = 1/\lambda$  i przechodzi przez początek układu współrzędnych. A zatem parametr skalujący  $\lambda$  nie wpływa na współliniowość punktów na WPG.

Jeżeli założenie  $Z$  nie jest spełnione przez dane doświadczalne, to duże wartości  $d_j$  będą odchodziły się od linii prostej na WPG. Te kontrasty uważa się za istotnie różne od pozostałych. Tworzą one podgrupę jednorodnych kontrastów. Wydzielając te kontrasty z pozostałych, możemy sporządzić nowy WPG. Wymaga to wyznaczenia nowych kwantyli rozkładu gamma. Wykonując nowy WPG, wydzielamy drugą podgrupę kontrastów jednorodnych, które najsilniej odchylają się od pozostałych. Ta procedura wydzielenia jednorodnych podgrup kontrastów może być kontynuowana wielokrotnie.

#### 4.2. Badanie jednorodności macierzy kowariancji.

Niech znowu  $A, B, \dots, K$  będą  $k$  czynnikami doświadczalnymi, lecz teraz odpowiednio o liczbie poziomów  $m_1, m_2, \dots, m_k$ . Każdej kombinacji  $A_i B_j \dots K_k, i = 1, \dots, m_1$  przypisujemy  $r$  replikacji. Łącznie mamy  $m = m_1 \cdot m_2 \cdot \dots \cdot m_k$  kombinacji. Przyjmujemy, że na każdej jednostce doświadczalnej obserwowanych jest jednocześnie  $p$  cech. Dla  $i$ -tej kombinacji  $i = (i_1, \dots, i_k)$  wyznaczamy macierz kowariancji  $S_i^*, i = 1, \dots, m$ . Zakładamy, że  $S_i^*$  są macierzami dodatnio półokreślonymi gdyż dopuszczamy sytuację, gdy  $r \leq p$ .

Badane cechy mogą być wyrażone w różnych mianach. Wymaga to przeskalowania macierzy  $S_i^*$  przez pewną macierz  $A$ , której wyboru można dokonać jak podano w punkcie 4.1. Otrzymamy wtedy macierze  $S_i = S_i^* A, i = 1, \dots, m$ . Z macierzy  $S_i$  wyznaczamy wartości własne, powiedzmy  $c_{i1} \geq c_{i2} \geq \dots \geq c_{it} > 0$ , gdzie  $t = r(S_i)$ , z których budujemy funkcje  $F_i(c_{i1}, \dots, c_{it})$ . Takimi funkcjami mogą być

$$F_i(c_{i1}, \dots, c_{it}) = \begin{cases} \sum_{k=1}^t c_{ik} = \text{tr}(S_i) , \\ \left( \prod_{k=1}^t c_{ik} \right)^{1/t} , \text{ gdy } t = r(S_i) < p , \\ \left( \prod_{k=1}^p c_{ik} \right)^{1/p} , \text{ gdy } r(S_i) = p , \\ c_{\max}(c_{i1}, \dots, c_{it}) = c_{i1} , \\ t / \sum_{k=1}^t (1/c_{ik}) , \text{ gdy } t = r(S_i) < p . \end{cases}$$

Przez  $Z_1$  będziemy teraz rozumieć założenie, że  $p$ -wymiarowe obserwacje wektorowe są wzajemnie stochastycznie niezależne z identyczną choć nieznaną macierzą kowariancji  $\Sigma$  i brakiem oddziaływania efektów głównych i interakcyjnych.

Niech  $a_i = \text{tr}(S_i) = \text{tr}(S_i^* A)$ . Określamy rozkład  $a_i$  przy założeniu  $Z_1$ . Rozkład ten przybliżamy rozkładem gamma. Zauważmy, że  $S_i^* = Z_i Z_i' = Y' Q_i' Q_i Y$ , gdzie  $Y : n \times p$  jest macierzą  $p$ -wymiarowych obserwacji doświadczalnych, natomiast  $Q_i : v \times n$  jest taką macierzą, że  $Q_i Q_i' = I_v$  oraz  $Q_i Q_i'' = 0$ ,  $i \neq i'$ . Tutaj przez  $v$  oznaczamy liczbę stopni swobody

macierzy  $S_i$ . Mamy więc  $a_i = \text{tr}(Z_i Z_i' A) = \sum_{j=1}^v z_{ij}' A z_{ij}$ , gdzie  $z_{ij}$  jest  $j$ -tą kolumną  $Z_i$ .

Przy założeniu  $Z_1$  wektory  $z_{ij}$ ,  $i = 1, \dots, m$ ;  $j = 1, \dots, v$  traktujemy jako próbę  $p$ -wymiarowych wektorów wzajemnie niezależnych, każdy o rozkładzie  $N_p(0, \Sigma)$ . Tym samym  $a_i$  traktujemy jako sumy wzajemnie niezależnych dodatnio półokreślonych form kwadratowych a także  $a_1 \dots a_m$  są niezależne (Gnanadesikan i Lee, 1970). Rozkład form kwadratowych przybliżamy rozkładem gamma  $G(\lambda, \eta/v)$ . WPG sporządzamy za pomocą zbioru par  $\{(\tilde{q}_j(\alpha_j), a_{(j)}), j = 1, \dots, m\}$ , gdzie  $a_{(1)} \leq a_{(2)} \leq \dots \leq a_{(m)}$ , natomiast  $\tilde{q}_j(\alpha_j)$  są kwantylami w rozkładzie  $G(\lambda, \eta/v)$ . Interpretację przeprowadzamy analogicznie jak w punkcie 4.1.

WPG może być wykorzystany także do badania efektów głównych i interakcyjnych w doświadczeniach czynnikowych typu  $m^k$ ,  $m \geq 3$ . W tych doświadczeniach efekty główne mają  $v = n - 1$  stopni swobody, a efekty interakcyjne mają  $v = (m - 1)^{q+1}$  stopni swobody, gdzie  $q$  jest rzędem interakcji.

*Przykład 2.* Zastosowanie WPG przedstawiamy na danych zaczerpniętych z monografii Ahrensa i Lautera (1979). W doświadczeniu dwuczynnikowym: A – rasy zwierząt użytkowych ( $a = 3$  poziomy) i B – zabiegi stymulujące ich wzrost ( $b = 3$  poziomy) badano  $p = 2$  cechy, a mianowicie przyrost na masie ciała w dwóch następujących czasowo po sobie okresach wzrostu. Dla każdej rasy i każdego zabiegu przeprowadzono pomiary na 4 zwierzętach. Obserwacje poszczególnych kombinacji  $A_i B_j$  obu czynników wyrażają dwuwymiarowe wektory wierszowe:



$A_1B_1$ :	(4,5)	(4,3)	(8,8)	(6,5);
$A_1B_2$ :	(6,5)	(6,6)	(8,11)	(5,7);
$A_1B_3$ :	(17,13)	(10,8)	(13,10)	(8,6);
$A_2B_1$ :	(6,9)	(5,5)	(8,6)	(7,9);
$A_2B_2$ :	(9,12)	(7,6)	(6,5)	(5,8);
$A_2B_3$ :	(12,10)	(10,8)	(10,9)	(9,8);
$A_3B_1$ :	(8,10)	(7,7)	(10,8)	(9,11);
$A_3B_2$ :	(11,13)	(9,8)	(8,7)	(7,9);
$A_3B_3$ :	(13,11)	(11,8)	(11,9)	(10,7).

Zamierzamy odpowiedzieć na pytanie, czy obserwowalne dwuwymiarowe wektory dla kombinacji spełniają warunek jednorodności macierzy kowariancji.

Z danych doświadczalnych mamy:  $n=9$ ,  $p=2$ ,  $v=3$  oraz następujące macierze kowariancji dla kombinacji  $A_jB_j$  i ich ślady:

$$A_1B_1: \begin{bmatrix} 3.667 & 3.500 \\ 3.500 & 4.250 \end{bmatrix}, \quad a_1 = 7.917,$$

$$A_1B_2: \begin{bmatrix} 1.583 & 2.583 \\ 2.583 & 6.917 \end{bmatrix}, \quad a_2 = 8.500,$$

$$A_1B_3: \begin{bmatrix} 15.333 & 11.667 \\ 11.667 & 8.917 \end{bmatrix}, \quad a_3 = 24.250,$$

$$A_2B_1: \begin{bmatrix} 1.667 & 0.500 \\ 0.500 & 4.250 \end{bmatrix}, \quad a_4 = 5.917,$$

$$A_2B_2: \begin{bmatrix} 2.917 & 3.583 \\ 3.583 & 9.593 \end{bmatrix}, \quad a_5 = 12.500,$$

$$A_2B_3: \begin{bmatrix} 1.583 & 1.083 \\ 1.083 & 0.917 \end{bmatrix}, \quad a_6 = 2.500,$$

$$A_3B_1: \begin{bmatrix} 1.667 & 0.667 \\ 0.667 & 3.333 \end{bmatrix}, \quad a_7 = 5.000,$$

$$A_3B_2: \begin{bmatrix} 2.917 & 3.417 \\ 3.417 & 6.917 \end{bmatrix}, \quad a_8 = 9.834,$$

$$A_3B_3: \begin{bmatrix} 1.583 & 2.083 \\ 2.083 & 2.917 \end{bmatrix}, \quad a_9 = 4.500.$$

Ślady macierzy kowariancji zawierają się w przedziale  $\langle 2.5, 24.25 \rangle$ , przy czym skupiają się one po lewej stronie. Oceną macierzy kowariancji  $\Sigma$  jest macierz

$$\begin{bmatrix} 3.657 & 3.231 \\ 3.231 & 5.333 \end{bmatrix}.$$

Porządkujemy w ciąg niemalejący ślady  $x_j = a_{(j)} : 2.5, 4.5, 5, 5.917, 7.917, 8.5, 9.833, 12.5, 24.25$ . Wyznaczamy zgodnie z sugestią podaną w rozdziale 2 oceny parametrów rozkładu  $G(\lambda, \eta/3)$  w oparciu o ciąg  $x_j$ . Ustalamy  $K=9$ ,  $M=5$  i  $K/M = 1.8$ . Dla ustalenia ocen  $\lambda$  i  $\eta/3$  obliczamy wielkości pomocnicze

$$P = \left( \prod_{j=1}^M x_j \right)^{1/M} / x_M = \left( \prod_{j=1}^5 x_j \right)^{0.2} / 7.917$$

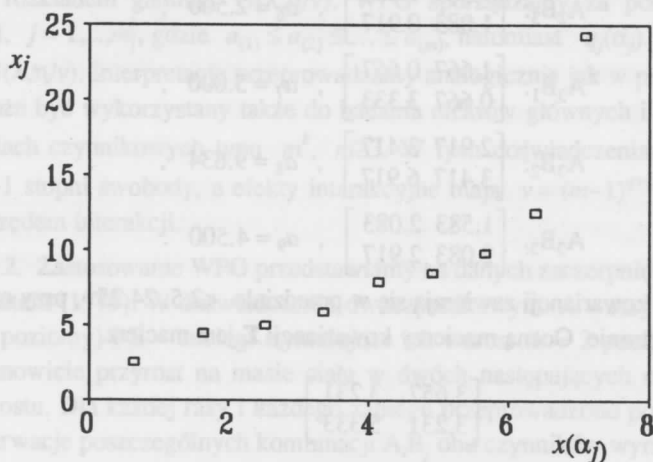
oraz

$$S = \sum_{j=1}^M x_j / M x_M = \sum_{j=1}^5 x_j / (5 \cdot 7.917),$$

które wyniosły  $P = 0.6104$  i  $S = 0.6526$ . Korzystając z tablic podanych przez Wilka i in. (1962, s.541) znajdujemy  $\hat{\gamma} = \hat{\eta}/3 = 1.1824$  i  $\hat{\lambda} = \hat{\eta}/(3\hat{\mu}x_M) = 1.1824/(1.0328 \cdot 7.917) = 0.1446$ . Wyznaczamy dla rozkładu  $G(0.1446, 1.1824)$  kwantyle  $\tilde{q}_j(\alpha_j)$ ,  $j = 1, \dots, 9$ , przyjmując  $\alpha_j = (j-0.5)/9$ , tzn.  $a = 0.5$ . Ograniczamy się do dolnych wartości tych kwantyli, które wyznaczamy następująco

$$\tilde{q}_j(\alpha_j) = [\alpha_j \hat{\gamma} \Gamma(\hat{\gamma})]^{1/\hat{\gamma}} = (\alpha_j \cdot 1.1824 \cdot 0.924)^{1/1.1824} = (\alpha_j \cdot 1.09254)^{0.8457}.$$

Szczegółowe obliczenia dostarczyły kwantyli  $\tilde{q}_j(\alpha_j)$ : 0.647, 1.638, 2.523, 3.353, 4.147, 4.914, 5.660, 6.388 i 7.101. Na rysunku 1 przedstawiono WPG dla zbioru par  $\{(0.647, 2.5), (1.638, 4.5), (2.523, 5), (3.353, 5.917), (4.147, 7.917), (4.914, 8.5), (5.66, 9.833), (6.388, 12.5), (7.101, 24.25)\}$ . Rysunek 1 pokazuje, że odstającym punktem jest  $(7.101, 24.25)$ , czyli macierz kowariancji wyznaczona z obserwacji wektorowych dla kombinacji  $A_1 B_3$  wskazuje na odstępstwo od jednorodności wszystkich macierzy kowariancji.



Rys. 1. Wykres probabilistyczny gamma

## LITERATURA

- Ahrens H., Läuter J. (1979). *Wielowymiarowa analiza wariancji*. PWN Warszawa.
- Blom G. (1958). *Statistical estimates and transformed beta-variables*. Stockholm.
- Box G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems – I. *Ann. Math. Statist.* 25, 290–302.
- Caliński T., Kaczmarek Z. (1973). Metody kompleksowej analizy doświadczeń wielocechowych. W: *Trzecie Colloquium Metodologiczne z Agro-Biometrii*, PAN i PTB, Warszawa, 258–320.
- Gnanadesikan R., Lee E.T. (1970). Graphical techniques for internal comparisons amongst equal degree of freedom groupings in multiresponse experiments. *Biometrika* 57, 229–237.
- Gnanadesikan R. (1977). *Methods for statistical data analysis of multivariate observations*. Wiley and Sons, New York.
- Patnaik P.B. (1949). The non-central  $\chi^2$  and F-distributions and their approximations. *Biometrika* 36, 202–232.
- Wagner W. (1990a). Zastosowanie wykresów probabilistycznych w analizie wariancji. *Listy Biometryczne* 27, 47–60.
- Wagner W. (1990b). Test normalności wielowymiarowej Shapiro-Wilka i jego zastosowanie w doświadczalnictwie rolniczym. *Roczniki Akademii Rolniczej w Poznaniu, Rozprawy Naukowe*, z. 197.
- Wilk M.B., Gnanadesikan R. (1964). Graphical methods for internal comparisons in multiresponse experiments. *Ann. Math. Statist.* 35, 613–631.
- Wilk M.B., Gnanadesikan R. (1968). Probability plotting methods for the analysis of data. *Biometrika* 55, 1–17.
- Wilk M.B., Gnanadesikan R., Huyett (1962). Estimation of parameters of the gamma distribution using order statistics. *Biometrika* 49, 525–545.

*Praca wpłynęła 10 września 1988;  
w wersji ostatecznej 9 września 1990*

### A note on applications of gamma probability plots in the multivariate analysis of variance

#### Summary

The paper presents different applications of gamma probability plots in the multivariate analysis of variance. The plots can be used to investigate contrasts of the vectors of main effects and the interaction effects with one degree of freedom in the analysis of factor design of type  $2^k$ , and also the main effects and the interaction effects in the factor design with equal degrees of freedom. They can also be used to investigate homogeneity of the covariance matrices with equal degrees of freedom.

**Key words:** gamma distribution, probability plots, factorial design, contrast, covariance matrix