Listy Biometryczne
Vol. XXIV (1987) No. 2
Biometrical Letters

ZASTOSOWANIE DWUSTOPNIOWEJ PROCEDURY STEINA W ANALIZIE WARIANCJI

Augustyn Markiewicz

Zakład Metod Matematycznych i Statystycznych,
Akademia Rolnicza, Wojska Polskiego 28, 60-637 Poznań

Markiewicz A., 1987. Application of Stein's two stage procedure in analysis of variance. Listy Biometryczne XXIV, z. 2, Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza, Poznań (Adam Mickiewicz University Press) pp. 47-54, PL ISSN 0458-0036.

The paper presents an application of a Stein two-stage procedure to an analysis of variance with unequal variances. This procedure gives an exact test independent of unknown variances with controllable power and in the case of multiple comparisons gives confidence intervals of controllable length. Properties of the test are presented and illustrated by a numerical example.

1. WSTĘP

Procedury testowe klasycznej analizy wariancji oparte są na założeniu normalności, niezależności i równości wariancji błędów. Badania odporności testu F wykazują, że odchylenia od normalności mają niewielki wpływ na poprawność wnioskowania o nieznanych średnich, natomiast odchylenia od niezależności lub równości wariancji wpływają dość znacznie na poprawność wnioskowania, szczególnie w przypadku niejednakowych liczebności prób.

W przypadku niejednakowych wariancji nie istnieje procedura jednostopniowa wielokrotnych porównań niezależna od nieznanych wariancji. Istnieją jedynie pewne przybliżone metody, których obszerny przegląd można znaleźć w pracach Hochberga (1976) i Tamhane (1979). Z drugiej strony należy stwierdzić, że jeśli nawet założenie o równości wariancji jest spełnione, to moc testu F zależy od nieznanej wspolnej wariancji, co utrudnia odpowiednie zaplanowanie eksperymentu.

W pracy przedstawiono dwustopniową procedurę analizy wariancji prowadzącą do dokładnego testu, którego moc nie zależy od nieznanych

---

Słowa kluczowe: analiza wariancji, nierówne wariancje, kontrolowana moc, wielokrotne porównania, przedziały ufności o stałej długości

wariancji i jest kontrolowana. W przypadku natomiast wielokrotnych porównań realizowane są one poprzez konstrukcję przedziałów ufności, których długości są niezależne od nieznanych wariancji i nie przekraczają z góry zadanej wielkości.

Procedurę dwustopniową wprowadził po raz pierwszy Stein (1945) przyjmując założenie o równości wariancji. Bishop i Dudewicz (1978), (1981) uogólnili tę procedurę na przypadek nierównych wariancji, a poza tym pokazali jej zastosowanie w wielozmiennej analizie wariancji. Hochberg (1975) i Bishop (1979) wykorzystali tę uogólnioną procedurę przy wielokrotnych porównaniach, natomiast w ogólnym modelu liniowym zastosował ją Bishop (1978). W wielowymiarowej analizie wariancji zastosowanie tej procedury pokazał Saborowski (1983).

2. PROCEDURA TESTOWA

Model klasyfikacji jednokierunkowej można przedstawić następująco :

$$X_{ij} = \mu_i + e_{ij} \qquad (i=1,2,\ldots,k;\ j=1,2,\ldots), \tag{1}$$

gdzie μ_i jest nieznaną średnią i-tej populacji, natomiast e_{ij} są niezależnymi zmiennymi losowymi, o których zakłada się, że e_{ij} $N(0,\sigma_i)$, gdzie $\sigma_i > 0$ są nieznane. Brak górnej granicy dla wskaźnika j w opisie modelu (1) oznacza, że z każdej populacji można pobrać próbę o dowolnie dużej, skończonej liczebności. Celem jest weryfikacja hipotezy H_o o równości k średnich populacyjnch μ_i,

$$H_o = \mu_1 = \mu_2 = \ldots = \mu_k.$$

Dwustopniowa procedura prowadząca do weryfikacji powyższej hipotezy H_o przebiega w następujący sposób.

Krok 1. Eksperymentator pobiera wstępne n_o (≥ 2) elementowe próbki losowe z każdej populacji. Na podstawie uzyskanych obserwacji oblicza się z każdej próbki nieobciążone oceny nieznanych wariancji σ_i^2,

$$s_i^2 = \sum_{j=1}^{n_o} (X_{ij} - \bar{X}_i)^2/(n_o - 1) \qquad (i=1,2,\ldots,k) \quad ,$$

gdzie $\bar{X}_i = \sum_{i=1}^{n_o} X_{ij}/n_o$. Następnie wyznacza wymaganą liczebność każdej z próbek losowych według następującego wzoru:

$$N_i = \max\{n_o + 1,\ [s_i^2/z] + 1\} \ , \qquad (i=1,2,\ldots,k) \quad , \tag{2}$$

gdzie $z>0$ jest uprzednio wybraną stałą (sposób wyboru z będzie przedstawiony później), natomiast symbol [a] oznacza największą liczbę całkowitą mniejszą od liczby rzeczywistej a.

Krok 2. Eksperymentator pobiera dodatkowo $N_i - n_o$ obserwacji z każdej populacji. Na podstawie N_i elementowych próbek oblicza uogólnione średnie próbkowe $\tilde{X}_i$ według wzoru podanego przez Steina (1945)

$$\tilde{X}_i = \sum_{j=1}^{N_i} a_{ij} X_{ij} \qquad (i=1,2,\dots,k) \quad , \tag{3}$$

gdzie a_{ij} spełniają warunki :

$$\sum_{j=1}^{N_j} a_{ij} = 1 \ , \quad a_{i1} = a_{i2} = \dots = a_{in_o} \ , \quad s_i^2 \sum_{j=1}^{N_i} a_{ij}^2 = z \ .$$

Przy wyznaczaniu średnich $\tilde{X}_i$ można wykorzystać współczynniki a_{ij} zaproponowane przez Bishopa i Dudewicza (1978) określone następująco:

$$a_{i1} = a_{i2} = \dots = a_{in_o} = [1 - (N_i - n_o) b_i]/n_o \quad ,$$

$$a_{in_o+1} = \dots = a_{iN_i} = b_i \quad ,$$

$$b_i = \{1 + [n_o(N_i z - s_i^2)/(N_i - n_o)s_i^2]^{1/2}\}/N_i \quad .$$

Do weryfikacji hipotezy H_o stosuje się teraz statystykę testową

$$\tilde{F} = \sum_{i=1}^{k} (\tilde{X}_i - \tilde{X}_.)^2/z \quad ,$$

gdzie $\tilde{X}_. = \sum_{i=1}^{k} \tilde{X}_i/k$. Wybór statystyki testowej jest uzasadniony interpretacją stałej z. Mianowicie, z odgrywa taką samą rolę jak σ^2/N w analizie wariancji ze znanymi i równymi wariancjami błędów σ^2 oraz z równymi liczebnościami prób N. W przypadku tym używa się statystyki testowej χ^2 będącej funkcją jedynie średnich próbkowych oraz σ^2/N. Statystyka testowa $\tilde{F}$ ma dokładnie tę samą postać z tym, że średnie próbkowe zastąpione są przez uogólnione średnie $\tilde{X}_i$ a σ^2/N przez z.

Hipoteza H_o zostaje odrzucona na poziomie istotności α wtedy i tylko wtedy, gdy

$$\tilde{F} > \tilde{F}^{\alpha}_{k,n_o} \quad ,$$

gdzie $\tilde{F}^{\alpha}_{k,n_o}$ jest określone następująco

$$P_{H_o}(\tilde{F} > \tilde{F}^{\alpha}_{k,n_o}) = \alpha \quad . \tag{4}$$

Przy założeniu prawdziwości hipotezy zerowej rozkład prawdopodobieństwa statystyki testowej $\tilde{F}$ jest taki sam jak rozkład zmiennej losowej $Q_o = \sum_{i=1}^{k}(t_i - \bar{t})^2$, gdzie t_i, i=1,2,...,k, są niezależnymi zmiennymi losowymi o identycznych rozkładach t-Studenta z n_o-1 stopniami swobody, a $\bar{t} = 1/k\sum_{i=1}^{k} t_i$. Własność ta wynika stąd, że każda zmienna losowa $t_i = (\tilde{X}_i - \mu_i)/z^{1/2}$ ma rozkład t-Studenta z n_o-1 stopniami swobody, co zostało wykazane przez Steina (1945), jak również przez Zacksa (1971) oraz Ghosha (1975). A zatem statystyka $\tilde{F}$ ma taki sam rozkład jak zmienna losowa

$$Q = \sum_{i=1}^{k} [t_i - \bar{t} + (\mu_i - \bar{\mu})/z^{1/2}]^2 , \qquad (5)$$

a przy założeniu prawdziwości hipotezy zerowej, zgodnie z którą $\mu_i = \bar{\mu}$ dla i=1,2,...,k, rozkłady statystyki $\tilde{F}$ i zmiennej losowej Q_o są takie same.

Z postaci rozkładu zmiennej losowej Q wynika, że zarówno rozkład statystyki testowej $\tilde{F}$ jak i moc testu nie zależą od nieznanych wariancji σ_i^2. Ponadto, co zostało wykazane np. w pracy Bishopa i Dudewicza (1981), moc testu jest kontrolowana poprzez dobór stałej z. Mianowicie, jeżeli $\mu_i \neq \bar{\mu}$ dla pewnego i=1,2,...,k, wtedy

$$\lim_{z \to 0} P(\tilde{F} > \tilde{F}^{\alpha}_{k,n_o}) = 1 .$$

W oparciu o rozkład zmiennej losowej Q określonej wzorem (5) można tak dobrać stałą z, że dla zadanej alternatywy określonej poprzez wektor $(\mu_1, \mu_2, \ldots, \mu_k) \neq (\bar{\mu}, \bar{\mu}, \ldots, \bar{\mu})$, moc testu będzie równa z góry zadanej wartości β, czyli takie z, że

$$P(Q > \tilde{F}^{\alpha}_{k,n_o}) = \beta . \qquad (6)$$

Procedury dokładnego wyznaczania dystrybuant oraz kwantyli rozkładów zmiennych losowych Q i Q_o nie są dotychczas znane. Natomiast znanych jest kilka aproksymacji rozkładu zmiennej losowej Q_o, któryych przegląd można znaleźć w pracy Bishopa i in. (1978). Hochberg (1975) zaproponował aproksymację rozkłdu zmiennej losowej Q_o rozkładem zmiennej losowej $lF_{l,m}$, gdzie $F_{l,m}$ jest zmienną losową o centralnym rozkładzie F Snedecora z l i m stopniami swobody, przy czym l i m są tak dobrane, że

$$E(Q_o) = E(lF_{l,m}) \quad \text{oraz} \quad E(Q_o^2) = E[(lF_{l,m})^2] .$$

Bishop i Dudewicz (1978) przedstawili aproksymację rozkładu zmiennej losowej Q rozkładem zmiennej losowej

$$[(n_o-1)/(n_o-3)]\chi^2_{k-1}(\Delta) \quad , \text{ gdzie } \Delta = \sum_{i=1}^{k} (\mu_i - \bar{\mu})^2/z \quad ,$$

która dla $\Delta=0$ jest nieco mniej dokładna od aproksymacji Hochberga, ale jest bardziej ogólna i wygodniejsza w użyciu. W przypadku zastosowania tej aproksymacji można tak dobierać stałą z, żeby moc testu była równa β dla alternatyw takich, że

$$\sum_{i=1}^{k} (\mu_i - \bar{\mu})^2 = \delta > 0 .$$

Stałą z wyznacza się wtedy w oparciu o wzór (6) posługując się np. tablicą 25 ze zbioru Pearsona i Hartleya (1972).

3. WIELOKROTNE PORÓWNANIA

Dla modelu (1) można wyprowadzić jednoczesne przedziały ufności o z góry zadanej długości dla wszystkich kontrastów stosując dwustopniową procedurę pobierania próby opisaną w części pierwszej. Na podstawie prac Hochberga (1975) oraz Bishopa (1979) rodzina co najmniej $100(1-\alpha)\%$ jednoczesnych przedziałów ufności typu Scheffe'go dla wszystkich kontrastów $\sum_{i=1}^{k} c_i\mu_i$ ma postać

$$\left\{ \sum_{i=1}^{k} c_i\tilde{X}_i \pm (z \sum_{i=1}^{k} c_i^2 \tilde{F}^{\alpha}_{k,n_o})^{1/2} \quad , \forall (c_1,\dots,c_k)' \in R^k \quad i \quad \sum_{i=1}^{k} c_i = 0 \right\} ,$$

gdzie $\tilde{X}_i$ i $\tilde{F}^{\alpha}_{k,n_o}$ określone są wzorami (3) i (4). Dla zapewnienia zadanej długości $2d$ przedziału ufności dla kontrastu $\sum_{i=1}^{k} c_i\mu_i$ stałą z dobiera się tak, żeby $d^2 = z \sum_{i=1}^{k} c_i^2 \tilde{F}^{\alpha}_{k,n_o}$

W przypadku porównywania średnich parami można posłużyć się jednoczesnymi przedziałami ufności typu Tukey'a. Rodzina $100(1-\alpha)\%$ jednoczesnych przedziałów ufności dla wszystkich różnic par średnich $\mu_i-\mu_j$ jest następująca:

$$\{\tilde{X}_i-\tilde{X}_j \pm g^{\alpha}_{k,n_o-1} z^{1/2} \quad , \quad 1\le i<j\le k \} \quad ,$$

gdzie g_{k,n_o-1} jest $(1-\alpha)$-tym kwantylem rozkładu g_{k,n_o-1} rozstępu k niezależnych zmiennych losowych o rozkładach t-Studenta z n_o-1 stopniami swobody. Tablice wartości g^{α}_{k,n_o-1} podał Hochberg (1976a), natomiast Wilcox (1983) zamieścił ich rozszerzoną wersję oraz wskazał na dobrą aproksymację rozkładu zmiennej losowej g_{k,n_o-1} odpowiednim rozkładem studentyzowanego rozstępu (definicja, tablice patrz np. Miller 1966). Dla zadanej z góry długości przedziałów ufności równej $2d$ stałą z należy dobrać tak, żeby $z = (g^{\alpha}_{k,n_o-1})^2/d^2$.

Rodziny jednoczesnych przedziałów ufności dla wszystkich funkcji liniowych $\sum_{i=1}^{k} l_i\mu_i$, gdzie $(l_1,l_2,\dots,l_k)' \in R^k$, można skonstruować w oparciu o pracę Hochberga (1976) - przedziały typu Tukeya, oraz pracę Tamhane (1977) - przedziały typu Scheffe'go.

4. WLASNOŚCI PROCEDURY

Liczebności prób N_i pobieranych z poszczególnych k populacji są, jak to wynika ze wzoru (2), zmiennymi losowymi, których rozkłady prawdopodobieństwa zależą od liczebności prób wstępnych n_o, od stałej z oraz od nieznanych wariancji. Wzory umożliwiające wyznaczenie rozkładu prawdopodobieństwa oraz wartości oczekiwanej końcowej liczebności próby N podał Stein (1945). Moshman (1958) na podstawie wzorów Steina zaproponował metodę doboru liczebności próby wstępnej n_o opartą na uwzględnieniu wpływu n_o na wartość oczekiwaną końcowej liczebności próby E(N) oraz na rozproszenie zmiennej losowej N mierzone najmniejszą liczbą całkowitą N_p taką, że $P(N \le N_p) \ge p$, dla różnych wartości z, σ^2 oraz p. Cohen i Sackrowitz (1984) przedstawili metodę doboru n_o dla różnych wartości σ^2 w oparciu o ryzyko uwzględniające koszt pobrania obserwacji. W przypadku braku informacji a priori o σ^2 można przyjąć na podstawie powyższych prac oraz pracy Bishopa i Dudewicza (1981), że najbardziej odpowiednia liczebność próby wstępnej powinna wynosić co najmniej 15, zwłaszcza gdy wariancja σ^2 nie jest zbyt mała.

Bishop i Dudewicz (1981) pokazali, że procedura ta może być stosowana w przypadku nierównych wariancji σ_i^2, jak również w przypadku, gdy wariancje są równe, porównując ją numerycznie z odpowiednimi procedurami jednostopniowymi. Stwierdzili również, że dwustopniowe pobieranie próby jest mniej efektywne od metod sekwencyjnch w sensie końcowej liczebności próby, ale procedury sekwencyjne nie zapewniają np. uzyskania przedziału ufności o zadanej długości i na dokładnie określonym poziomie istotności. Ponadto, w praktyce, możliwości stosowania procedur sekwencyjnych są bardziej ograniczone.

Ramkaran (1983) badając dwustopniowe procedury typu Steina wykazał ich odporność na umiarkowane odstępstwa od założenia normalności.

5. PRZYKŁAD

Celem eksperymentu było porównanie średnich wysokości czterech odmian żyta w końcowej fazie wzrostu, tzn. weryfikacja hipotezy

$$H_o = \mu_1 = \mu_2 = \mu_3 = \mu_4 \quad .$$

Z powodu niemożliwości przyjęcia założenia o równości wariancji zdecydowano się na użycie procedury dwustopniowej, przyjmując poziom istotności $\alpha = 0{,}05$ oraz liczebność próby wstępnej $n_o = 15$. Eksperymentator zdecydował ponadto, że w przypadku niespełnienia H_o, mierzonego wielkością $\delta = \sum_{i=1}^{4} (\mu_i - \bar{\mu})^2 \ge 1$, moc testu β winna wynosić co najmniej 0,9. Oznacza to, że gdyby zróżnicowanie prawdziwych średnich wysokości, mierzone wielkością δ, było większe niż 1, wtedy prawdopodobieństwo odrzucenia hipotezy H_o byłoby równe co najmniej 0,9. Stała z odpowiadająca przyjętym założeniom wynosi 0,06 natomiast wartość krytyczna testu $\tilde{F}_{4,15}^{0,05} = 9{,}75$. Wartości te zostały odczytane z tablic Bishopa i Dudewicza (1978) dla $\beta = 0{,}896$.

Eksperyment przebiegał w następujący sposób.

W pierwszym etapie pobrano losowo po 15 roślin każdej odmiany spośród dużej liczby roślin rosnących na czterech poletkach. Zmierzono ich wysokości z dokładnością do jednego milimetra i wyznaczono liczebności N_i ze wzoru (2).

W drugim etapie pobrano losowo z każdego poletka odpowiednio $N_i - n_o$ roślin i zmierzono ich wysokości.

Uzyskane wartości N_i, s_i^2, a_i, b_i oraz $\tilde{X}_i$ przedstawia poniższa tabela:

i	N_i	s_i^2[dcm^2]	a_i	b_i	$\tilde{X}_i$[dcm]
1	60	3,594	0,01545	0,01707	14,871
2	73	4,337	0,01102	0,01439	14,874
3	29	1,695	0,02905	0,04030	13,648
4	31	1,851	0,02991	0,03444	11,703

Wartość statystyki testowej $\tilde{F}$ wyniosła 111,9696, a zatem hipoteza H_o została odrzucona. Po odrzuceniu hipotezy zerowej dla wybranych kontrastów zostały wyznaczone jednoczesne przedziały ufności typu Scheffe'go. W celu wskazania, które średnie różnią się istotnie przedziały te zostały wyzznaczone na podstawie tej samej próby i dla tych samych wartości parametrów z oraz α. Kontrasty oraz uzyskane dla nich przedziały ufności podane są poniżej.

$\mu_1 + \mu_2 - \mu_3 - \mu_4$ (2,864 ; 5,924)
$\mu_1 - \mu_2$ (-1,085 ; 1,078)
$\mu_1 - \mu_3$ (0,141 ; 2,305)
$\mu_3 - \mu_4$ (0,863 ; 3,026) .

Przykład został opracowany na podstawie danych zebranych w Ogrodzie Botanicznym PAN.

LITERATURA

Bishop, T.A. (1978). A Stein two - sample procedure for the general linear model with unequal error variances. *Commun.Statist.*, A7(5), 495-507.

Bishop, T.A. (1979). Some results on simultaneous inference for analysis of variance with unequal variances. *Technometrics* 21, 337-340.

Bishop, T.A., Dudewicz, E.J. (1978). Exact analysis of variance with unequal variances. *Technometrics* 20, 419-430.

Bishop, T.A., Dudewicz, E.J. (1981). Heteroscedastic ANOVA. *Sankyā* ser. B 43, 40-57.

Bishop, T.A., Dudewicz, E.J., Juritz, J., Stephens, M.A. (1978). Percentage points of quadratic forms in Student t variates. *Biometrika* 65, 435-439.

Cohen, A., Sackrowitz, H.B. (1984). Bayes double sample estimation procedures. *Ann. Statist.* 12, 1035-1049.

Ghosh, B.K. (1975). A two-stage procedure for the Behrens-Fisher problem.

J. Amer. Statist. Assoc. 70, 457-462.

Hochberg, Y. (1975). Simultaneous inference under Behrens-Fisher conditions - a two sample approach. *Commun. Statist.* A 4 (12), 1109-1119.

Hochberg, Y. (1976). A modification of the T - method of multiple comparisons for a one-way layout with unequal variances. *J. Amer. Statist. Assoc.* 71, 200-203.

Miller, R.G. (1966). *Simultaneous statistical inference.* McGraw-Hill, New York.

Moshman, J. (1958). A method for selecting the size of the initial sample in Stein's two sample procedure. *Ann. Math. Statist.* 29, 1271-1275.

Pearson, E.S., Hartley, H.O. (1972). *Biometrika tables for statisticians.* Vol. II. University Press, Cambridge.

Ramkaran (1983). The robustness of Stein's two-stage procedure. *Ann. Statist.* 11, 1251-1256.

Tamhane, A.C. (1977). Multiple comparisons in model I one-way ANOVA with unequal variances. *Commun. Statist.* A 6 (1), 15-32.

Tamhane, A.C. (1979). A comparison of procedures for multiple comparisons of means with unequal variances. *J. Amer. Statist. Assoc.* 74, 471-480.

Saborowski, J. (1983). Stein's two-stage test in linear models. *Biom. J.* 25, 459-467.

Stein, C. (1945). A two-sample test for a linear hypothesis whose pover is independent of the variance. *Ann. Math. Statist.* 16, 243-258.

Wilcox, R.R. (1983). A table of percentage points of the range of independent t variables. *Technometrics* 25, 201-204.

Zacks, S. (1971). *The theory of statistical inference.* John Wiley, New York.

Praca wpłynęła 5 grudnia 1986 :

w wersji ostatecznej 17 listopada 1987.