

## **Krzywe wzrostu – przegląd metod i nowe problemy**

**Mirosława Wesołowska-Janczarek**

Instytut Zastosowań Matematyki, Akademia Rolnicza,  
20-950 Lublin, Akademicka 13

### **Streszczenie**

Praca zawiera przegląd znanych metod krzywych wzrostu oraz tematyki najnowszych badań w tej dziedzinie. Ukazujące się w ostatnich latach prace rozszerzają znane metody w kierunku zmiany założeń, a więc poszerzają możliwości ich zastosowań oraz sugerują powiązania tych zagadnień z innymi, na przykład z procesami stochastycznymi.

### **1. Wstęp**

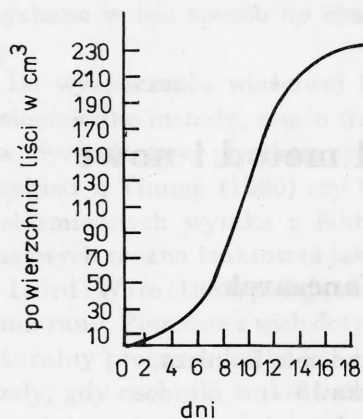
Wzrost jest jednym z najbardziej podstawowych procesów biologicznych obserwowanych zarówno u poszczególnych osobników, jak i całych populacji. Matematyczny opis przebiegu tego procesu interesował eksperymentatorów z różnych dziedzin nauki jak na przykład biologii, medycyny, genetyki, biochemii, biofizyki, a także ekonometrii czy też demografii.

Pierwszą publikacją zawierającą opis matematyczny wzrostu jak podaje Rash (1988) była praca T. R. Malthusa "Essay on the principle of population", London, 1798, natomiast pierwszym, który sugerował użycie kryterium najmniejszych kwadratów dla oceny współczynników w dopasowywaniu liniowej krzywej był Legendre w 1805 roku.

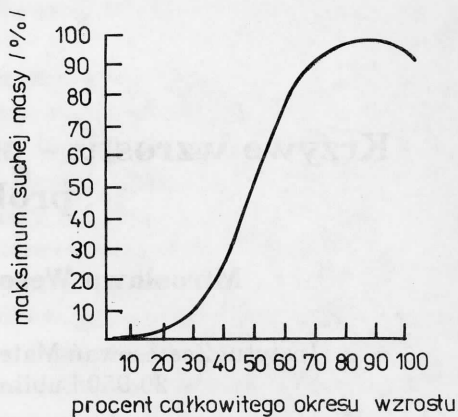
Od tego czasu opis przebiegu wielu procesów w czasie (w tym także wzrostu), a w szczególności metod ich analizy, był tematem szeregu badań i publikacji.

Przebieg wzrostu roślin jest dobrze znany przyrodnikom. P.F. Wareing i I.D.J. Philips w książce "Wzrost i różnicowanie się roślin", PWN, W-wa 1985 podają

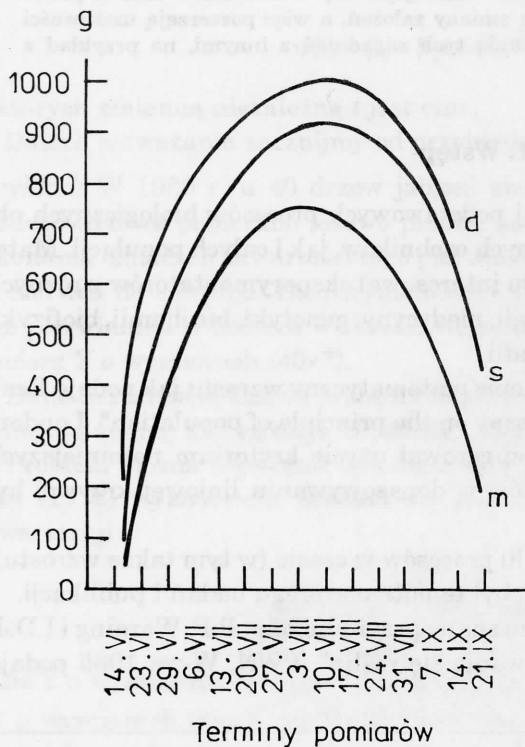
przebieg wzrostu roślin na przykładzie wzrostu liści ogórka (rys. 1) i przyrostu suchej masy roślin jęczmienia podczas okresu wegetacji (rys. 2).



**Rys. 1.** Krzywa wzrostu powierzchni liścia ogórka



**Rys. 2.** Przyrost suchej masy roślin jęczmienia podczas okresu wegetacji



W pracy St. Roztropowicza "Zmiany w strukturze i warunkach siedliska łąnu ziemniaków powodowane zmienną obsadą roślin" opublikowanej w materiałach konferencji na temat "Obsada i produktywność roślin uprawnych", Puławy 1988, opisana jest dynamika wzrostu roślin ziemniaka i zamieszczony rysunek przedstawiający przyrost masy nadziemnej dla roślin wyrastających z sadzoniaków małych (m), średnich (s) i dużych (d).

**Rys. 3.** Dynamika wzrostu roślin ziemniaka wyrastających z sadzoniaków małych (m), średnich (s) i dużych (d)

Jak wynika z zamieszczonych powyżej rysunków przebieg zmian w czasie może być opisany różnego typu funkcją. Na rysunku 3 są to wielomiany stopnia drugiego, natomiast wykresy na rys. 1 i 2 są szczególnymi przypadkami krzywych Richardsa, które wyrażają się wzorem [patrz: Encyclopedia of Statistical Sciences (1982) Vol.8, John Wiley, str. 128]

$$f(t) = a(1 + sbc^t)^\lambda \quad \text{dla } t > 0,$$

gdzie  $s = \pm 1$ ,  $a, b, c \in R_+$ , przy czym dla  $s = +1$ ,  $\lambda < 0$ , a dla  $s = -1$   $\lambda \geq 0$  ( $t$  jest zmienną określającą czas). Szczególnymi przypadkami tych krzywych są:

zmodyfikowana wykładnicza  $f(t) = a(1 - bc^t)$  ,

logistyczna  $f(t) = \frac{a}{1 + bc^t}$  ,

Gompertz'a  $f(t) = ae^{-bc^t}$  .

W literaturze statystycznej problem krzywych wzrostu obejmuje opisy procesów, które dadzą się wyrazić wielomianami dowolnego stopnia i mieści się w grupie modeli liniowych. Opis biologicznych procesów przy użyciu krzywych Richardsa można znaleźć w książce Causton i Venus (1981), natomiast problemy dotyczące innych typów krzywych można znaleźć u Rasha (1988), czy też w książce Barda (1974) dotyczącej nieliniowej estymacji.

Dalsze rozważania w tej pracy dotyczą tylko krzywych wzrostu opisanych wielomianami.

## 2. Metoda wielomianowych krzywych wzrostu dla danych niesklasyfikowanych

Opisu matematycznego przebiegu procesu wzrostu dokonuje się przez dopasowanie krzywej na podstawie wyników pomiarów uzyskanych w badaniach empirycznych. Dopasowanie modelu natomiast wymaga estymacji parametrów i ustalenia stopnia wielomianu.

Krzywe wzrostu można rozważać dla poszczególnych osobników lub dla całej populacji.

W pierwszym przypadku obserwacje stanowią ciąg  $y_t$  pomiarów cechy  $Y$  w poszczególnych chwilach  $t$ . Szukana krzywa wzrostu jest funkcją regresji względem  $t$

$$y_t = y(t) + e_t .$$

Estymacji dokonuje się najczęściej metodą najmniejszych kwadratów. Każdy taki ciąg obserwacji jest szeregiem czasowym (Wójcik, Baudański, 1989).

Estymacja krzywej wzrostu dla populacji wymaga wykonania pomiarów badanej cechy u  $n$  jednostek powtarzanych w  $p$  punktach czasowych  $t_1, t_2, \dots, t_p$ . Uzyskane w ten sposób  $np$  obserwacji ustawia się w macierz  $Y$  o wymiarach  $n \times p$ .

Do wyznaczenia właściwej krzywej wzrostu w tym przypadku stosuje się wielozmienne metody, często traktowane jako szczególny przypadek metod stosowanych do tzw. powtarzanych pomiarów (repeated measurements) jak na przykład u Timma (1980) czy Crowdera i Handa (1991). Zastosowanie metod wielozmiennych wynika z faktu, że pomiary tej samej cechy w  $p$  punktach czasowych można traktować jako  $p$  cech, które są wzajemnie skorelowane.

Laird i Ware (1982) podają różnice między krzywymi wzrostu, a powtarzaniem pomiarów. Pierwsze z nich dotyczą przypadków, gdy badany i porównywany jest naturalny proces wzrostu bez interwencji w jego przebieg. Drugie rozważa się wtedy, gdy osobniki traktowane są różnymi zabiegami czy też warunkami, co powoduje zmiany wartości cechy w czasie i po ustaniu ich działania cechy te dążą do ustabilizowania swoich wartości. Nie jest to jednak sztywna zasada.

Wielozmienną metodę krzywych wzrostu stosuje się w przypadku, gdy dopasowywane krzywe są wielomianami stopnia  $q-1$  postaci

$$\psi(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_{q-1} t^{q-1}, \quad (2.1)$$

w których zmienną niezależną  $t$  jest czas.

Dalsze rozważania zaczniemy od przykładu:

*Przykład:* W 1989 r. u 40 drzew jabłoni zbadano zawartość Ca w liściach. Z każdego drzewa pobierano losowo próbkę liści z różnych części korony, w celu dokonania oznaczeń zawartości tego pierwiastka, w 7 równoodległych terminach od czerwca do sierpnia. Eksperymentatora interesowała dynamika zawartości tego pierwiastka w liściach w okresie wegetacji. Uzyskane wyniki zestawiono w macierz  $Y$  o wymiarach  $(40 \times 7)$ .

Dane te stanowią losową próbę  $n=40$  jednostek eksperymentalnych, które nie podlegają żadnej klasyfikacji. U każdej z nich wykonano  $p=7$  pomiarów.

Problem można rozwiązać opierając się na metodzie podanej przez Kshirsagara (1972) wyznaczając średnie dla poszczególnych terminów oraz macierz kowariancji

$$\bar{y}_{p1} = \frac{1}{n} Y'E \quad \text{oraz} \quad S_{pp} = Y'(I_n - \frac{1}{n}E)Y, \quad (2.2)$$

gdzie  $Y$  o wymiarach  $(n \times p)$  jest macierzą obserwacji,  $I_n$  jest macierzą jednostkową o wymiarach  $(n \times n)$ , zaś  $E$  jest macierzą złożoną z samych jedynek o odpowiednich wymiarach.

Przy założeniu normalności rozkładu

$$\sqrt{n} \mathbf{y} \sim N_p(\sqrt{n} \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}), \quad (2.3)$$

gdzie

$$\mathbf{X}_{p,q} = \begin{pmatrix} 1 & t_1 & \dots & t_1^{q-1} \\ 1 & t_2 & \dots & t_2^{q-1} \\ \cdot & \cdot & \dots & \cdot \\ 1 & t_p & \dots & t_p^{q-1} \end{pmatrix} \quad \text{i} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_{q-1} \end{pmatrix}, \quad (2.4)$$

$\mathbf{S}$  ma  $p$ -wymiarowy rozkład Wisharta

$$\mathbf{S} \sim W_p(n-1, \boldsymbol{\Sigma}) \quad (2.5)$$

oraz  $\bar{\mathbf{y}}$  i  $\mathbf{S}$  są niezależne. Estymator  $\hat{\boldsymbol{\beta}}$  można wyznaczyć minimalizując wyrażenie

$$\text{Min}_{\boldsymbol{\beta}} (\bar{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})' \mathbf{S}^{-1} (\bar{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}) \quad (2.6)$$

jako

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}^{-1}\bar{\mathbf{y}}. \quad (2.7)$$

Zawiera on estymatory współczynników w równaniu krzywej wzrostu (2.1).

Przedziały ufności dla poszczególnych współczynników lub ich kombinacji liniowej  $\mathbf{b}'\boldsymbol{\beta}$  mają postać:

$$\mathbf{b}'\hat{\boldsymbol{\beta}} \pm \left\{ \left( 1 + \frac{T_{p-q}^2}{f} \right) \frac{v}{f - (p-q)} F_{\alpha; 1, f-p+q} \right\}^{1/2}, \quad (2.8)$$

gdzie

$$\frac{T_{p-q}^2}{f} = n(\bar{\mathbf{y}} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{S}^{-1} (\bar{\mathbf{y}} - \mathbf{X}\hat{\boldsymbol{\beta}}); \quad f = n-1; \quad q = r(\mathbf{X}); \quad \frac{1}{f}v = \mathbf{b}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{b}$$

przy

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{f}\mathbf{S}.$$

Wreszcie obszar ufności dla  $\psi(t)$  dany jest przez

$$\hat{\psi}(t) \pm \left\{ \left( 1 + \frac{T_{p-q}^2}{f} \right) \frac{vq}{n-p} F_{\alpha; q, n-p} \right\}^{1/2}, \quad (2.9)$$

gdzie  $\hat{\psi}(t) = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2 + \dots + \hat{\beta}_{q-1} t^{q-1}$ .

Granice ufności są wyznaczone właściwie tylko wtedy, gdy dopasowany model jest adekwatny, co można sprawdzić testując hipotezę  $H_0: E(\bar{\mathbf{y}}) = \mathbf{X}\boldsymbol{\beta}$  przeciw  $H_1: E(\bar{\mathbf{y}}) \neq \mathbf{X}\boldsymbol{\beta}$  za pomocą funkcji testowej:

$$F_{p-q, f-(p-q)+1} = \frac{f-(p-q)+1}{p-q} \cdot \frac{T_{p-q}^2}{f} . \quad (2.10)$$

W rozważanym tu przykładzie uzyskano:

$$\bar{\mathbf{y}} = [1.00 \quad 1.04 \quad 1.10 \quad 1.11 \quad 1.15 \quad 1.20 \quad 1.20] ,$$

$$\mathbf{S} = \begin{pmatrix} 0.0738 & 0.0777 & 0.0815 & 0.0769 & 0.0765 & 0.0979 & 0.0638 \\ 0.0777 & 0.0969 & 0.0931 & 0.0855 & 0.0873 & 0.1069 & 0.0672 \\ 0.0815 & 0.0931 & 0.1150 & 0.0953 & 0.1029 & 0.1137 & 0.0742 \\ 0.0769 & 0.0855 & 0.0953 & 0.0977 & 0.0934 & 0.1144 & 0.0752 \\ 0.0765 & 0.0873 & 0.1029 & 0.0934 & 0.1060 & 0.1141 & 0.0684 \\ 0.0979 & 0.1069 & 0.1137 & 0.1144 & 0.1141 & 0.1545 & 0.0877 \\ 0.0638 & 0.0672 & 0.0742 & 0.0752 & 0.0686 & 0.0877 & 0.1356 \end{pmatrix} .$$

Przy stałych odcinkach czasowych przyjmujemy  $t_1 = 1, \dots, t_7 = 7$ .

Niech  $q=3$ , czyli dopasowywana krzywa jest stopnia drugiego. Wtedy macierz  $\mathbf{X}$  jest postaci:

$$\mathbf{X}' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 4 & 9 & 16 & 25 & 36 & 49 \end{pmatrix} .$$

Otrzymujemy:  $\hat{\boldsymbol{\beta}}' = [0.9143 \quad 0.0459 \quad -0.0012]$  i sprawdzamy testem adekwatności krzywej:

$$\frac{T_{p-q}^2}{f} = 2.6584; \quad \text{zatem } F_{\text{obl}} = 23.926, \quad \text{przy } F_{0.05; 4, 36} = 2.6335.$$

$H_0$  odrzucamy, obniżamy stopień wielomianu, przyjmując  $q=2$  i dopasowujemy funkcję liniową.  $\bar{\mathbf{y}}$  i  $\mathbf{S}$  są takie same, natomiast

$$\hat{\boldsymbol{\beta}}' = [0.9962 \quad 0.0299] .$$

Tym razem  $F_{\text{obl}} = 0.008$  przy  $F_{0.05; 5, 35} = 2.4851$ , co potwierdza wybór modelu

$$\psi(t) = 0.9962 + 0.0299t .$$

Jeżeli odległości między punktami czasowymi są stałe, to elementami macierzy  $\mathbf{A}$  mogą być współczynniki wielomianów ortogonalnych.

Przedstawiona tu metoda może być stosowana, gdy badane jednostki eksperymentalne nie podlegają żadnej klasyfikacji. W praktycznych przypadkach ten

warunek nie zawsze jest spełniony. W przykładzie zilustrowanym na rys. 3 badane ziemniaki należały do jednej z trzech grup. Problem polegał na wyznaczeniu krzywych wzrostu dla każdej z grup oraz porównaniu tych krzywych dla rozpatrywanych grup. Rozwiązanie tego problemu można uzyskać stosując metody przedstawione poniżej.

### 3. Ogólny model krzywych wzrostu. Estymacja parametrów i testowanie hipotez

Ogólny wielozmienny model liniowy krzywych wzrostu podali Potthoff i Roy (1964) w postaci

$$\mathbf{Y} = \mathbf{XBT} + \mathbf{E} \quad , \quad (3.1)$$

gdzie  $\mathbf{Y}$  o wymiarach  $(n \times p)$  jest macierzą obserwacji,  $\mathbf{X}$  o wymiarach  $(n \times m)$  jest znaną macierzą układu (bez kolumny jedynek) taką, że  $r(\mathbf{X}) = m$ ,  $\mathbf{B}$  jest macierzą nieznaną stałych parametrów (współczynników w równaniach krzywych) o wymiarach  $(m \times q)$ ,  $\mathbf{T}$  o wymiarach  $(q \times p)$  jest znaną macierzą określającą wewnętrzną strukturę obserwacji ( $r(\mathbf{T}) = q$ ,  $q \leq p$ ), a  $\mathbf{E}$  o wymiarach  $(n \times p)$  jest macierzą błędów losowych.

Model ten może być potraktowany jako połączenie dwóch modeli

1°. dla układu eksperymentalnego

$$\mathbf{y}_1 = \mathbf{X} \begin{matrix} \beta_1^0 \\ \beta_{m1}^0 \end{matrix} + \mathbf{e}_1 \quad , \quad (3.1a)$$

gdzie  $\mathbf{y}_1$  – wektor kolumnowy pojedynczych pomiarów dla  $n$  jednostek eksperymentalnych;  $\mathbf{X}$  – macierz układu,  $\beta_1^0$  – wektor nieznaną parametrów ( $\beta_1^0 = [\beta_{11}^0, \beta_{12}^0, \dots, \beta_{1m}^0]$ ),  $\mathbf{e}_1$  – wektor błędów losowych.

2°. dla regresji

$$\mathbf{y}_2 = \begin{matrix} \beta_2^0 \\ \beta_{1q}^0 \end{matrix} \mathbf{T} + \mathbf{e}_2 \quad , \quad (3.1b)$$

gdzie  $\mathbf{y}_2$  – wektor wierszowy obserwacji u tej samej jednostki eksperymentalnej w  $p$  punktach czasowych,  $\beta_2^0$  – wektor współczynników w równaniu krzywej, ( $\beta_2^0 = [\beta_{20}^0, \beta_{21}^0, \dots, \beta_{2,q-1}^0]$ ),  $\mathbf{T}$  jest macierzą typu Vandermonde'a postaci

$$\mathbf{T}_{qp} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ t_1 & t_2 & \dots & t_p \\ t_1^2 & t_2^2 & \dots & t_p^2 \\ \cdot & \cdot & \dots & \cdot \\ t_1^{q-1} & t_2^{q-1} & \dots & t_p^{q-1} \end{pmatrix} \quad , \quad (3.2)$$

natomiast  $\mathbf{e}_2$  jest wektorem błędów losowych.

Łącząc je otrzymamy model postaci

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta}_1^0 \boldsymbol{\beta}_2^0 \mathbf{T} + \mathbf{E} ,$$

gdzie macierz

$$\boldsymbol{\beta}_1^0 \boldsymbol{\beta}_2^0 = \begin{pmatrix} \beta_{11}^0 \beta_{20}^0 & \beta_{11}^0 \beta_{21}^0 & \dots & \beta_{11}^0 \beta_{2,q-1}^0 \\ \beta_{12}^0 \beta_{20}^0 & \beta_{12}^0 \beta_{21}^0 & \dots & \beta_{12}^0 \beta_{2,q-1}^0 \\ \vdots & \vdots & \dots & \vdots \\ \beta_{1m}^0 \beta_{20}^0 & \beta_{1m}^0 \beta_{21}^0 & \dots & \beta_{1m}^0 \beta_{2,q-1}^0 \end{pmatrix} .$$

Wprowadzając oznaczenie

$$\mathbf{B} = \begin{pmatrix} \beta_{10} & \beta_{11} & \dots & \beta_{1,q-1} \\ \beta_{20} & \beta_{21} & \dots & \beta_{2,q-1} \\ \vdots & \vdots & \dots & \vdots \\ \beta_{m0} & \beta_{m1} & \dots & \beta_{m,q-1} \end{pmatrix} \quad (3.3)$$

otrzymujemy model w postaci (3.1), dla którego przyjmuje się założenia:

$$\mathbf{E}(\mathbf{Y}) = \mathbf{XBT} , \quad (3.4)$$

wiersze macierzy  $\mathbf{E}$  są wzajemnie niezależne i każdy jest  $p$ -wymiarowym wektorem o  $p$ -wymiarowym identycznym rozkładzie normalnym z macierzą kowariancji  $\boldsymbol{\Sigma}_{pp}$  i wartością oczekiwaną  $\mathbf{0}$  czyli  $\mathbf{e}_i \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$ ;  $i=1,2,\dots,n$ , macierz  $\boldsymbol{\Sigma}$  jest dodatnio określona.

Model (3.1) można stosować gdy:

1. Jednostki eksperymentalne stanowią jednorodną grupę i nie są traktowane żadnymi zabiegami (jak w przykładzie powyżej). Wtedy  $\mathbf{X}' = [1, 1, \dots, 1]$ ,  $\mathbf{B} = [\beta_0, \beta_1, \dots, \beta_{q-1}]$ , a macierz  $\mathbf{T}$  ma postać daną w (3.2).

2. Jednostki eksperymentalne podzielone są na  $m$  grup niekoniecznie równolicznych (podlegają jednokierunkowej klasyfikacji). Wtedy  $\mathbf{X}$  jest macierzą o wymiarach  $(n \times m)$ , zaś  $\mathbf{B}$  i  $\mathbf{T}$  określone są odpowiednio do (3.3) i (3.2). Przykład znajduje się w pracy Baksalary, Caliński, Kala (1980).

3. Jednostki eksperymentalne podlegają klasyfikacji dwukierunkowej. Wtedy  $\mathbf{X}$  jest macierzą układu,

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_1 \\ \dots \\ \mathbf{B}_2 \end{pmatrix} ,$$

$\mathbf{T}$  jest jak w (3.2).  $\mathbf{B}_1$  zawiera współczynniki krzywych dla pierwszej klasyfikacji, zaś  $\mathbf{B}_2$  dla drugiej klasyfikacji. Przykład podano w pracy Baksalary, Caliński, Kala (1983).



4. Model wieloreakcyjny tzn. uwzględniający kilka cech jednocześnie, mierzonych u tych samych  $n$  jednostek w tych samych  $p$  punktach czasowych. Każda z cech może być opisana wielomianem innego stopnia. Na przykład biorąc pod uwagę dwie cechy przyjmuje się  $\mathbf{Y} = [\mathbf{Y}_1 : \mathbf{Y}_2]$ , gdzie  $\mathbf{Y}_1$  jest macierzą obserwacji dla pierwszej cechy, a  $\mathbf{Y}_2$  dla drugiej,  $\mathbf{X}$  – macierz układu,  $\mathbf{B} = [\mathbf{B}_1 : \mathbf{B}_2]$ , przy  $q = q_1 + q_2$  oraz

$$\mathbf{T}_{q,2p} = \begin{pmatrix} \mathbf{T}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_2 \end{pmatrix},$$

gdzie  $\mathbf{T}_1$  i  $\mathbf{T}_2$  są postaci (3.2).

Zastosowanie modelu (3.1) podane w punktach 3 i 4 można rozszerzyć na dowolną klasyfikację i układ eksperymentalny.

Podstawowe zagadnienia rozważane dla ogólnego modelu krzywych wzrostu to:

1°. Estymacja macierzy parametrów  $\mathbf{B}$ , co jest równoznaczne z dopasowaniem krzywych,

2°. Testowanie wielozmiennych hipotez obejmujące między innymi ustalenie stopnia wielomianu i porównywanie krzywych dla różnych grup. Ogólna postać wielozmiennej hipotezy jest

$$H_0 : \mathbf{ABV} = \mathbf{0} \text{ przeciw } H_1 : \mathbf{ABV} \neq \mathbf{0} \text{ przy } r(\mathbf{A}) = s \text{ i } r(\mathbf{V}) = u, \quad (3.5)$$

3°. Wyznaczanie obszarów ufności dla krzywych wzrostu.

#### Metoda Potthoffa i Roya (PR)

Rozwiązanie powyższych problemów zaproponowali Potthoff i Roy (1964) podając następującą metodę.

Niech  $\mathbf{G}_{pp}$  będzie symetryczną, dodatnio określoną macierzą taką, aby  $\mathbf{TG}^{-1}\mathbf{T}'$  była nieosobliwa i aby  $\mathbf{C}_1 = \mathbf{G}^{-1}\mathbf{T}'(\mathbf{TG}^{-1}\mathbf{T}')^{-1}$  miała  $r(\mathbf{C}_1) = q$ . Za pomocą macierzy  $\mathbf{C}_1$  dokonuje się transformacji modelu (3.1) do zwykłego wielozmiennego modelu liniowego. Ponieważ  $\mathbf{TC}_1 = \mathbf{I}_q$ , więc  $\mathbf{YC}_1 = \mathbf{XBTC}_1 + \mathbf{EC}_1$  i oznaczając  $\mathbf{Y}_1 = \mathbf{YC}_1$  otrzymujemy model

$$\mathbf{Y}_1 = \mathbf{XB} + \mathbf{E}. \quad (3.6)$$

Zatem  $\mathbf{Y}_1$  o wartości oczekiwanej  $\mathbf{E}(\mathbf{Y}_1) = \mathbf{XB}$  ma niezależne wiersze o  $q$ -wymiarowych rozkładach normalnych z macierzami kowariancji  $\mathbf{\Sigma}_{qq} = \mathbf{C}_1'\mathbf{\Sigma}\mathbf{C}_1$ . Z ogólnej

teorii wielozmiennego modelu liniowego wynika, że estymator macierzy parametrów  $\mathbf{B}$  ma postać

$$\hat{\mathbf{B}} = (\mathbf{X}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{G}^{-1}\mathbf{T}'(\mathbf{T}\mathbf{G}^{-1}\mathbf{T}')^{-1}, \quad (3.7)$$

natomiast macierze sum kwadratów dla hipotezy i dla błędu są odpowiednio

$$\mathbf{S}_H = (\mathbf{A}\hat{\mathbf{B}}\mathbf{V})' [\mathbf{A}(\mathbf{X}\mathbf{X})^{-1}\mathbf{A}']^{-1}(\mathbf{A}\hat{\mathbf{B}}\mathbf{V}), \quad \mathbf{S}_E = \mathbf{V}'\mathbf{Y}'_1[\mathbf{I}_n - \mathbf{X}(\mathbf{X}\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}_1\mathbf{V}. \quad (3.8)$$

Z ogólnej teorii wynika, że przy prawdziwej hipotezie  $H_0$  określonej w (3.5)  $\mathbf{S}_H$  i  $\mathbf{S}_E$  mają niezależne rozkłady  $\mathbf{S}_H \sim W_p(s, \mathbf{V}'\Sigma_1\mathbf{V})$  i  $\mathbf{S}_E \sim W_u(n-m, \mathbf{V}'\Sigma_1\mathbf{V})$ . Problemem jest jednak wybór macierzy  $\mathbf{G}$ .

Potthoff i Roy pokazali, że optymalnym wyborem byłoby  $\mathbf{G} = \Sigma$ . Jednakże  $\Sigma$  jest nieznanne. Zastąpienie  $\Sigma$  przez  $\mathbf{S}$ , gdzie  $\mathbf{S} = \mathbf{Y}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}$  nie zachowuje własności związanych z rozkładami, ze względu na to, że  $\mathbf{S}$  zależy od  $\mathbf{Y}$ .

Inną możliwością jest wybór  $\mathbf{G} = \mathbf{I}_p$ , wtedy  $\mathbf{Y}_1 = \mathbf{Y}\mathbf{T}'(\mathbf{T}\mathbf{T}')^{-1}$ . Jeżeli macierz  $\mathbf{T}$  będzie zawierać współczynniki wielomianów ortogonalnych, to  $\mathbf{T}\mathbf{T}' = \mathbf{I}_q$  i  $\mathbf{Y}_1 = \mathbf{Y}\mathbf{T}'$ . Jest to wynik J. Wisharta (1938).

#### Metoda Rao-Khatri (R-K)

Alternatywny sposób rozwiązania problemów 1<sup>o</sup>–3<sup>o</sup> podali Rao (1965) i Khatri (1966). Proponują oni model warunkowy, aby uniknąć wyboru macierzy  $\mathbf{G}$ .

Dokonuje się transformacji modelu (3.1) wybierając macierz  $\mathbf{C}_{pp}$  taką, że

$\mathbf{C} = [\mathbf{C}_1 \mid \mathbf{C}_2]$  jest nieosobliwa oraz  $r(\mathbf{C}_1) = q$ ,  $\mathbf{T}\mathbf{C}_1 = \mathbf{I}_q$  oraz  $r(\mathbf{C}_2) = p-q$  i  $\mathbf{T}\mathbf{C}_2 = \mathbf{0}$ . Wtedy

$$\mathbf{Y}_c = \mathbf{Y}\mathbf{C} = \mathbf{Y}[\mathbf{C}_1 \mid \mathbf{C}_2] = [\mathbf{Y}\mathbf{C}_1 \mid \mathbf{Y}\mathbf{C}_2] = [\mathbf{Y}_1 \mid \mathbf{Y}_2]$$

oraz

$$\mathbf{E}(\mathbf{Y}_1) = \mathbf{X}\mathbf{B}\mathbf{T}\mathbf{C}_1 = \mathbf{X}\mathbf{B}, \quad \mathbf{E}(\mathbf{Y}_2) = \mathbf{X}\mathbf{B}\mathbf{T}\mathbf{C}_2 = \mathbf{0}$$

i warunkowy model może być zapisany jako model analizy kowariancji w postaci

$$\mathbf{E}(\mathbf{Y}_1 | \mathbf{Y}_2) = \mathbf{X}\mathbf{B} + \mathbf{Y}_2 \mathbf{\Gamma}_{p-q,q} = [\mathbf{X} : \mathbf{Y}_2] \begin{pmatrix} \mathbf{B} \\ \mathbf{\Gamma} \end{pmatrix}, \quad (3.9)$$

gdzie  $\mathbf{\Gamma}_{p-q,q}$  jest macierzą nieznanych współczynników regresyjnych przy zmiennych towarzyszących. Oznaczając  $\mathbf{Z} = \mathbf{Y}_2$  otrzymujemy znany wielozmienny model analizy kowariancji, gdzie  $\mathbf{Z}$  jest macierzą wartości zmiennych towarzyszących.  $r(\mathbf{X} : \mathbf{Z}) = m + p - q$ , gdyż  $\mathbf{X}$  ma  $m$  liniowo niezależnych kolumn i  $\mathbf{Z}$

jest macierzą z liniowo niezależnymi kolumnami. Kolumny  $\mathbf{Z}$  są też liniowo niezależne od kolumn  $\mathbf{X}$ . Ponadto wiersze  $\mathbf{Y}_1$  mają niezależne normalne wielozmienne warunkowe rozkłady z jednakową macierzą kowariancji.

Przy tych założeniach, stosując metodę analizy kowariancji, otrzymuje się estymatory najmniejszych kwadratów, które są również estymatorami największej wiarygodności (Khatri, 1966) postaci

$$\hat{\Gamma} = (\mathbf{C}'_2 \mathbf{S} \mathbf{C}_2)^{-1} \mathbf{C}'_2 \mathbf{S} \mathbf{C}_1$$

natomiast

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}\mathbf{S}^{-1} \mathbf{T}'(\mathbf{T}\mathbf{S}^{-1}\mathbf{T}')^{-1} \quad \text{przy} \quad \mathbf{S} = \mathbf{Y}'(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y}. \quad (3.10)$$

Jeżeli  $p \neq q$ , to należy użyć  $p-q$  zmiennych towarzyszących. Estymatory  $\hat{\mathbf{B}}$  uzyskane metodami PR i R-K są takie same, gdy  $\mathbf{G}=\mathbf{S}$ . Jeżeli  $\mathbf{G} = \mathbf{I}_p$ , metoda PR daje takie same rezultaty jak metoda analizy kowariancji (R-K) przy pominiętych zmiennych towarzyszących. Jest tak w przypadku, gdy kowariancje między kolumnami macierzy  $\mathbf{Y}_1$  i  $\mathbf{Y}_2$  są zerami.

Jeżeli  $p=q$ , problem wyboru macierzy  $\mathbf{C}$  nie istnieje.

Jak wybrać macierz  $\mathbf{C}$ ? Grizzle i Allen (1969) proponują jako jeden z możliwych wyborów przyjmując  $\mathbf{C}_1 = \mathbf{T}'$ , gdzie macierz  $\mathbf{T}$  zawiera współczynniki wielomianów ortogonalnych do stopnia  $q-1$  przy  $\mathbf{T}\mathbf{T}' = \mathbf{I}_q$ , natomiast macierz  $\mathbf{C}_2$  będzie zawierać pozostałe współczynniki wielomianów ortogonalnych uzupełniających do stopnia  $p-1$ .

Równoważność ocen uzyskanych obydwoma metodami, gdy  $\mathbf{G}=\mathbf{S}$  i przy użyciu wszystkich zmiennych towarzyszących wykazali Grizzle i Allen (1969) oraz Lee Y.H.K. (1974), a Baksalary, Corsten, Kala (1978) rozszerzyli to na przypadek, gdy użyte są wszystkie zmienne towarzyszące.

Macierze uogólnionych sum kwadratów do testowania hipotezy  $H_0 : \mathbf{ABV} = \mathbf{0}$  uzyskane metodą analizy kowariancji są postaci

$$\mathbf{S}_h = \mathbf{V}'\mathbf{Y}'_1 \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{A}'(\mathbf{A}\mathbf{R}\mathbf{A}')^{-1} \mathbf{A}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}_1 \mathbf{V}, \quad (3.11)$$

gdzie

$$\mathbf{R} = (\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}[\mathbf{S}^{-1} - \mathbf{S}^{-1} \mathbf{T}'(\mathbf{T}\mathbf{S}^{-1}\mathbf{T}')^{-1}] \mathbf{Y}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

i

$$\mathbf{Y}_1 = \mathbf{Y}\mathbf{S}^{-1} \mathbf{T}'(\mathbf{T}\mathbf{S}^{-1}\mathbf{T}')^{-1}$$

oraz

$$\mathbf{S}_e = \mathbf{V}'(\mathbf{T}\mathbf{S}^{-1}\mathbf{T}')^{-1} \mathbf{V} \quad (3.12)$$

przy  $v_h = r(\mathbf{A}) = q$ ,  $v_e = n - r(\mathbf{X}) - (p - q) = n - m - (p - q)$ .

Macierze  $\mathbf{S}_H$  i  $\mathbf{S}_h$  oraz  $\mathbf{S}_E$  i  $\mathbf{S}_e$  określone w (3.8), (3.11) i (3.12) nie są identyczne. Natomiast  $\mathbf{S}_h$  i  $\mathbf{S}_e$  przy prawdziwej  $H_0$  mają niezależne rozkłady Wisharta z  $v_h$  i  $v_e$  stopniami swobody oraz wspólną macierzą kowariancji i nie zależą od zmiennych towarzyszących  $\mathbf{Y}_2$ . Można więc stosować znane testy wielozmienne.

### Metoda Tubbsa, Levisa i Durana

Timm (1980) cytuje Tubbsa, Levisa i Durana (1975), którzy otrzymali estymatory parametrów w modelu krzywych wzrostu metodą największej wiarygodności. Są one równoważne estymatorom PR, gdy  $\mathbf{G} = \mathbf{S}$  i mają postać

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{S}^{-1}\mathbf{T}'(\mathbf{T}\mathbf{S}^{-1}\mathbf{T}')^{-1}.$$

Podane przez wymienionych autorów uogólnione sumy kwadratów są:

$$\mathbf{S}_h = \mathbf{V}'\mathbf{Y}'_1\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}']^{-1}\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_1\mathbf{V}$$

i

$$\mathbf{S}_e = \mathbf{V}'(\mathbf{T}\mathbf{S}^{-1}\mathbf{T}')\mathbf{V}$$

przy  $v_h = r(\mathbf{A})$ ,  $v_e = n - r(\mathbf{X})$  gdzie  $\mathbf{S} = \mathbf{Y}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}$  i  $\mathbf{Y}_1 = \mathbf{Y}\mathbf{S}^{-1}\mathbf{T}'(\mathbf{T}\mathbf{S}^{-1}\mathbf{T}')^{-1}$ .

Procedura R-K daje wyniki asymptotycznie równoważne wynikom Tubbsa i innych.

Praktycznie najczęściej stosowana jest metoda Rao-Khatri.

Testowanie hipotezy  $H_0: \mathbf{ABV} = \mathbf{0}$  można dokonać za pomocą jednego z testów:

$$1) \Lambda - \text{Wilksa: } \Lambda = \frac{|\mathbf{S}_e|}{|\mathbf{S}_e + \mathbf{S}_h|},$$

$$2) T^2 - \text{Hotellinga: } T_o^2 = v_e \text{tr}(\mathbf{S}_h \cdot \mathbf{S}_e^{-1}),$$

$$3) \theta - \text{Roy'a, największego pierwiastka charakterystycznego: } \theta = \frac{\lambda_1}{1 + \lambda_1}.$$

Testowanie adekwatności modelu można również traktować jako szczególny przypadek ogólnej hipotezy  $H_0: \mathbf{ABV} = \mathbf{0}$ .

Oznaczając  $\mathbf{B}_1 = \mathbf{BT}$  mamy  $E(\mathbf{Y}) = \mathbf{XB}_1$ , a przy  $\mathbf{TC}_2 = \mathbf{0}$

$$E(\mathbf{Y}_2) = \mathbf{XB}_1\mathbf{C}_1 = \mathbf{XBTC}_2 = \mathbf{0}.$$

Jeżeli  $r(\mathbf{X}) = m$ , czyli  $\mathbf{X}$  jest pełnego rzędu, to  $\mathbf{XB}_1\mathbf{C}_2 = \mathbf{0}$  wtedy i tylko wtedy, gdy  $\mathbf{B}_1\mathbf{C}_2 = \mathbf{0}$ . Model (3.1) spełnia to założenie, a więc jego adekwatność może być testowana przez weryfikowanie hipotezy, że  $E(\mathbf{Y}_2) = \mathbf{0}$ , co można zapisać

$H_0 : \mathbf{I}_p \mathbf{B}_1 \mathbf{C}_2 = \mathbf{0}$  dla modelu, gdy  $E(\mathbf{Y}) = \mathbf{X}\mathbf{B}_1$ . Jest to testowanie ogólnej hipotezy przy  $\mathbf{A} = \mathbf{I}_p$ .

Obszary ufności najczęściej rozważa się dla estymowalnych funkcji parametrycznych postaci

$$\varphi = \mathbf{a}'\mathbf{B}\mathbf{d} , \quad (3.13)$$

gdzie  $\mathbf{a}$  jest wektorem  $(m \times 1)$ , a  $\mathbf{d}$  wektorem  $(q \times 1)$ .

W zależności od postaci wektorów  $\mathbf{a}$  i  $\mathbf{d}$  można wyznaczać przedziały ufności dla poszczególnych parametrów czy ich różnicy albo obszary ufności dla krzywej lub różnicy krzywych. W przypadku wyznaczania obszarów ufności przyjmuje się, że wektor  $\mathbf{d}$  powinien być  $\mathbf{d}_t = [1, t, t^2, \dots, t^{q-1}]'$ . Należy pamiętać, że dowolna funkcja parametryczna będzie estymowalna jeżeli macierze  $\mathbf{X}$  i  $\mathbf{T}$  są pełnego rzędu.

W najbardziej ogólnej formie, wykorzystując test maksymalnego pierwiastka Roy'a dla testowania hipotezy  $H_0 : \mathbf{A}\mathbf{B}\mathbf{V} = \mathbf{0}$  przy  $r(\mathbf{X}) = m$ , jednocześnie  $(1-\alpha)100\%$  przedziały ufności dla kombinacji liniowej  $\mathbf{a}'\mathbf{A}\mathbf{B}\mathbf{V}\mathbf{d}$  są postaci (Seber, 1984)

$$\mathbf{a}'\mathbf{A}\mathbf{B}\mathbf{V}\mathbf{d} \in \left\{ \mathbf{a}'\hat{\mathbf{A}}\mathbf{B}\mathbf{V}\mathbf{d} \pm \sqrt{\frac{\theta_\alpha}{1-\theta_\alpha} \mathbf{a}'\mathbf{F}\mathbf{a} \cdot \mathbf{d}'\mathbf{S}_e\mathbf{d}} \right\} \quad (3.14)$$

dla wszystkich  $\mathbf{a}$  i  $\mathbf{d}$ , gdzie

$$\mathbf{F} = \mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}' + \mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}[\mathbf{S}^{-1} - \mathbf{S}^{-1}\mathbf{T}'(\mathbf{T}\mathbf{S}^{-1}\mathbf{T}')\mathbf{T}\mathbf{S}^{-1}]\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}' ,$$

zaś  $\theta_\alpha$  odczytana jest z tablic statystyki maksymalnego pierwiastka Roy'a, które można znaleźć na przykład w książce Sebera (1984) na str. 593-598, przy  $s_o = \min(u, s)$ ,  $v_1 = \frac{1}{2}(|u-s|-1)$ ,  $v_2 = \frac{1}{2}[n-m-(p-q)-u-1]$ .

W szczególnym przypadku, gdy dane podlegają pojedynczej klasyfikacji (Balsalary, Caliński, Kala, 1980), przedział ufności dla estymowalnej funkcji parametrycznej (3.13) jest postaci

$$\varphi \in \left\{ \hat{\varphi} \pm \sqrt{\frac{1}{n-m-p+q} F_{\alpha/2; 1, n-m-p+q} \cdot \mathbf{a}'\mathbf{R}\mathbf{a} \cdot \mathbf{d}'(\mathbf{T}\mathbf{S}^{-1}\mathbf{T}')^{-1}\mathbf{d}} \right\} ,$$

gdzie

$$\mathbf{R} = (\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\{\mathbf{S}^{-1} - \mathbf{S}^{-1}\mathbf{T}'(\mathbf{T}\mathbf{S}^{-1}\mathbf{T}')^{-1}\mathbf{T}\mathbf{S}^{-1}\}\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} ,$$

natomiast  $100(1-\alpha)\%$  obszar ufności dla funkcji  $\varphi(t) = \mathbf{a}'\mathbf{B}\mathbf{d}_t$  dla wszystkich wartości  $t$  wyraża się wzorem

$$\varphi(t) \in \left\{ \hat{\varphi}(t) \pm \sqrt{\frac{q}{n-m-p+1} F_{\alpha/2; q, n-m-p+1} \cdot \mathbf{a}'\mathbf{R}\mathbf{a} \cdot \mathbf{d}_t'(\mathbf{T}\mathbf{S}^{-1}\mathbf{T}')^{-1}\mathbf{d}_t} \right\} .$$

Korzystając z tego ostatniego wzoru można wyznaczyć również obszar ufności dla różnicy  $\varphi_1(t) - \varphi_2(t)$ .

#### 4. Krzywe wzrostu z losowymi współczynnikami

Jeżeli zmienność między jednostkami eksperymentalnymi jest duża i liczba pomiarów dla każdej z nich nie jest taka sama, poprzednio omówiona metoda nie da się zastosować. Wynika to stąd, że założenia przyjęte dla błędów losowych w modelu (3.1) nie są spełnione.

Zagadnienie niezrównoważonych krzywych wzrostu pierwszy rozważał Rao (1965). Od tego czasu szereg autorów podejmuje badania nad modelem krzywych wzrostu przy innych niż dotąd założeniach. Uzyskane zostają pewne wyniki metodami iteracyjnymi, Bayesowskimi i w oparciu o predykcję brakujących obserwacji.

Metodę nieiteracyjną podali pierwsi Carter i Yang (1986). Mogła być ona jednakże stosowana tylko do danych niesklasyfikowanych.

Uogólnienie tej metody dla dowolnych danych przedstawili Vonesh i Carter (1987). Jest to metoda dwustopniowa, zwana krzywymi wzrostu z losowymi współczynnikami.

W poprzednio omówionej metodzie rozważane były modele (3.1a) i (3.1b). Były to modele stałe. Przedstawiona tu metoda obejmuje przypadek, gdy tylko model (3.1a) jest stały, natomiast współczynniki regresji w modelu (3.1b) są losowe (stąd nazwa metody). Uzyskanie estymatorów współczynników w równaniach krzywych wzrostu wymaga dwóch etapów.

Pierwszy to wyznaczenie estymatorów współczynników regresji w (3.1b), traktowanych chwilowo jako ustalone, dla kolejnych wektorów obserwacji w czasie odpowiadających poszczególnym jednostkom eksperymentalnym. Wektory te nie muszą mieć jednakowej liczby elementów.

W drugim natomiast, ze względu na to, że wyznaczone na podstawie losowych obserwacji współczynniki regresji są losowe, wykorzystując je można utworzyć wielozmienny model liniowy. W tym stałym modelu parametrami są właśnie nieznanne współczynniki w równaniach krzywych wzrostu, a macierz układu uwzględnia przynależność obserwacji do poszczególnych grup.

Dokładny opis kolejnych etapów tej metody jest następujący.

1°. "Wewnątrz jednostek eksperymentalnych"

U każdej z  $n$  jednostek eksperymentalnych dokonuje się pomiarów cechy w  $r_i$  punktach czasowych, uzyskując wektory obserwacji  $\mathbf{y}_i$ , dla których model ma postać

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{e}_i, \quad i = 1, 2, \dots, n, \quad (4.1)$$

gdzie  $\mathbf{y}'_i = (y_{1i}, y_{2i}, \dots, y_{r_i, i})$  jest wektorem pomiarów na  $i$ -tej jednostce,  $\mathbf{X}_i$  - jest macierzą  $(r_i \times p)$  znanych zmiennych objaśniających taką, że  $r(\mathbf{X}_i) = p < r_i$ , a  $\beta_i$  jest  $(p \times 1)$ -wektorem nieobserwowalnych losowych współczynników regresji. Wektory  $\mathbf{e}_i$  o wymiarach  $(r_i \times 1)$  mają niezależne rozkłady  $N(\mathbf{0}, \sigma^2 \mathbf{I})$ .

2°. "Pomiędzy jednostkami eksperymentalnymi"

Model ma teraz postać

$$\beta = \mathbf{Z}\mathbf{A} + \xi, \quad (4.2)$$

gdzie  $\beta = (\beta_1, \beta_2, \dots, \beta_n)'$  jest  $(n \times p)$  macierzą losowych współczynników regresji,  $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n)'$  jest  $(n \times q)$  macierzą układu, to znaczy  $i$ -ty wiersz macierzy  $\mathbf{Z}$  oznaczony przez  $\mathbf{Z}'_i$  ( $i = 1, \dots, n$ ) wskazuje przynależność badanej jednostki do jednej z  $q$  grup ( $q < n$ ).  $\mathbf{A}$  jest  $(q \times p)$  macierzą stałych parametrów zwanych przez autorów lokacyjnymi.  $\xi = (\xi_1, \xi_2, \dots, \xi_n)'$  jest  $(n \times p)$  macierzą losowych błędów. O wektorach  $\xi_i$  zakłada się, że mają niezależne rozkłady normalne  $N_p(\mathbf{0}, \Psi)$  i są niezależne od  $\mathbf{e}_i$ .

Model (4.2) można wyrazić w postaci wektorowej

$$\beta_i = (\mathbf{I}_p \otimes \mathbf{Z}'_i) \lambda + \xi_i \quad \text{dla } i = 1, \dots, n, \quad (4.3)$$

gdzie  $\lambda$  jest  $(pq \times 1)$  - wektorem ułożonym z kolumn macierzy  $\mathbf{A}$  jedna pod drugą.

Elementy macierzy  $\mathbf{A}$  (lub wektora  $\lambda$ ) są współczynnikami poszukiwanych krzywych wzrostu. Należy zatem wyznaczyć ich estymatory.

Sprowadza się to do wyznaczenia metodą najmniejszych kwadratów zwykłych estymatorów poszczególnych  $\beta_i$ , oznaczonych dalej przez  $\mathbf{b}_i$ , a następnie podstawienia ich do (4.3) i oszacowania stąd nieznanych elementów macierzy  $\mathbf{A}$ .

Mamy więc:

$$\mathbf{b}_i = (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{y}_i, \quad i = 1, 2, \dots, n, \quad (4.4)$$

z macierzą kowariancji  $\mathbf{V}_i = \Psi + \sigma^2 (\mathbf{X}'_i \mathbf{X}_i)^{-1}$ . Model krzywych wzrostu z losowymi współczynnikami ma więc postać

$$\mathbf{b}_i = (\mathbf{I}_p \otimes \mathbf{Z}'_i) \lambda + \mathbf{v}_i, \quad i = 1, 2, \dots, n, \quad (4.5)$$

gdzie  $\mathbf{v}_i$  mają niezależne rozkłady  $N_p(\mathbf{0}, \mathbf{V}_i)$  i estymator  $\lambda$  będzie wyrażony przez  $\mathbf{b}_i$ .

Oznaczamy przez  $\mathbf{0}$  wektor złożony z  $1 + p(p+1)/2$  parametrów  $\sigma^2$  i różnych elementów macierzy  $\Psi$ .

W dalszej analizie można rozważać dwa przypadki.

### 1. Znana wariancja (znane elementy wektora $\mathbf{0}$ ).

Uogólnioną metodą najmniejszych kwadratów uzyskuje się najlepszy liniowy nieobciążony estymator  $\lambda$  w postaci

$$\hat{\lambda}(\mathbf{0}) = \left[ \sum_{i=1}^n \{ \mathbf{W}_i \otimes \mathbf{Z}_i \mathbf{Z}_i' \} \right]^{-1} \sum_{i=1}^n \{ \mathbf{W}_i \otimes \mathbf{Z}_i \} \mathbf{b}_i, \quad (4.6)$$

gdzie  $\mathbf{W}_i = \mathbf{V}_i^{-1}$  z macierzą kowariancji

$$\Omega(\mathbf{0}) = \left[ \sum_{i=1}^n \{ \mathbf{W}_i \otimes \mathbf{Z}_i \mathbf{Z}_i' \} \right]^{-1}. \quad (4.7)$$

### 2. Nieznana wariancja.

Należy najpierw wyznaczyć nieobciążoną ocenę elementów  $\mathbf{0}$ , a następnie podstawić ją do (4.6).

Metodą momentów wyznacza się nieobciążony estymator  $\psi$  postaci

$$\tilde{\psi} = \mathbf{S}_{bb} - \hat{\sigma}^2(n-q)^{-1} \sum_{i=1}^n [1 - \mathbf{Z}_i'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}_i] (\mathbf{X}_i' \mathbf{X}_i)^{-1},$$

gdzie

$$\mathbf{S}_{bb} = \mathbf{B}'[\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{B} / (n-q) \quad \text{i} \quad \mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)'. \quad (4.8)$$

Ponieważ  $\tilde{\psi}$  może nie być dodatnio określone przyjmuje się

$$\hat{\psi} = \begin{cases} \tilde{\psi} & \text{gdy } \hat{\zeta} > \hat{\sigma}_2^2, \\ \mathbf{S}_{bb} - \hat{\zeta}(n-q)^{-1} \sum_{i=1}^n [1 - \mathbf{Z}_i'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}_i] (\mathbf{X}_i' \mathbf{X}_i)^{-1} & \text{gdy } \hat{\zeta} \leq \hat{\sigma}_2^2, \end{cases} \quad (4.9)$$

gdzie  $\hat{\zeta}$  jest najmniejszym pierwiastkiem równania

$$|\mathbf{S}_{bb} - \zeta(n-q)^{-1} \sum_{i=1}^n [1 - \mathbf{Z}_i'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}_i] (\mathbf{X}_i' \mathbf{X}_i)^{-1}| = 0,$$

natomiast

$$\hat{\sigma}^2 = \left[ \sum_{i=1}^n (r_i - p) \right]^{-1} \sum_{i=1}^n (r_i - p) S_i^2, \quad (4.10)$$

przy



$$S_i^2 = \mathbf{y}'_i [\mathbf{I} - \mathbf{X}_i (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i] \mathbf{y}_i / (r_i - p) ,$$

jest nieobciążonym estymatorem  $\sigma^2$ .

Estymator  $\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\theta}})$  wyznacza się metodą uogólnionych najmniejszych kwadratów podstawiając za  $\boldsymbol{\theta}$  estymator  $\hat{\boldsymbol{\theta}}$  uzyskany z  $\hat{\sigma}^2$  i  $\hat{\boldsymbol{\psi}}$ . Ma on postać

$$\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\theta}}) = \left[ \sum_{i=1}^n (\hat{\mathbf{W}}_i \otimes \mathbf{Z}_i \mathbf{Z}'_i) \right]^{-1} \sum_{i=1}^n (\hat{\mathbf{W}}_i \otimes \mathbf{Z}_i) \mathbf{b}_i , \quad (4.11)$$

gdzie  $\hat{\mathbf{W}}_i = [\hat{\boldsymbol{\psi}} + \hat{\sigma}^2 (\mathbf{X}'_i \mathbf{X}_i)^{-1}]^{-1}$ , a estymatorem jego macierzy kowariancji jest

$$\hat{\boldsymbol{\Omega}}(\hat{\boldsymbol{\theta}}) = \left[ \sum_{i=1}^n (\hat{\mathbf{W}}_i \otimes \mathbf{Z}_i \mathbf{Z}'_i) \right]^{-1} .$$

Estymator  $\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\theta}})$  jest zgodnym i asymptotycznie efektywnym estymatorem  $\boldsymbol{\lambda}$  przy  $n \rightarrow \infty$ .

Testowanie hipotez zależy w dużym stopniu od wielkości  $n$ , czyli ilości jednostek wybranych do badania. Przy modelu (4.5) ogólna wielozmienna hipoteza liniowa  $H_0 : \mathbf{C} \boldsymbol{\lambda} \mathbf{A} = \mathbf{0}$  może być zapisana w postaci  $H'_0 : (\mathbf{A}' \otimes \mathbf{C}) \boldsymbol{\lambda} = \mathbf{0}$ , gdzie macierz  $\mathbf{C}$  o wymiarach  $(v_1 \times q)$  ma  $r(\mathbf{C}) = v_1 \leq q$ , zaś macierz  $\mathbf{A}$  o wymiarach  $(p \times v_2)$  ma  $r(\mathbf{A}) = v_2 \leq p$ . Do testowania tej hipotezy zostanie użyta statystyka

$$T^2(\hat{\boldsymbol{\theta}}) = [\mathbf{A}' \otimes \mathbf{C}] \hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\theta}}) [(\mathbf{A}' \otimes \mathbf{C}) \hat{\boldsymbol{\Omega}}(\hat{\boldsymbol{\theta}}) (\mathbf{A}' \otimes \mathbf{C})]^{-1} [(\mathbf{A}' \otimes \mathbf{C}) \hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\theta}})] . \quad (4.12)$$

Ma ona przy prawdziwej hipotezie  $H'_0$  i przy dużej liczebności  $n$  przybliżony rozkład  $\chi^2$  z  $v = v_1 v_2$  stopniami swobody.

Jeżeli  $\min(r_i)$  będzie duże, to przy prawdziwej  $H_0$  statystyka (4.12) ma rozkład  $T^2$ -Hotellinga, a więc  $T^2(\hat{\boldsymbol{\theta}}) / (n - q)$  należy porównać z odpowiednią wartością  $T^2_{\alpha}$  odczytaną z tablic (z parametrami  $s_o = \min(v_1, v_2)$ ,  $m_o = (|v_1 - v_2| - 1) / 2$  i  $n_o = [(n - q) - v_2] / 2$ ). Alternatywnie, korzystając z aproksymacji funkcją  $F$ ,  $H_0$  można testować używając statystyki

$$F(\hat{\boldsymbol{\theta}}) = \frac{2s_o n_o - s_o + 2}{s_o v_1 v_2} \cdot \frac{1}{n - q} \cdot T^2(\hat{\boldsymbol{\theta}}) ,$$

która dla dużych  $\min(r_i)$  ma przybliżony centralny rozkład  $F$  z  $v_1 = v_1 v_2$  i  $v_2 = 2s_o n_o - s_o + 2$  stopniami swobody.

Ilustracją tej metody niech będzie następujący przykład.

*Przykład:* Dane uzyskane z Instytutu Szczegółowej Uprawy Roli i Roślin AR w Lublinie dotyczyły badania u rośliny *Sida Hermaphrodita Rusby* przebiegu zmian

wysokości roślin (w cm) w pierwszym pokosie w pięciu następujących terminach: 1–13.05, 2–23.05, 3–2.06, 4–12.06 i 5–22.06.1987 roku. W badaniach zostały uwzględnione cztery grupy roślin wyhodowanych z: odcinków korzeniowych sadzonych wiosną w odległości 40 cm (KW), wysiewu nasion wiosną w ilości 4 kg/ha (NW), odcinków korzeniowych sadzonych jesienią w odległości 40 cm (KJ) oraz wysiewu nasion jesienią w ilości 4 kg/ha (NJ). Rozstawa rzędów wynosiła 70 cm. W każdej grupie poddano badaniu 40 wybranych losowo roślin. Uzyskane wyniki zestawiono w macierz  $\mathbf{Y}$  o wymiarach  $(160 \times 5)$ .

Celem badania było ustalenie kształtu funkcji opisującej przebieg zmian wysokości roślin *Sida* w pierwszym pokosie dla każdej z grup oraz porównanie ich między sobą.

Stosując opisaną w tym rozdziale metodę dla wielomianu stopnia drugiego ( $p=3$ ) mamy:  $n=160$ , dla każdego  $i=1, \dots, 160$  liczba pomiarów w czasie jest stała, a więc  $r_i=5$  oraz  $\mathbf{X}_i = \mathbf{X}$  przy czym

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ t_1 & t_2 & t_3 & t_4 & t_5 \\ t_1^2 & t_2^2 & t_3^2 & t_4^2 & t_5^2 \end{bmatrix}$$

i ze względu na jednakowe długości odcinków czsowych przyjmuje się  $t_1 = 1, \dots, t_5 = 5$ . Ponadto liczba grup obserwacji  $q=4$ .

Wykorzystując do obliczeń program komputerowy uzyskano następujące wyniki:

$$\hat{\sigma}^2 = 57.588 ,$$

macierz  $\hat{\Psi}$

$$\hat{\Psi} = \begin{bmatrix} 95.285 & -104.585 & 18.464 \\ -104.587 & 114.812 & -20.196 \\ 18.464 & -20.196 & 3.845 \end{bmatrix}$$

oraz wektor  $\hat{\lambda}$

$$\hat{\lambda}' =$$

$$[5.742 \ 9.019 \ 6.327 \ 12.849 \ 3.327 \ 1.574 \ 2.748 \ -2.149 \ 3.040 \ 3.678 \ 3.284 \ 4.060]$$

Daje to następujące równania dla poszczególnych sposobów hodowli:

$$\text{KW: } y = 5.742 + 3.327x + 3.040x^2 ,$$

$$\text{NW: } y = 9.019 + 1.574x + 3.678x^2 ,$$

$$\text{KJ: } y = 6.327 + 2.748x + 3.284x^2 ,$$

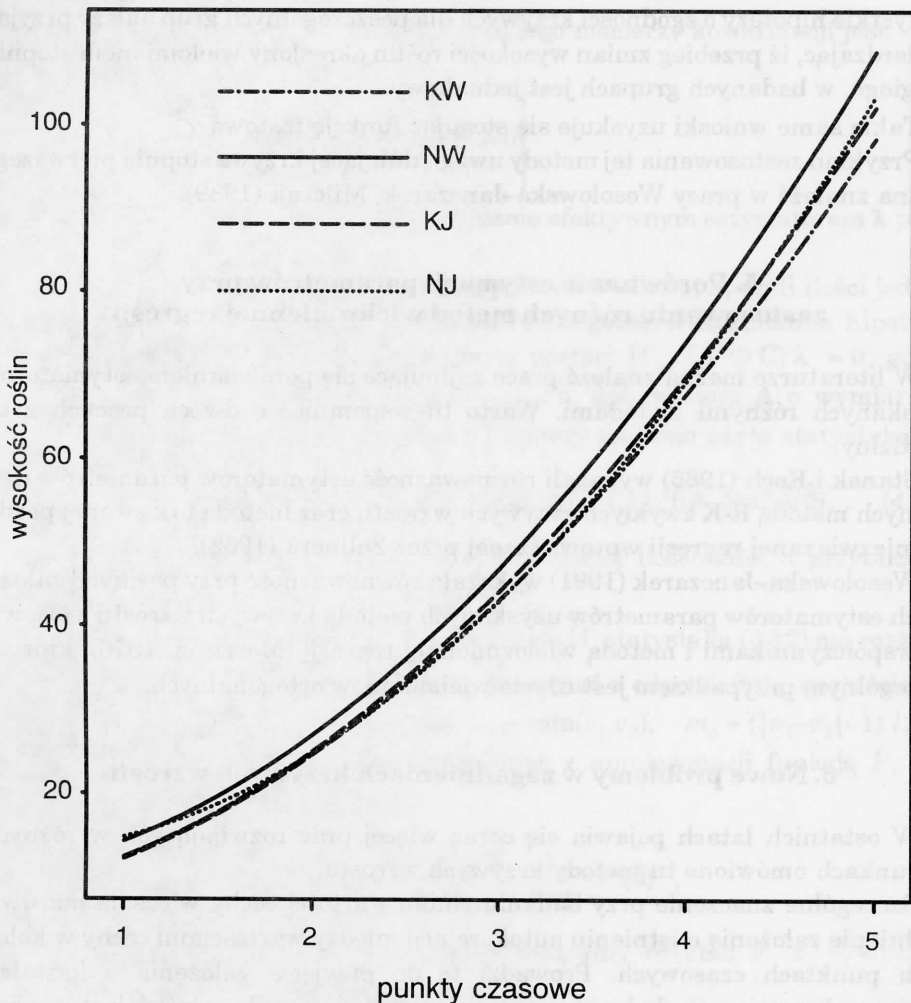
$$\text{NJ: } y = 12.849 - 2.149x + 4.060x^2 .$$

Wykresy uzyskanych krzywych znajdują się na Rys. 4.

Sprawdzenie zgodności przebiegu zmian w czasie w poszczególnych grupach wykonano weryfikując hipotezy  $H_0: C\Lambda A = 0$  przyjmując  $A = I_3$  i kolejno

$$1^\circ. \quad C = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

dla porównania krzywych KW i NW, NW i KJ oraz KJ i NJ,



**Rys. 4.** Zmiany wysokości roślin *Sida* w pierwszym pokosie przy wybranych sposobach ich hodowli: KW, NW, KJ i NJ.

$$2^{\circ}. \quad C = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

dla porównania krzywych KW i KJ oraz

$$3^{\circ}. \quad C = [1 \ 0 \ 0 \ -1]$$

dla porównania KW i NJ.

Odpowiednie obliczone wartości funkcji testowych  $F$  są następujące

$$F_{9;455}^{\circ} = 1.559; \quad F_{6;304}^{\circ} = 0.947 \quad \text{oraz} \quad F_{3;153}^{\circ} = 1.428 .$$

Wszystkie hipotezy o zgodności krzywych dla poszczególnych grup należy przyjąć stwierdzając, iż przebieg zmian wysokości roślin określony wielomianem stopnia drugiego, w badanych grupach jest jednakowy.

Takie same wnioski uzyskuje się stosując funkcję testową  $\chi^2$ .

Przykład zastosowania tej metody uwzględniającej krzywe stopnia pierwszego można znaleźć w pracy Wesołowska-Janczarek, Milczak (1989).

## **5. Porównanie estymacji parametrów przy zastosowaniu różnych metod wielozmiennej regresji**

W literaturze można znaleźć prace zajmujące się porównaniem estymatorów uzyskanych różnymi metodami. Warto tu wspomnieć o dwóch pracach z tej dziedziny.

Stanek i Koch (1985) wykazali równoważność estymatorów parametrów uzyskanych metodą R-K zwykłych krzywych wzrostu oraz metodą tak zwanej pozornie niezwiązanej regresji wprowadzonej przez Zellnera (1962).

Wesołowska-Janczarek (1991) wykazała równoważność przy pewnych założeniach estymatorów parametrów uzyskanych metodą krzywych wzrostu z losowymi współczynnikami i metodą wielozmiennej regresji (Morrison, 1976), których szczególnym przypadkiem jest użycie wielomianów ortogonalnych.

## **6. Nowe problemy w zagadnieniach krzywych wzrostu**

W ostatnich latach pojawia się coraz więcej prac rozwijających w różnych kierunkach omówione tu metody krzywych wzrostu.

Szczególne znaczenie przy badaniu zmian wartości cechy w czasie ma uwzględnienie założenia o istnieniu autokorelacji między wartościami cechy w kolejnych punktach czasowych. Prowadzi to do przyjęcia założenia o kształcie macierzy kowariancji. Jednym z częściej uwzględnianych założeń jest seryjna struktura kowariancyjna. Jak stwierdza Lee (1988) wydaje się, że jest ona jedną z najbardziej naturalnych postaci macierzy kowariancji przy rozważaniu danych

dotyczących krzywych wzrostu. Dane te uzyskuje się bowiem przez powtarzanie pomiarów tak jak w bardzo krótkich szeregach czasowych.

Dodatkowym argumentem przemawiającym za przyjęciem takiej struktury macierzy kowariancji jest to, że ilość parametrów, które należy w niej estymować jest mniejsza niż w macierzy bez określonej seryjnej struktury i nie zależy od ilości uwzględnionych punktów czasowych.

Przyjmując założenie w modelu (3.1) jak dla stacjonarnego szeregu czasowego z autoregresją rzędu  $p-1$ , Lee (1988) podaje estymatory współczynników w równaniach krzywych wzrostu dla modelu stałego z nieznaną macierzą kowariancji w następującej postaci.

W modelu (3.1) z macierzą kowariancji

$$\Sigma = \sigma^2 C \quad \text{gdzie } C = (\rho^{|i-j|}) \quad \text{przy } i, j = 1, 2, \dots, p \quad (6.1)$$

estymator największej wiarygodności  $\hat{B}$  jest postaci:

$$\hat{B} = (X'X)^{-1} X'Y\hat{C}^{-1} T'(T\hat{C}^{-1}T')^{-1}, \quad (6.2)$$

gdzie

$$\hat{C} = (\hat{\rho}^{|i-j|})$$

oraz

$$\hat{\sigma}^2 = \frac{1}{pn} \{ \text{tr}[T\hat{C}^{-1}S\hat{C}^{-1}T'(T\hat{C}^{-1}T')^{-1}] + \text{tr}[Z'YZ(Z'\hat{C}Z)^{-1}] \}, \quad (6.3)$$

gdzie  $Z$  jest znaną macierzą o wymiarach  $(p, p-q)$ ,  $r(Z) = p-q$  i taką, że  $TZ=0$ , zaś

$$S = Y'[I - X(X'X)^{-1}X']Y.$$

Estymator największej wiarygodności współczynnika autokorelacji  $\rho$  otrzymuje się przez maksymalizację funkcji

$$L_{\max}(\rho) = (\hat{\sigma}^2(\rho))^{-pn/2} (1 - \rho^2)^{-n(p-1)/2}, \quad (6.4)$$

gdzie  $\hat{\sigma}^2(\rho)$  jest wyrażone wzorem (6.3), w którym  $\hat{\rho}$  zostało zastąpione przez  $\rho$ .

Innego typu szczególną postać macierzy kowariancji dla błędu zakłada Azza-lini (1987). Jest ona zgodna z macierzą kowariancji dla stacjonarnego procesu AR( $k$ ) oraz ARMA(1.1). Przy tych założeniach rozważana jest estymacja parametrów w modelach stałym i z losowymi współczynnikami. Traktując model stały jako szczególny przypadek modelu z losowymi współczynnikami, został podany test decydujący o wyborze jednego z tych modeli. Także Jones i Boadi-Boateng (1991) przedstawili analizę krzywych wzrostu przy założeniu autoregresji pierwszego rzędu ale niejednakowych odstępach czasu.

Badaniem założenia dodatniej określoności macierzy kowariancji w ogólnym stałym modelu krzywych wzrostu, kryteriami wyboru modelu przy różnej postaci

macierzy kowariancji uwzględniającej seryjną strukturę oraz testem równości macierzy kowariancji dla  $k$  populacji zajmuje się Lee (1991). Zerbe i Walker (1977) podają test porównujący krzywe wzrostu dla poszczególnych grup, gdy dopasowane krzywe są różnego stopnia. Analizę krzywych wzrostu, gdy obserwacje u poszczególnych osobników dokonywane były w niejednakowych punktach czasowych przedstawił Hui (1984).

Estymację krzywych wzrostu przy brakujących obserwacjach ocenianych metodą analizy kowariancji oraz przy zastosowaniu algorytmu EM zaprezentował Liski (1985). Wyborowi zmiennych towarzyszących oraz testowaniu hipotez o zbędności niektórych zmiennych w metodzie R-K poświęcili swoją pracę Fujikoshi i Rao (1991).

Kilka prac poświęconych jest wieloreakcyjnym krzywym wzrostu. Sprawdzeniem poprawności założenia o strukturze macierzy kowariancji takiej jak dla krzywych wzrostu z losowymi współczynnikami w przypadku badania przebiegu zmian kilku cech równocześnie zajmują się Chinchilli i Carter (1984). Natomiast Lundbye-Christensen (1991) przedstawia analizę wielocechowych krzywych wzrostu przy różnych strukturach macierzy kowariancji i różnych wartościach oczekiwanych w zastosowaniu do badań medycznych.

Osobną grupę stanowią prace poświęcone predykcji w modelach krzywych wzrostu. Można wymienić tu niektóre z nich: Lee i Geisser (1972), Rao (1977), Geisser (1980), Rao i Baudream (1985), Liski i Nummi (1988), Lee (1988).

Przedstawiony tu przegląd tematyki nie wyczerpuje pełnej bibliografii prac dotyczących krzywych wzrostu, nie było to bowiem głównym celem tej pracy.

## Podziękowanie

Autorka chciałaby tą drogą podziękować profesorowi Wiktorowi Oktabie za zaproszenie jej do wygłoszenia referatu na XXII Colloquium Metodologicznym z Agrobiometrii w Olsztynie, które przyczyniło się do powstania tej pracy.

## LITERATURA

- Azzalini A. (1987). Growth curves analysis for patterned covariance matrices. In "New Perspectives in Theoretical and Applied Statistics". Puri M.L., Vilaplana J. P and Wertz W. (ed.), John Wiley, New York, 63-73.
- Baksalary J., Caliński T., Kala R. (1980). Estymacja krzywych wzrostu i jej zastosowanie do oceny odmian gatunków o plonowaniu wielokrotnym, *Biuletyn Oceny Odmian*. Tom VIII, 1(12), 167-182.
- Baksalary J., Caliński T., Kala R. (1983). Estymacja krzywych wzrostu w układzie bloków kompletnych. *Biuletyn Oceny Odmian*. Tom X, 1(15), 105-118.
- Baksalary J., Corsten L.C.A. Kala, R. (1978). Reconciliation of two different views on estimation of growth curve parameters. *Biometrika* **65**, 662-665.
- Bard Y. (1974). *Nonlinear Parameter Estimation*. Academic Press, New York.

- Carter R.L., Yang M.C.K. (1986). Large sample inference in random coefficient regression models. *Communications in Statistics - Theory and Methods* **8**, 2507-2526.
- Causton D.R., Venus J.C. (1981). *The Biometry of Plant Growth*. Edward Arnold, London.
- Crowder M. J., Hand D. J. (1991). *Analysis of Repeated Measures*. Chapman and Hall, London.
- Chinchilli Y.M., Carter W.H.Jr. (1984). A likelihood ratio test for a patterned covariance matrix in a multivariate growth curve model. *Biometrics* **40**, 151-156.
- Fujikoshi, Y., Rao C.R. (1991). Selection of covariables in growth curve model. *Biometrika* **78**, 779-785.
- Geisser S. (1980). Growth curve analysis. In *Handbook of Statistics*, Vol. 1, P.R. Krishnaiah (ed), North-Holland, Amsterdam, 89-115.
- Grizzle J.E., Allen D.M. (1969). Analysis of growth and dose response curves. *Biometrics* **25**, 357-381.
- Hui S.L. (1984). Curve fitting for repeated measurements made at irregular time-points. *Biometrics* **40**, 691-697.
- Jones R.H., Boadi-Boateng F. (1991). Unequally spaced longitudinal data with AR(1) serial correlation. *Biometrics* **43**, 161-175.
- Khatri G.G. (1966). A note on a manova model applied to problems growth curve. *Annals of the Institute of Statistical Mathematics* **18**, 75-86.
- Kshirsagar A.M. (1972). *Multivariate Analysis*. Marcel Dekker, Inc. New York.
- Laird N.M., Ware J.H. (1982). Random effects models for longitudinal data. *Biometrics* **38**, 963-974.
- Lee J.C. (1988). Prediction and estimation of growth curves with special covariance structures. *JASA* **83**, 432-440.
- Lee J.C. (1991) Tests and model selection for the general growth curve model. *Biometrics* **47**, 147-159.
- Lee Y.H.K. (1971). A note on Rao's reduction of Potthoff and Roy's generalized linear model. *Biometrika* **61**, 349-351.
- Lee J.C., Geisser S. (1972). Growth curve prediction. *Sankhyā Ser. A*, **34**, 393-412.
- Liski E.P. (1985). Estimation from incomplete data in growth curves models. *Commun. Statist. - Simula. Computa.* **14**, 13-27.
- Liski E.P., Nummi T. (1988). Prediction in Growth Curve Models. Report A 203 University of Tampere, Finland.
- Lundbye-Christensen S. (1991). A multivariate growth curve model for pregnancy. *Biometrics* **47**, 637-657.
- Morrison D.F. (1976). *Multivariate Statistical Methods*. 2nd ed., McGraw-Hill, New York.
- Potthoff R.F., Roy S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51**, 313-326.
- Rao C.R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika* **52**, 447-458.

- Rao C.R. (1977). Prediction of future observations with special reference to linear models. In *Multivariate Analysis IV*, Krishnaiah P.R. (ed), North-Holland Publishing Co, 193-208.
- Rao C.R., Boudreau R. (1985). Prediction of future observations in a factor analytic type growth model. In *Multivariate Analysis VI*, Krishnaiah P.R. (ed), Elsevier Science Publishers B.V., 449-466.
- Rash D. (1988). Growth curve analysis - A review and recent results. Probleme der angewandten Statistik. In *Proceedings of the International Conference on Population Mathematics*. Part I. Akademie der Landwirtschaftswissenschaften der DDR Forschungszentrum für Tierproduktion, Dummerstorf-Rostock.
- Seber G.A.F. (1984). *Multivariate Observations*, John Wiley, New York.
- Stanek III E.J., Koch G.G. (1985). The equivalence of parameter estimates from growth curve models and seemingly unrelated regression models. *The American Statistician*. **39**, 149-152.
- Timm N.H. (1980). Multivariate analysis of variance of repeated measurements. In *Handbook of Statistics*., Vol 1, Krishnaiah P.R. (ed), 41-87.
- Tubbs J.D., Levis T.O., Duran, B.S. (1975). A note on the analysis of the MANOVA model and its application to growth curves. *Comm. Statist.* **4**, 643-653.
- Vonesh E. F., Carter R.L. (1987). Efficient inference for random coefficient growth curve models with unbalanced data. *Biometrics* **43**, 617-628.
- Wesołowska-Janczarek M., Milczak H. (1989). Uwagi o zastosowaniu jedno- i wielozmiennej metody do wyznaczania dynamiki zawartości wapnia w liściach jabłoni. *XIX Coll. z Agrobiometrii*, PAN, Warszawa, 200-209.
- Wesołowska-Janczarek M. (1991). Equivalence of the estimators obtained by two multivariate regression methods. In *Proceedings of the 10th International Conference on Probability and Mathematical Statistics*. Pazman, A. and Volaufova, J. (ed), 75-79.
- Wójcik A.R., Ludański Z. (1989). *Planowanie i wnioskowanie statystyczne w doświadczalnictwie*. PWN, Warszawa.
- Zellner A. (1962). An efficient method of estimating seemingly unrelated regression and tests for aggregation bias. *J. Amer. Statist. Assoc.* **57**, 348-368.
- Zerbe G.O., Walker S.H. (1977). A randomization test for comparison of groups of growth curves with different polynomial design matrices. *Biometrics* **33**, 653-657.

*Praca wpłynęła 18 stycznia 1993;  
w wersji ostatecznej 25 października 1993*

## Growth curves – a review and new problems

### Summary

A review of the known growth curve methods and some of the new subjects of this topic are presented in this paper. The connections of the growth curve methods with the stochastic processes and other statistical problems are suggested.

**Key words:** growth curve, random coefficients, estimation of parameters, tests, confidence regions.