Uniwersytet Mikołaja Kopernika
w Toruniu
Wydział Matematyki i Informatyki

Oleksandr Zaihraiev

Teoria optymalnego planowania doświadczeń

Wykład do wyboru

Toruń 2000

Plan wykładu do wyboru – 30 godzin

Teoria Optymalnego Planowania Doświadczeń

1. Analiza regresji liniowej:
   – estymacja parametrów regresji w modelu klasycznym;
   – sprawdzanie prostych hipotez o parametrach regresji w modelu klasycznym;
   – ogólny przypadek.

2. Planowanie doświadczeń w modelach regresji liniowej:
   – plany dyskretne (ścisłe i uogólnione) i ciągłe;
   – kryteria optymalności planów eksperymentu;
   – twierdzenia równoważności i niektóre ich konsekwencje;
   – plany D-optymalne dla regresji wielomianowej;
   – metody numeryczne konstrukcji planów optymalnych.

3. Doświadczenia czynnikowe:
   – plany czynnikowe (pełne i ułamkowe);
   – relacje generujące i określające.

4. Wybrane pytania analizy wariancji:
   – analiza jednoczynnikowa;
   – analiza dwoczynnikowa.

Literatura

1. Rafajłowicz E. *Algorytmy planowania eksperymentu z implementacjami w środowisku Mathematica*, 1996. Rozdz. 2–4.
2. Silvey S.D. *Wnioskowanie statystyczne*, 1978. Rozdz. 3.
3. Zieliński R. *Siedem wykladów wprowadzających do statystyki matematycznej*, 1990. Rozdz. 6.
4. Rao C.R., Toutenburg H. *Linear Models. Least Squares and Alternatives*, 1995. Rozdz. 3,5.
5. Rao C.R. *Modele liniowe statystyki matematycznej*, 1982. Rozdz. 1,4.
6. Borovkov A.A. *Matematiczeskaja statistika. Dopolnit. glavy*, 1984. Rozdz. 1.3,1.4.
7. Wilks S. *Matematiczeskaja statistika*, 1967. Rozdz. 10.
8. Oktaba W. *Elementy statystyki matematycznej i metodyka doświadczalnictwa*, 1977. Rozdz. 11–13.
9. Shah K.R., Sinha B.K. *Theory of Optimal Designs*, 1989. Rozdz. 1,2.
10. Krasovskij G.I., Filaretov G.F. *Planirovanie eksperimenta*, 1982. Rozdz. 3–5.
11. Seber G. *Linejnyj regressionnyj analiz*, 1980. Rozdz. 3,4,9.
12. Daniel C. *Primenenie statistiki v promyszlennom eksperimente*, 1979. Rozdz. 3,11.
13. Scheffe H. *Dispersionnyj analiz*, 1980. Rozdz. 3–5.

Spis treści

1 Historia. Klasyczny model regresji liniowej. Twierdzenie Gaussa-Markova. 3

2 Estymacja parametru σ^2 w modelu klasycznym. Sprawdzanie prostych hipotez o współczynnikach regresji. 11

3 Przypadek niespełnienia założeń modelu klasycznego regresji liniowej. 17

4 Planowanie doświadczeń. Plany dyskretne i ciągłe. Macierz informacji i macierz kowariancji. Własności macierzy informacji. 23

5 Kryteria optymalności. Przykłady. 27

6 Twierdzenia równoważności, przykłady ich używania. 34

7 D-optymalne plany dyskretne uogólnione dla regresji wielomianowej. 42

8 Metody numeryczne konstrukcji planów optymalnych. 45

9 Doświadczenia czynnikowe. 54

10 Wybrane pytania analizy wariancji. 60

11 Pytania do teoretycznej części egzaminu 67

12 Zadania do praktycznej części egzaminu 70

1 Historia. Klasyczny model regresji liniowej. Twierdzenie Gaussa-Markova.

Historia

Teoria optymalnego planowania doświadczeń (TOPD) jest nauką porównalno mlodą. Jej formowanie zaczelo się $60-70$ lat temu, najpierw głównie na zapotrzebowanie metod statystyki stosowanej w rolnictwie, biologii, medycynie. Założycielem teorii uważany jest angielski statystyk *R.A.Fisher*, a podstawowej pracej – jego książka *The Design of Experiments*, (London, 1935). W ogóle, cały szereg prac Fishera byl poświęcony metodam analizy statystycznej i planowania doświadczeń w rolnictwie. Z tych prac potem wzieli swój początek takie działy statystyki matematycznej, związane z TOPD, jak analiza wariancji i planowanie czynnikowe. Pracy Fishera w zasadzie udowodnili banalną istyne : racjonalne planowanie doświadczeń daje efekt nie mniejszy niż optymalna obróbka danych.

Istnienie dużej ilości zadań praktycznych stymulowalo rozwój matematycznej teorii optymalnego doświadczenia. Znaczące wnioski do TOPD byli zrobione w latach 50-ch oraz 60-ch przez statystyków amerykańskich Box'a i Kiefera. Ten ostatni uważa się za twórce klasycznej TOPD, której też będziemy się zajmowaly.

Oczywiście, TOPD bardzo rozrosla się za miniony okres. Będziemy się starać ogarnąc klasyczne wyniki jak również i względnie nowe.

Literatura proponowana zawiera cały szereg pozycji dostępnych w bibliotece Wydziału (patrz Plan wykładu). Niestety ze względu na bardzo ograniczony wybór niema takiej książki, którą można by bylo polecić jako podstawową. Najbardziej blizkimi do treści kursu są ksiażki:

- F.Pukelsheim *Optimal Design of Experiments*, New York, 1993;
- S.M. Ermakov, A.A. Żigljavskij *Matematiczeskaja Teorija Optimalnogo Eksperimenta*, Moskva, 1987.

Ogólnie mówiąc, TOPD jest podrozdzialem Statystyki Matematycznej. W trakcie wykładu będą nam także potrzebne pewne fakty z :

- Algebry liniowej;
- Metod optymalizacji.

Klasyczny model regresji liniowej

Zajmiemy się teraz problemem, który bardzo często się powstaje przy badaniu pewnych zjawisk i obróbce danych. Jest to problem znależenia zależności pomiędzy pewnymi wielkościami, czy też zmiennymi. Najczęściej jedyną możliwościa wykrycia prawdziwej zależności jest prowadzenie eksperymentu i tylko w nielicznych przypadkach zależność ta może być uzyskana teoretycznie.

Zwykle postępujemy tutaj w sposób następujący. Niech istnieje pewna zależność pomiędzy zmiennej y (zmiennej zależnej) i $m \geq 1$ zmiennymi $x^{(1)}, \ldots, x^{(m)}$ (zmiennymi niezależnymi). Przykłady zależności : wydajność ziemi od różnych nawozów mineralnych, poziom zysku banku od ilości klientów, wysokości inwestycji, ilości kredytów i t.d.

Załóżmy, że możemy przeprowadzić n eksperymentów, n pomiarów wielkości y w zależności od pewnych wielkości $x^{(1)}, \ldots, x^{(m)}$. Żeby zadanie poszukiwania zależności bylo sensowne (rozwiązywalne), zwięzymy to poszukiwanie do pewnej rodziny funkcji H, tj. będziemy uważać, że prawdziwa zależność opisuje się funkcjej z klasy H. Oczywiście, powstaje przy tym poważny problem wyboru rodziny funkcji H. Pewnymi wskazówkami dla takiego wyboru mogą być : pewna wiedza o charakterze zależności lub poszukiwanie rozwiązania jaknajprostszego.

Najczęściej, za rodzine funkcji H bierze się rodzina parametryczna, tj.

$$H = \{g(x, \theta),\ \theta \in R^k,\ x = (x^{(1)}, \ldots, x^{(m)}) \in R^m\},$$

gdzie $g\ :\ R^m \times R^k \to R$ jest zadaną funkcją. Tzn. uważa się, że prawdziwą zależnościa jest $y = g(x, \theta)$ dla pewnej wartości parametru θ. Funkcja $g(x, \theta)$ nazywa się *funkcjej regresji* (nazwa ta jest historyczna; nadal ją angielski statystyk F.Galton badając problemy dziedziczności). Taka zależność odzwierciedla różną role argumentów x i θ funkcji regresji. Uważamy, że x odpowiada za warunki przeprowadzenia eksperymentów i może być wybrane na początku. Natomiast θ jest argumentem nieznanym. W przypadku opisanym problem znależenia nieznanej zależności sprowadza się do znależenia θ.

Jeśli funkcja g jest liniowa względem θ, to regresja nazywa się *liniowej*, w przeciwnym wypadku – *nieliniowej*. Ogólna postać funkcji regresji liniowej, to :

$$g(x, \theta) = \sum_{j=1}^{k} \theta_j f_j(x),$$

gdzie $f_j : R^m \to R$ są funkcjami znanymi (w najprostszym przypadku, kiedy $f_j(x) = x^{(j)}$, mamy $g(x, \theta) = \sum_{j=1}^{k} \theta_j x^{(j)}$).

Opisana zależność jest zależnościa *determinystyczną*. Bardzo rzadko jednak mamy do czynienia z danymi, które dokladnie pasują do takiej zależności. Ale okazuje się, że podobne podejście z malą modyfikacją można stosować i w przypadku najczęściej spotykanym, kiedy dane nie pasują do danej zależności. Ta modyfikacja ma nazwe zależności *statystycznej*. Otóż uważamy, że wartości y są obserwowane z dokladnościa do pewnych losowych składników e, nazywanych *błędami pomiarów*. Tak naprawde, żródla tych błędów mogą być bardzo różne : od błędów losowych spowodowanych nieprecyzyjnościa urządzień pomiarowych do błędów spowodowanych nieprawidlowościa wyboru klasy H. W przypadku opisanym problem znależenia zależności sprowadza się do znależenia takiego θ, dla którego funkcja regresji opisuje otrzymane dane dość dobrze (lub wręcz najlepiej).

Więc w przypadku regresji liniowej, zakladamy, że mamy obserwacje

$$y_i = \sum_{j=1}^{k} \theta_j f_j(x_i) + e_i, \qquad i = 1, 2, \ldots, n,$$

gdzie $\theta = (\theta_1, \ldots, \theta_k)^T$ jest wektorem nieznanych parametrów rzeczywistych (współczynników regresji), $\{f_j\}$ są znanymi funkcjami, $\{x_i\}$ są zadanymi wektorami, $\{e_i\}$ są błędami pomiarów. Co do tych błędów, często zakładają, że

$$Ee_i = 0, \; Ee_i e_j = \begin{cases} \sigma^2 & i = j \\ 0 & i \neq j \end{cases},$$

gdzie $\sigma > 0$ jest nieznaną stalą. Tj. macierz kowariancji $Cov\ e$ wektora $e = (e_1, \ldots, e_n)^T$ ma postać $\underline{Cov\ e = \sigma^2 I_n}$, gdzie I_n jest macierzą jednostkową wymiaru $n \times n$. Poznaczmy $y = (y_1, \ldots, y_n)^T$,

$$F = \begin{pmatrix} f_1(x_1) & \ldots\ldots & f_k(x_1) \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ f_1(x_n) & \ldots\ldots & f_k(x_n) \end{pmatrix}$$

- prostokątna macierz wymiaru $n \times k$, która się nazywa *macierzą planu*. Będziemy dalej rozważać przypadek, kiedy $n \geq k$ i macierz F ma pełny rząd, tj. $\underline{r(F) = k}$. Dwa warunki podkreślone definują tak zwany *model klasyczny*.

W nowych poznaczeniach mamy $y = F\theta + e$, przy czym $Ey = F\theta$, $Cov\ y = \sigma^2 I_n$. Więc, model klasyczny zwykle się zapisuje za pomocą trójki $(y, F\theta, \sigma^2 I_n)$,

gdzie y jest wektorem obserwacji, $F\theta$ jest jego wartością oczekiwaną, $\sigma^2 I_n$ jest jego macierzą kowariancji.

Następnym problemem po wyborze typu funkcji regresji $g(x,\theta)$ jest znalezienie takich wartości parametru θ, dla których funkcja regresji będzie dość dobrze opisywać otrzymane dane. Tutaj będziemy potrzebowaly kilka pojeć ze Statystyki Matematycznej.

Definicja 1.1 *Każde odwzorowanie mierzalne z R^n do R^k (lub ze zbioru Y wartości y do zbioru Θ wartości θ) niezależne od nieznanych parametrów nazywamy statystyką.*

Definicja 1.2 *Każdą statystykę $\hat{\theta}(y)$, której wartości traktujemy jako przybliżenia dla θ, nazywamy estymatorem parametru θ.*

Definicja 1.3 *Estymator $\hat{\theta}(y)$ nazywa się liniowym, jeśli funkcja $\hat{\theta}(y)$ jest funkcjej liniowej zmiennej y.*

Definicja 1.4 *Estymator $\hat{\theta}(y)$ nazywa się nieobciążonym, jeśli*

$$E\hat{\theta}(y) = \theta, \qquad \forall\theta.$$

Definicja 1.5 *Estymator $\hat{\theta}(y)$ nazywamy najlepszym liniowym nieobciążonym estymatorem (NLNE), jeśli jest on liniowym i nieobciążonym oraz $Cov\ \hat{\theta}(y) \leq Cov\ \theta^*(y)$ dla każdego innego estymatora liniowego nieobciążonego $\theta^*(y)$.*

Uwaga. Nierówność $A \leq B$ pomiędzy dwoma kwadratowymi i symetrycznymi macierzami A i B tutaj i dalej rozumie się w tym sensie, że macierz $B - A$ jest nieujemnie określona.

Uwaga. Relacja $Cov\ \hat{\theta}(y) \leq Cov\ \theta^*(y)$ jest równoważna relacji $\forall l \in R^k \quad Var\ l^T\hat{\theta}(y) \leq Var\ l^T\theta^*(y)$. Rzeczywiście,

$$Cov\,\hat{\theta}(y) \leq Cov\,\theta^*(y) \iff \forall l \in R^k \quad l^T E(\hat{\theta}-\theta)(\hat{\theta}-\theta)^T l \leq l^T E(\theta^*-\theta)(\theta^*-\theta)^T l$$

$$\iff \forall l \in R^k \quad E(l^T(\hat{\theta}-\theta))^2 \leq E(l^T(\theta^*-\theta))^2 \iff \forall l \in R^k \quad Var\, l^T\hat{\theta} \leq Var\, l^T\theta^*.$$

Dla rozwiązania problemu estymacji zastosujmy tak zwaną *metodę najmniejszych kwadratów* (MNK), wprowadzoną jeszcze w XIX wieku przez Gaussa i Legandre'a. Polega ona na tym, że mając pomiary $\{(x_i, y_i)\}$ wybieramy $\hat{\theta}$ takie, że

$$\hat{\theta} = Arg\ \min_{\theta}\ (y - F\theta)^T(y - F\theta) = Arg\ \min_{\theta}\ \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{k}\theta_j f_j(x_i)\right)^2$$

tzn.

$$\sum_{i=1}^{n}\left(y_i-\sum_{j=1}^{k}\widehat{\theta}_j f_j(x_i)\right)^2=\min_{\theta}\sum_{i=1}^{n}\left(y_i-\sum_{j=1}^{k}\theta_j f_j(x_i)\right)^2.$$

Jaki jest sens w tej metodzie ? Otóż wybieramy w rodzinie $H=\{\sum_{j=1}^{k}\theta_j f_j(x),\ \theta\in R^k,\ x\in R^m\}$ taką funkcję, dla której minimizuje się suma kwadratów różnić pomiędzy obserwacjami $\{y_i\}$ a wartościami funkcji regresji w punktach $\{x_i\}$ (błąd średniokwadratowy).

Uwaga. Poznaczmy $f_j=(f_j(x_1),\ldots,f_j(x_n))^T\in R^n,\ j=1,2,\ldots,k$. Niech $\mathcal{F}=\overline{\{f_1,\ldots,f_k\}}$ – k-wymiarowa podprestrzeń R^n rozpięta nad wektorami $\{f_j\}$. Wtedy

$$\sum_{j=1}^{k}\theta_j f_j(x_i)=\sum_{j=1}^{k}\theta_j f_j,\qquad \theta\in R^k$$

jest dowolnym wektorem z $\mathcal{F}$, a $\|y-\sum_{j=1}^{k}\theta_j f_j\|$ jest odległościa w przestrzeni R^n pomiędzy wektorem y a wektorem z podprzestrzeni $\mathcal{F}$. MNK polega na wyborze takiego wektora $\sum_{j=1}^{k}\widehat{\theta}_j f_j\in\mathcal{F}$, który minimalizuje tą odległość. Z Algebry liniowej wiadomo, że $\sum_{j=1}^{k}\widehat{\theta}_j f_j=P_{\mathcal{F}}y$ (rzut y na $\mathcal{F}$).

Twierdzenie Gaussa-Markova. *W ramach klasycznego modelu regresji $(y,F\theta,\sigma^2 I_n)$ następujące stwierdzenia są sprawiedliwe :*

1. *MNK estymatorem dla θ jest liniowy estymator $\widehat{\theta}=(F^TF)^{-1}F^Ty$;*
2. *$E\widehat{\theta}=\theta$, a więc estymator MNK $\widehat{\theta}$ jest nieobciążony, oraz $Cov\ \widehat{\theta}=\sigma^2(F^TF)^{-1}$;*
3. *MNK estymator $\widehat{\theta}$ jest NLNE dla θ.*

Dowód. 1. Poznaczmy $Q(\theta)=(y-F\theta)^T(y-F\theta)$. Żeby znaleźć minimum $Q(\theta)$ policzmy pochodne cząstkowe :

$$\frac{\partial Q(\theta)}{\partial\theta_l}=\frac{\partial}{\partial\theta_l}\left(\sum_{i=1}^{n}\left(y_i-\sum_{j=1}^{k}\theta_j f_j(x_i)\right)^2\right)=$$

$$2\sum_{i=1}^{n}\left(y_i-\sum_{j=1}^{k}\theta_j f_j(x_i)\right)\frac{\partial}{\partial\theta_l}\left(y_i-\sum_{j=1}^{k}\theta_j f_j(x_i)\right)=$$

$$2\sum_{i=1}^{n}\left(y_i-\sum_{j=1}^{k}\theta_j f_j(x_i)\right)(-f_l(x_i))=2\left(\sum_{i=1}^{n}\sum_{j=1}^{k}\theta_j f_j(x_i)f_l(x_i)-\sum_{i=1}^{n}y_i f_l(x_i)\right)$$

$$\Longrightarrow \left(\frac{\partial Q(\theta)}{\partial \theta_1}, \ldots, \frac{\partial Q(\theta)}{\partial \theta_k}\right)^T = 2\left(F^T F\theta - F^T y\right).$$

Przyrównując wszystkie pochodne cząstkowe do zera, otrzymujemy równanie $F^T F\theta = F^T y$ nazywane *układem równań normalnych*. Przy naszych założeniach macierz $F^T F$ jest odwracalna (**Zadanie domowe.** Uzasadnić tą teze.) i układ równań normalnych posiada tylko jedno wskazane rozwiązanie.

Uwaga. Macierz $P_{\mathcal{F}}$ rzutu ortogonalnego na podprzestrzeń $\mathcal{F}$ z liniowo niezależną bazą $\{f_j\}_{j=1}^k$ ma postać $P_{\mathcal{F}} = F(F^T F)^{-1}F^T$ (przypominamy : $F = (f_1,\ f_2,\ \ldots,\ f_k)$). Spawdzmy to.
(i) Musimy sprawdzić, że $P_{\mathcal{F}} z = z$ dla $z \in \mathcal{F}$. Rzeczywiście, jeśli $z \in \mathcal{F}$, to istnieje $a \in \mathcal{F}$ takie, że $z = \sum_{j=1}^k a_j f_j = Fa$. Wtedy $F(F^T F)^{-1}F^T z = F(F^T F)^{-1}F^T Fa = Fa = z$.
(ii) Musimy też sprawdzić, że $P_{\mathcal{F}} z = 0$ dla $z \perp \mathcal{F}$. Rzeczywiście, jeśli $z \perp \mathcal{F}$, to $f_j^T z = 0$ dla wszystkich j, więc $F^T z = 0 \implies F(F^T F)^{-1}F^T z = 0$.
A więc zgadza się to, że $P_{\mathcal{F}} y = F\hat{\theta}$.

2. Mamy

$$E\hat{\theta} = E\left((F^T F)^{-1}F^T y\right) = E\left((F^T F)^{-1}F^T F\theta + (F^T F)^{-1}F^T e\right) = \theta,$$

$$Cov\ \hat{\theta} = E(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T = E\left((F^T F)^{-1}F^T e\right)\left((F^T F)^{-1}F^T e\right)^T =$$

$$(F^T F)^{-1}F^T Cov\ eF(F^T F)^{-1} = \sigma^2(F^T F)^{-1}F^T F(F^T F)^{-1} = \sigma^2(F^T F)^{-1}.$$

3. Niech $\theta^*(y)$ jest dowolnym estymatorem liniowym nieobciążonym dla θ. Wtedy istnieje $k \times n$-macierz L taka, że $\theta^* = Ly$ oraz $E\theta^* = LF\theta = \theta\ \forall\theta \implies LF = I_k$ – warunek nieobciążoności estymatora θ^*. Udowodnimy, że $Cov\ \theta^* \geq Cov\ \hat{\theta}$. Rozważmy macierz $A = (F^T F)^{-1}F^T - L$. Mamy $AF = I_k - I_k = 0$ (z warunku nieobciążoności). Teraz

$$Cov\ \theta^* = E(Ly - \theta)(Ly - \theta)^T = E(Le)(Le)^T = LCov\ eL^T = \sigma^2 LL^T =$$

$$\sigma^2\left((F^T F)^{-1}F^T - A\right)\left((F^T F)^{-1}F^T - A\right)^T =$$

$$\sigma^2\left((F^T F)^{-1} - (F^T F)^{-1}F^T A^T - AF(F^T F)^{-1} + AA^T\right) =$$

$$Cov\ \hat{\theta} + \sigma^2 AA^T \geq Cov\ \hat{\theta},$$

ponieważ $\sigma^2 AA^T \geq 0$. $\square$

Wniosek 1.1 *W ramach klasycznego modelu regresji estymator* $\sum_{j=1}^{k} \widehat{\theta}_j f_j(x)$ *jest liniowym i nieobciążonym estymatorem dla wartości funkcji regresji* $\sum_{j=1}^{k} \theta_j f_j(x)$ *w zadanym punkcie* x *mającym minimalną wariancje spośród wszystkich estymatorów liniowych i nieobciążonych.*

Przykład 1. Dopasować prostą do punktów $(x_i, y_i) = (0,0),\ (1/2, 1/4),\ (1,1)$ za pomocą MNK.

Mamy $g(x,\theta) = \theta_1 + \theta_2 x\ (m = 1,\ k = 2),\ n = 3$. Wtedy

$$F = \begin{pmatrix} 1 & 0 \\ 1 & 1/2 \\ 1 & 1 \end{pmatrix}, \qquad y = \begin{pmatrix} 0 \\ 1/4 \\ 1 \end{pmatrix} \qquad \Longrightarrow$$

$$F^T F = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1/2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1/2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 3/2 \\ 3/2 & 5/4 \end{pmatrix}, \qquad det\ F^T F = 3/2,$$

$$(F^T F)^{-1} = \begin{pmatrix} 5/6 & -1 \\ -1 & 2 \end{pmatrix}, \qquad F^T y = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1/2 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1/4 \\ 1 \end{pmatrix} = \begin{pmatrix} 5/4 \\ 9/8 \end{pmatrix},$$

$$\widehat{\theta} = (F^T F)^{-1} F^T y = \begin{pmatrix} 5/6 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 5/4 \\ 9/8 \end{pmatrix} = \begin{pmatrix} -1/12 \\ 1 \end{pmatrix} \qquad \Longrightarrow$$

najlepszej pasującej do danych rozważanych prostej, zgodnie z MNK, jest $y = -1/12 + x$.

Przykład 2. Rozważmy to samo zadanie w przypadku dowolnej ilości n punktów. Wtedy

$$F = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \qquad F^T F = \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix},$$

$$det\ F^T F = n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2,$$

$$(F^T F)^{-1} = (det\ F^T F)^{-1} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix},$$

$$F^T y = \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix},$$

$$\widehat{\theta} = (F^T F)^{-1} F^T y = (det\ F^T F)^{-1} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix} =$$

$$(det\ F^T F)^{-1} \begin{pmatrix} \sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i y_i \sum_{i=1}^n x_i \\ n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i \end{pmatrix} \qquad \Longrightarrow$$

$$\widehat{\theta}_1 = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}, \qquad \widehat{\theta}_2 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}.$$

Równoważnie możemy zapisać

$$\widehat{\theta}_2 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2}, \qquad \widehat{\theta}_1 = \bar{y} - \widehat{\theta}_2 \bar{x},$$

gdzie $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,\ \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.

Jeśli wybieramy takie punkty $\{x_i\}$, że $\bar{x} = 0$, to wszystkie formuly się upraszczają i mamy

$$\widehat{\theta}_1 = \bar{y}, \qquad \widehat{\theta}_2 = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i}{\frac{1}{n} \sum_{i=1}^n x_i^2}.$$

2 Estymacja parametru σ^2 w modelu klasycznym. Sprawdzanie prostych hipotez o współczynnikach regresji.

Estymacja parametru σ^2 w modelu klasycznym

Poznaczmy

$$\varepsilon = y - F\hat{\theta} = y - F(F^TF)^{-1}F^Ty = y - P_{\mathcal{F}}y = (I_n - P_{\mathcal{F}})y.$$

Wektor ε nazywa się *wektorem reszt*. Jest to różnica pomiędzy zaobserwowanymi wartościami a predykcjej funkcji regresji w odpowiednich punktach (błąd predykcji).

Uwaga. Jak łatwo zrozumieć, w przypadku $n = k$ mamy $\varepsilon = 0$. Oznacza to dokladne dopasowanie modelu do danych. Naprzykład jeśli $g(x, \theta) = \theta_1 + \theta_2 x$ $(m = 1, k = 2)$, mamy tylko dwie obserwacje i możemy poprowadzić prostą przez te dwa punkty bezbłędnie. Ale nie znaczy to wcale, że wykonaliśmy zadanie znależenia zależności perfekcyjnie. Przecież gdyby mielibyśmy jeszcze obserwacje, nie musialy by one pasować do tego samego modelu. Więc oznacza to tylko, że mamy zamalo informacji. Na dowód tego mówi i to, że nie możemy w tym przypadku oszacować σ^2 (jak zobaczymy dalej).

Otóż dalej w tym Rozdziale zakładamy, że $n > k$.

Lemat 2.1 *Statystyka* $s^2 = \frac{1}{n-k}\varepsilon^T\varepsilon$ *jest nieobciążonym estymatorem dla* σ^2.

Dowód. Mamy

$$\varepsilon = (I_n - P_{\mathcal{F}})(F\theta + e) = (I_n - P_{\mathcal{F}})e.$$

Macierz $G = I_n - P_{\mathcal{F}}$ spełnia warunki : $G = G^T$, $G^2 = G$. Jest to tzw. *macierz idempotentna* (jest to także macierz rzutu na podprzestrzeń $\mathcal{F}^{\perp}$). Więc,

$$\varepsilon^T\varepsilon = e^TG^TGe = e^TGe$$

oraz

$$E\varepsilon^T\varepsilon = E(tr\, Gee^T) = tr\,(GEee^T) = \sigma^2 tr\, G = \sigma^2\,(tr\ I_n - tr\ P_{\mathcal{F}}) = \sigma^2(n-k). \square$$

Zadanie domowe. Sprawdzić, że $x^TAx = tr\ Axx^T, \quad tr\ AB = tr\ BA.$

Przykład. Niech $g(x,\theta) = \theta_1 + \theta_2 x$ $(m=1, k=2)$ oraz $\sum_{i=1}^n x_i = 0 \implies$ mamy estymatory jak z Przykładu 2 Rozdzialu 1. Więc,

$$\varepsilon = y - F\hat{\theta} = \begin{pmatrix} y_1 - \hat{\theta}_1 - \hat{\theta}_2 x_1 \\ \ldots\ldots\ldots\ldots\ldots \\ y_n - \hat{\theta}_1 - \hat{\theta}_2 x_n \end{pmatrix} \implies s^2 = \frac{1}{n-2}\sum_{j=1}^n \left(y_j - \hat{\theta}_1 - \hat{\theta}_2 x_j\right)^2$$

$$= \frac{1}{n-2}\sum_{j=1}^n \left(y_j - \frac{1}{n}\sum_{i=1}^n y_i - x_j \sum_{i=1}^n x_i y_i \left(\sum_{i=1}^n x_i^2\right)^{-1}\right)^2 .$$

Definicja 2.1 *Niech $y = (y_1, \ldots, y_n)^T$ są wartościami zaobserwowanymi i łączny rozkład wektora y ma gęstość $p_\theta(u_1, \ldots, u_n)$, gdzie θ jest wektorem nieznanych parametrów. Wtedy funkcja*

$$L(y,\theta) = p_\theta(y_1, \ldots, y_n)$$

nazywa się funkcjej wiarogodności. Statystyka

$$\breve{\theta}(y) = Arg \max_\theta L(y,\theta)$$

nazywa się estymatorem największej wiarogodności (ENW) dla θ.

Uwaga. Często łatwiej jest szukać nie $\max_\theta L(y,\theta)$, lecz $\max_\theta \ln L(y,\theta)$.

Twierdzenie 2.1 *Jeśli w klasycznym modelu regresji liniowej zmienne losowe $\{e_i\}$ mają rozkłady normalne, to MNK estymator $\hat{\theta}$ jest ENW dla θ, natomiast $\frac{n-k}{n}s^2 = \frac{1}{n}\varepsilon^T\varepsilon$ jest ENW dla σ^2.*

Dowód. Mamy $y \sim N_n\left(F\theta, \sigma^2 I_n\right)$. Więc,

$$L(y,\theta,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(y-F\theta)^T(y-F\theta)\right).$$

Szukamy teraz maksimum funkcji $\ln L$ poprzez wyliczanie pochodnych cząstkowych i przyrównanie ich do zera. Mamy

$$\begin{cases} \frac{\partial \ln L}{\partial \theta} = \frac{1}{\sigma^2}F^T(y-F\theta) = 0 \\ \frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y-F\theta)^T(y-F\theta) = 0 \end{cases} \implies \begin{cases} \breve{\theta} = \hat{\theta} \\ \breve{\sigma^2} = \frac{1}{n}\varepsilon^T\varepsilon \end{cases} . \ \square$$

Uwaga. ENW dla σ^2 jest obciążony, ponieważ $E\breve{\sigma^2} = \frac{n-k}{n}\sigma^2$. Ale jest on asymptotycznie nieobciążony, ponieważ $E\breve{\sigma^2} \to \sigma^2$ kiedy $n \to \infty$.

Przykład. Niech $g(x,\theta) = \theta$ $(m = 0, k = 1)$, a więc $y_j = \theta + \xi_j$, tzn. $y \sim N_n\left(\theta, \sigma^2 I_n\right)$. Wtedy $F^T = (1,\ldots,1)$, $F^T F = n$, $F^T y = \sum_{i=1}^n y_i \implies \breve{\theta} = \widehat{\theta} = \frac{1}{n}\sum_{i=1}^n y_i = \bar{y}$ – średnia z próby. Natomiast,

$$\breve{\sigma^2} = \frac{1}{n}(y - F\widehat{\theta})^T(y - F\widehat{\theta}) = \frac{1}{n}\sum_{i=1}^n \left(y_i - \widehat{\theta}\right)^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2$$

– wariancja z próby. Otrzymujemy więc znany wynik : ENW dla θ i σ^2 w przypadku próby z rozkładu $N(\theta, \sigma^2)$ są odpowiednio średnia z próby i wariancja z próby.

Sprawdzanie prostych hipotez o współczynnikach regresji

Niech w klasycznym modelu regresji liniowej zmienne losowe $\{e_i\}$ mają rozkłady normalne. Wtedy $\widehat{\theta} \sim N_n(\theta, \sigma^2(F^T F)^{-1})$.

Załóżmy, że chcemy sprawdzić prostą hipoteze $H_i \;:\; \theta_i = a$ przeciwko $\theta_i \neq a$ (a jest zadane).

Jeśli H_i ma miejsce, to $\widehat{\theta}_i \sim N\left(a, \sigma^2((F^T F)^{-1})_{ii}\right)$.

Kiedy σ^2 jest znane, to przy H_i statystyka

$$t = \frac{\widehat{\theta}_i - a}{\sigma\sqrt{((F^T F)^{-1})_{ii}}} \sim N(0,1)$$

może być wybrana jako testowa dla sprawdzania danej hipotezy i mamy następująca procedure sprawdzania hipotezy H_i :

- zadajemy poziom ufności $1-\alpha$ (α jest prawdopodobieństwem odrzucić hipoteze kiedy jest ona prawdziwa);
- wyliczamy t_α takie, że $P\left(|\xi| < t_\alpha\right) = 1-\alpha,\ \xi \sim N(0,1)$;
- wyliczamy wartość statystyki t;
- jeśli $t \in (-t_\alpha, t_\alpha)$, to nie mamy podstaw odrzucić hipoteze H_i; w przeciwnym przypadku hipoteze H_i odrzucamy.

Kiedy σ^2 jest nieznane, to zamiast σ używamy s. Więc, bierzemy statystykę

$$\frac{\widehat{\theta}_i - a}{s\sqrt{((F^T F)^{-1})_{ii}}}$$

za testową. Znajdziemy jej rozkład.

Lemat 2.2 *Wartości własne macierzy idempotentnej są* 0 *lub* 1.

Dowód. Niech $\{\lambda_i\}$ oraz $\{x_i\}$ są odpowiednio wartościami własnymi i odpowiadającymi tym wartościam wektorami własnymi macierzy idempotentnej A. Mamy

$$A^2 x_i = AAx_i = A\lambda_i x_i = \lambda_i Ax_i = \lambda_i^2 x_i.$$

Z drugiej strony ponieważ $A^2 = A$,

$$A^2 x_i = Ax_i = \lambda_i x_i \implies \lambda_i^2 = \lambda_i. \ \square$$

Fakt 2.1 *Jeśli* $\xi_1, \ldots, \xi_n$ *są niezależnymi zmiennymi losowymi o jednakowym rozkładzie* $N(0,1)$, *to zmienna losowa* $\sum_{i=1}^n \xi_i^2$ *posiada rozkład o gęstości*

$$f(u) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} u^{n/2-1} e^{-u/2}, \ u > 0,$$

który nazywa się rozkładem χ_n^2 *– chi-kwadrat z* n *stopniami swobody.*

Lemat 2.3 *Niech* $\xi \sim N_n(0, I_n)$ *oraz* A *jest* $n \times n$*-macierzą idempotentną,* $r(A) = k \leq n$. *Wtedy zmienna losowa* $\xi^T A \xi$ *posiada rozkład* χ_k^2.

Dowód. Wykorzystamy spektralną dekompozycję symetrycznej macierzy A w postaci $A = C\Lambda C^T$, gdzie C – macierz ortogonalna, której kolumnami są wektory własne macierzy A, natomiast Λ – macierz diagonalna z wartościami własnymi $\{\lambda_i\}$ macierzy A na przekątnej, przy czym można uważać, że $\lambda_1 \geq \ldots \geq \lambda_n$. Zgodnie z Lematem 2.2 k pierwszych wartości spośród $\{\lambda_i\}$ sa jedynki, a pozostale zera. Więc,

$$\xi^T A \xi = \xi^T C \Lambda C^T \xi = \zeta^T \Lambda \zeta = \sum_{i=1}^k \zeta_i^2,$$

gdzie $\zeta = C^T \xi$. Wykorzystanie Faktu 2.1 konczy dowód, bo $\zeta \sim N_n(0, I_n)$. $\square$

Wniosek 2.1 *Jeśli w klasycznym modelu regresji liniowej zmienne losowe* $\{e_i\}$ *mają rozkłady normalne, to statystyka* $\varepsilon^T \varepsilon$ *posiada rozkład* $\sigma^2 \chi_{n-k}^2$.

Lemat 2.4 *Jeśli w klasycznym modelu regresji liniowej zmienne losowe* $\{e_i\}$ *mają rozkłady normalne, to statystyki* $\widehat{\theta}$ *i* $\varepsilon^T \varepsilon$ *są niezależne.*

Dowód. Mamy

$$\hat{\theta} = (F^T F)^{-1} F^T y = \theta + Ae, \quad \varepsilon^T \varepsilon = e^T G e,$$

gdzie $e \sim N_n\left(0, \sigma^2 I_n\right)$, $G = I_n - F(F^T F)^{-1} F^T$, $A = (F^T F)^{-1} F^T$. Wystarczy pokazać, że Ae i $e^T G e$ są niezależne.

Jak łatwo zrozumieć, $AG = 0$. Wykorzystamy teraz spektralną dekompozycję (patrz Lemat 2.3) $G = C \Lambda C^T$. Poznaczmy też $\zeta = C^T e$, $B = AC$. Wtedy $\zeta \sim N_n\left(0, \sigma^2 I_n\right)$ oraz

$$Ae = B\zeta, \qquad e^T G e = \zeta^T \Lambda \zeta = \sum_{i=1}^{n-k} \zeta_i^2 \qquad (\textit{Lemat 2.2}),$$

przy czym $B \Lambda C^T = 0 \Longrightarrow B\Lambda = 0$. To znaczy, że pierwsze $(n-k)$ kolumn macierzy B są zerami i

$$(B\zeta)_i = \sum_{j=n-k+1}^{n} B_{ij} \zeta_j,$$

a więc $B\zeta$ zależy tylko od $\zeta_{n-k+1}, \ldots, \zeta_n$. □

Fakt 2.2 *Jeśli $\xi \sim N(0,1)$, $\eta \sim \chi_n^2$, oraz ξ i η są niezależne, to zmienna losowa $\sqrt{n}\xi / \sqrt{\eta}$ posiada rozkład o gęstości*

$$f(u) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{u^2}{n}\right)^{-(n+1)/2}, \quad u \in R,$$

który nazywa się rozkładem t_n – Studenta o n stopniach swobody.

Z Wniosku 2.1, Lematu 2.4 i Faktu 2.2 wynika następujący wynik.

Lemat 2.5 *Jeśli w klasycznym modelu regresji liniowej zmienne losowe $\{e_i\}$ mają rozkłady normalne, to statystyka*

$$t = \frac{\hat{\theta}_i - a}{s\sqrt{((F^T F)^{-1})_{ii}}}$$

posiada rozkład t_{n-k}.

Dowód. Mamy

$$\frac{\hat{\theta}_i - a}{\sigma\sqrt{((F^T F)^{-1})_{ii}}} \sim N(0,1), \quad \frac{\varepsilon^T \varepsilon}{\sigma^2} \sim \chi_{n-k}^2 \Longrightarrow$$

$$\sqrt{n-k}\frac{\frac{\widehat{\theta}_i-a}{\sigma\sqrt{((F^TF)^{-1})_{ii}}}}{\sqrt{\frac{\varepsilon^T\varepsilon}{\sigma^2}}} \sim t_{n-k} \implies \frac{\widehat{\theta}_i-a}{\sqrt{\frac{1}{n-k}\varepsilon^T\varepsilon}\sqrt{((F^TF)^{-1})_{ii}}} \sim t_{n-k}. \ \square$$

Zadanie domowe. Napisać procedure sprawdzania hipotezy H_i w przypadku, kiedy σ^2 jest nieznane.

3 Przypadek niespełnienia założeń modelu klasycznego regresji liniowej.

Przypadek modeli, w których macierz kowariancji jest dodatnio określona ale nie jest diagonalną

W klasycznym modelu regresji liniowej zakładaliśmy, że macierz kowariancji $Cov\ y$ wektora obserwacji y ma postać $Cov\ y = \sigma^2 I_n$. Zrezygnujmy teraz z tego ograniczenia i załóżmy, że $Cov\ y = \sigma^2 W$, gdzie W jest znaną macierzą dodatnio określoną. Więc, mamy do czynienia z modelem $\left(y, F\theta, \sigma^2 W\right)$.

Łatwo jednak jest sprowadzić ten model do modelu klasycznego. Rzeczywiście, z dodatniej określoności W wynika, że zdefiniowana jest macierz $W^{-1/2}$. Niech $\tilde{y} = W^{-1/2}y$, $\tilde{F} = W^{-1/2}F$. Wtedy model $\left(y, F\theta, \sigma^2 W\right)$ jest równoważny modelu klasycznemu postaci $\left(\tilde{y}, \tilde{F}\theta, \sigma^2 I_n\right)$. Estymator MNK dla ostatniego modelu zgodnie z Twierdzeniem Gaussa-Markova (Rozdzial 1) zapisuje się w postaci

$$\tilde{\theta} = (\tilde{F}^T\tilde{F})^{-1}\tilde{F}^T\tilde{y} = (F^T W^{-1} F)^{-1} F^T W^{-1} y.$$

Estymator ten minimizuje po θ wartość

$$(\tilde{y} - \tilde{F}\theta)^T(\tilde{y} - \tilde{F}\theta) = (y - F\theta)^T W^{-1}(y - F\theta).$$

Układ równań normalnych w tym przypadku ma postać $F^T W^{-1} F\theta = F^T W^{-1} y$.

Estymator $\tilde{\theta}$ jest nieobciążony, a jego macierz kowariancji jest równa

$$Cov\ \tilde{\theta} = \sigma^2 (F^T W^{-1} F)^{-1}.$$

Estymator ten często się nazywa uogólnionym estymatorem MNK (dla odróżnienia od zwykłego estymatora MNK).

Uwaga. Zwykly nieobciążony estymator MNK równy $\hat{\theta} = (F^T F)^{-1} F^T y$ w tym przypadku ma macierz kowariancji

$$Cov\ \hat{\theta} = \sigma^2 (F^T F)^{-1} F^T W F (F^T F)^{-1}.$$

Zgodnie z Twierdzeniem Gaussa-Markova estymator $\tilde{\theta}$ jest NLNE. W szczególności, $Cov\ \tilde{\theta} \leq Cov\ \hat{\theta}$.

Wektor reszt definujemy tutaj jak poprzednio: $\varepsilon = y - F\tilde{\theta}$. Jak łatwo zrozumieć, teraz estymator $\tilde{s}^2 = \frac{1}{n-k}\varepsilon^T W^{-1}\varepsilon$ jest nieobciążonym dla σ^2. Również $\tilde{\theta}$ jest ENW dla θ, natomiast $\frac{n-k}{n}\tilde{s}^2$ jest ENW dla σ^2.

W przypadku sprawdzania prostych hipotez o współczynnikach regresji w tym modelu bazujemy procedure na następujących statystykach testowych:

- σ^2 jest znane:
$$t = \frac{\tilde{\theta}_i - a}{\sigma\sqrt{((F^T W^{-1} F)^{-1})_{ii}}} \sim N(0,1);$$
- σ^2 jest nieznane
$$t = \frac{\tilde{\theta}_i - a}{\tilde{s}\sqrt{((F^T W^{-1} F)^{-1})_{ii}}} \sim t_{n-k}.$$

Przypadek modeli, w których macierz planu jest niepełnego rzędu

Będziemy mówić tutaj o przypadku, kiedy model regresji liniowej $(y, F\theta, \sigma^2 W)$ jest taki, że $r(F) = p < k$. Dla takich modeli uogólniony estymator MNK też spełnia układ równań normalnych $F^T W^{-1} F\theta = F^T W^{-1} y$, ale teraz rozwiązanie tego układu nie jest jednoznaczne, bo z k równań zostanie nam tylko p równań, a niewiadomych mamy $k > p$.

Lemat 3.1 *Niech $\tilde{\theta}^{(1)}$ i $\tilde{\theta}^{(2)}$ są dwa rozwiązania układu równań normalnych (dwa uogólnionych estymatora MNK). Wtedy $F\tilde{\theta}^{(1)} = F\tilde{\theta}^{(2)}$.*

Dowód. Macierz F wymiaru $n \times k$ ma tylko p liniowo niezależnych kolumn spośród k. Rozważmy najpierw przypadek, kiedy pierwsze p kolumn są liniowo niezależne. Wtedy możemy zapisać

$$F = \begin{pmatrix} F_1 & F_2 \end{pmatrix} = \begin{pmatrix} F_1 & F_1 L \end{pmatrix}$$

dla niektórej $p \times (k-p)$-macierzy L, gdzie macierzy F_1 i F_2 mają odpowiednio wymiary $n \times p$ i $n \times (k-p)$. Mamy

$$F^T W^{-1} F\theta = F^T W^{-1} y \Longleftrightarrow \begin{pmatrix} F_1^T \\ L^T F_1^T \end{pmatrix} W^{-1} \begin{pmatrix} F_1 & F_1 L \end{pmatrix} \theta = \begin{pmatrix} F_1^T \\ L^T F_1^T \end{pmatrix} W^{-1} y$$

$$\iff \begin{pmatrix} F_1^T W^{-1} F_1 & F_1^T W^{-1} F_1 L \\ L^T F_1^T W^{-1} F_1 & L^T F_1^T W^{-1} F_1 L \end{pmatrix} \begin{pmatrix} \theta_{(1)} \\ \theta_{(2)} \end{pmatrix} = \begin{pmatrix} F_1^T W^{-1} y \\ L^T F_1^T W^{-1} y \end{pmatrix},$$

gdzie wektory $\theta_{(1)}$ i $\theta_{(2)}$ są odpowiednio wymiaru p i $k-p$. Współrzędne wektora $\theta_{(2)}$ możemy wybierać dowolnie, a dla nich jednoznacznie odnajdujemy współrzędne wektora $\theta_{(1)}$. Z powyższego układu zostaje nam tylko równanie

$$F_1^T W^{-1} F_1 \theta_{(1)} + F_1^T W^{-1} F_1 L \theta_{(2)} = F_1^T W^{-1} y$$
$$\iff \theta_{(1)} = (F_1^T W^{-1} F_1)^{-1} F_1^T W^{-1} y - L\theta_{(2)}.$$

Więc, dowolne rozwiązanie $\tilde{\theta}$ układu równań normalnych ma postać

$$\tilde{\theta} = \begin{pmatrix} (F_1^T W^{-1} F_1)^{-1} F_1^T W^{-1} y - Lz \\ z \end{pmatrix}, \qquad z \in R^{k-p}.$$

Dla różnych wartości $z \in R^{k-p}$ uzyskujemy różne uogólnione estymatory MNK.

Oczywiście,

$$F\tilde{\theta} = \begin{pmatrix} F_1 & F_1 L \end{pmatrix} \begin{pmatrix} \tilde{\theta}_{(1)} \\ \tilde{\theta}_{(2)} \end{pmatrix} = F_1 \tilde{\theta}_{(1)} + F_1 L \tilde{\theta}_{(2)} = F_1 (F_1^T W^{-1} F_1)^{-1} F_1^T W^{-1} y,$$

a więc $F\tilde{\theta}$ nie zależy od z, tzn. nie zależy od wyboru uogólnionego estymatora MNK.

Jeśli pierwsze p kolumn macierzy F nie są liniowo niezależne, to dokonajmy takiego przestawienia kolumn macierzy F, żeby były one liniowo niezależne. Przestawienie kolumn macierzy F wymaga też odpowiedniego przestawienia współrzędnych wektora θ i estymatora $\tilde{\theta}$. □

Uwaga. Jak łatwo policzyć,

$$E\tilde{\theta} = \begin{pmatrix} \theta_{(1)} + L(\theta_{(2)} - z) \\ z \end{pmatrix}, \quad Cov\ \tilde{\theta} = \sigma^2 \begin{pmatrix} (F_1^T W^{-1} F_1)^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

Więc, estymator $\tilde{\theta}$ jest obciążony, natomiast jego macierz kowariancji nie zależy od z.

Zauważmy, że w przypadku modeli niepełnego rzędu nie istnieje estymatora liniowego nieobciążonego dla wektora θ. Rzeczywiście, jeśli $\breve{\theta} = Ay$ i $\breve{\theta}$ jest nieobciążony, to musi być spełnione $E\breve{\theta} = AF\theta = \theta\ \forall\theta \implies AF = I_k$. Ale ta równość nie może być spełniona, bo $r(F) < k$.

Ale istnieją funkcje liniowe (kombinacje liniowe) nieznanych parametrów θ postaci $\tau = t^T\theta,\ t \in R^k$, dla których nieobciążone estymatory $\tilde{\tau} = b^T y,\ b \in R^n$, istnieją.

Definicja 3.1 *Mówimy, że funkcja $\tau = t^T\theta$ jest estymowalna, jeśli dla niej istnieje estymator liniowy nieobciążony $\tilde{\tau} = b^T y$.*

Oczywiście, $\tau = t^T\theta$ jest estymowalna $\iff$ t jest kombinacjej liniowej wierszy macierzy F, bo $E\tilde{\tau} = \tau \ \forall\theta \iff t^T = b^T F \iff t = F^T b$.

Twierdzenie 3.1 *W modelu $\left(y, F\theta, \sigma^2 W\right)$, $r(F) = p \leq k$, niech $\tilde{\theta}$ jest dowolnym rozwiązaniem układu równań normalnych. Wtedy, jeśli $\tau = t^T\theta$ jest estymowalna, to :*
i) estymator $t^T\tilde{\theta}$ nie zależy od wyboru $\tilde{\theta}$;
ii) $t^T\tilde{\theta}$ jest NLNE dla τ.

Dowód. Niech $\tau = t^T\theta$ jest estymowalna. Wtedy istnieje wektor $b \in R^n$ taki, że $t^T = b^T F$. Więc, $t^T\tilde{\theta} = b^T F\tilde{\theta}$ i zgodnie z Lematem 3.1 mamy i).

Bez straty ogólności będziemy zakładać, że pierwsze p kolumn macierzy F są liniowo niezależne. Wtedy z tego ż Lematu 3.1,

$$t^T\tilde{\theta} = b^T F\tilde{\theta} = b^T F_1(F_1^T W^{-1} F_1)^{-1} F_1^T W^{-1} y \implies$$

$$Et^T\tilde{\theta} = b^T F_1(F_1^T W^{-1} F_1)^{-1} F_1^T W^{-1} F\theta = b^T F\theta = t^T\theta,$$

$$Var\ t^T\tilde{\theta} = t^T Cov\ \tilde{\theta}t = \sigma^2 b^T F_1(F_1^T W^{-1} F_1)^{-1} F_1^T b.$$

Zauważmy, że ponieważ

$$F^T b = t \iff \begin{pmatrix} F_1^T \\ L^T F_1^T \end{pmatrix} b = t \iff \begin{pmatrix} F_1^T b \\ L^T F_1^T b \end{pmatrix} = \begin{pmatrix} t_{(1)} \\ t_{(2)} \end{pmatrix},$$

to $Var\ t^T\tilde{\theta}$ tak naprawde zależy od $t_{(1)}$ i nie zależy od wyboru b który może być niejednoznaczny.

Niech $a^T y$ jest dowolny inny estymator liniowy nieobciążony dla $\tau = t^T\theta$, tzn. wektor a spełnia warunek $t^T = a^T F$. Udowodnimy, że

$$Var\ t^T\tilde{\theta} \leq Var\ a^T y = E(a^T y - t^T\theta)^2 = E(a^T e)^2 = \sigma^2 a^T W a.$$

Poznaczmy

$$l^T = b^T F_1(F_1^T W^{-1} F_1)^{-1} F_1^T W^{-1} - a^T \implies l^T F = 0 \implies l^T F_1 = 0.$$

Mamy

$\sigma^2 a^T W a =$

$$\sigma^2 \left(b^T F_1 (F_1^T W^{-1} F_1)^{-1} F_1^T W^{-1} - l^T\right) W \left(W^{-1} F_1 (F_1^T W^{-1} F_1)^{-1} F_1^T b - l\right) =$$
$$\sigma^2 b^T F_1 (F_1^T W^{-1} F_1)^{-1} F_1^T b - b^T F_1 (F_1^T W^{-1} F_1)^{-1} F_1^T l -$$
$$l^T F_1 (F_1^T W^{-1} F_1)^{-1} F_1^T b + l^T W l = \sigma^2 b^T F_1 (F_1^T W^{-1} F_1)^{-1} F_1^T b + l^T W l$$
$$\geq \sigma^2 b^T F_1 (F_1^T W^{-1} F_1)^{-1} F_1^T b = Var\ t^T \tilde{\theta}. \ \square$$

Sprawdzanie prostych hipotez liniowych o współczynnikach regresji

Rozważmy model $\left(y, F\theta, \sigma^2 W\right)$, $r(F) = p \leq k$, i niech T jest $q \times k$-macierzą, $r(T) = q \leq p$. Zajmiemy się problemem estymacji wektora $\tau = T\theta \in R^q$. Poznaczmy wektory-wiersze macierzy T przez t_i^T, $i = 1, 2, \ldots, q$. Wtedy $\tau_i = t_i^T \theta$, $i = 1, 2, \ldots, q$. Załóżmy, że istnjeje $q \times n$-macierz Q taka, że $T = QF$. W tym przypadku, jak wiemy, q funkcji liniowych $t_i^T \theta$, $i = 1, \ldots, q$, są estymowalne.

Zgodnie z Twierdzeniem 3.1, NLNE dla $\tau = T\theta$ wyraża się wzorem $\tilde{\tau} = T\tilde{\theta}$, gdzie $\tilde{\theta}$ jest dowolnym uogólnionym estymatorem MNK dla wektora θ. Zauważmy, że $\tilde{\tau}$ nie zależy od wyboru $\tilde{\theta}$ i definuje się w sposób jednoznaczny.

Będziemy zakładać, że zmienne losowe $\{e_i\}$ mają rozkłady normalne. Wtedy $\tilde{\tau} \sim N_q \left(T\theta, TCov\ \tilde{\theta} T^T\right)$.

Fakt 3.1 *Jeśli $\xi \sim N_n(\mu, V)$ i macierz V jest dodatnio określona, to $(\xi - \mu)^T V^{-1} (\xi - \mu) \sim \chi_n^2$.*

Załóżmy, że chcemy sprawdzić hipoteze $H_0 \ : \ T\theta = \tau_0$, gdzie $\tau_0 \in R^q$ jest zadane, przeciwko $T\theta \neq \tau_0$. Bez straty ogólności będziemy znowu zakładać, że pierwsze p kolumn macierzy F są liniowo niezależne. Jeśli H_0 ma miejsce, to $\tilde{\tau} \sim N_q \left(\tau_0, \sigma^2 V\right)$, gdzie

$$V = Q \begin{pmatrix} F_1 & F_1 L \end{pmatrix} \begin{pmatrix} (F_1^T W^{-1} F_1)^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} F_1^T \\ L^T F_1^T \end{pmatrix} Q^T$$
$$= Q F_1 (F_1^T W^{-1} F_1)^{-1} F_1^T Q^T = T_1 (F_1^T W^{-1} F_1)^{-1} T_1^T, \qquad (1)$$

gdzie $\begin{pmatrix} T_1 & T_2 \end{pmatrix} = T$. Niech $r(T_1) = q$. Zgodnie z Faktem 3.1,

$$\frac{1}{\sigma^2} (\tilde{\tau} - \tau_0)^T V^{-1} (\tilde{\tau} - \tau_0) \sim \chi_q^2.$$

Więc, jeśli σ^2 jest znane, to statystyka

$$t = \frac{1}{\sigma^2}(\tilde{\tau} - \tau_0)^T V^{-1}(\tilde{\tau} - \tau_0)$$

może być wybrana za testową.

Załóżmy teraz, że σ^2 jest nieznane.

Fakt 3.2 *Jeśli ξ jest zmienną losową o rozkładzie χ^2_n, a η jest zmienną losową o rozkładzie χ^2_m i niezależną od ξ, to zmienna losowa $(m\xi)/(n\eta)$ posiada rozkład o gęstości*

$$f(u) = \frac{\Gamma\left(\frac{n+m}{2}\right)\left(\frac{n}{m}\right)^{n/2}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} u^{(n-2)/2}\left(1 + \frac{nu}{m}\right)^{-(n+m)/2}, \ u > 0,$$

który nazywa się rozkładem Fishera $F(n, m)$ o (n, m) stopniach swobody.

Statystyki $T\tilde{\theta}$ i $\varepsilon^T W^{-1}\varepsilon$ są niezależne (dowód jak w Lemacie 2.4).
Statystyka $\varepsilon^T W^{-1}\varepsilon$ posiada rozkład $\sigma^2\chi^2_{n-p}$ (patrz Wniosek 2.1).
Więc z Faktu 3.2,

$$\frac{1}{q\sigma^2}(\tilde{\tau} - \tau_0)^T V^{-1}(\tilde{\tau} - \tau_0) / \frac{\varepsilon^T W^{-1}\varepsilon}{(n-p)\sigma^2} \sim F(q, n-p),$$

tzn.

$$\frac{n-p}{q\varepsilon^T W^{-1}\varepsilon}(\tilde{\tau} - \tau_0)^T V^{-1}(\tilde{\tau} - \tau_0) \sim F(q, n-p),$$

gdzie V jest ze wzoru (1).

Mamy następującą procedure sprawdzania hipotezy H_0 :

- zadajemy poziom ufności α (jest to prawdopodobieństwo odrzucić hipoteze kiedy jest ona prawdziwa);
- wyliczamy t_α takie, że $P(\zeta \geq t_\alpha) = \alpha, \ \zeta \sim F(q, n-p)$;
- wyliczamy
$$t = \frac{n-p}{q\varepsilon^T W^{-1}\varepsilon}(\tilde{\tau} - \tau_0)^T V^{-1}(\tilde{\tau} - \tau_0);$$
- jeśli $t \in (0, t_\alpha)$, to nie mamy podstaw odrzucić hipoteze H_0, w przeciwnym przypadku hipoteze H_0 odrzucamy.

Uwaga. W przypadku $q = 1$ mamy hipoteze $t^T\theta = a$ przeciwko $t^T\theta \neq a$. Będzie to uogólnienie przypadku rozpatrywanego w Rozdziale 2. Zauważmy też, że $F(1, n-p) = t^2_{n-p}$.

4 Planowanie doświadczeń. Plany dyskretne i ciągłe. Macierz informacji i macierz kowariancji. Własności macierzy informacji.

Co to jest planowanie doświadczeń

Niech mamy model klasyczny regresji liniowej $\left(y, F\theta, \sigma^2 I_n\right)$, tzn. mamy obserwacje

$$y_i = \sum_{j=1}^{k} \theta_j f_j(x_i) + e_i,\ i = 1, 2, \ldots, n,\ n \geq k,$$

gdzie $\theta = (\theta_1, \ldots, \theta_k)^T$ jest wektorem nieznanych współczynników regresji, $e = (e_1, \ldots, e_n)^T$ jest takim wektorem losowym, że $Ee = 0$, $Cov\ e = \sigma^2 I_n$, $\sigma > 0$ jest nieznane, $\{x_i\}_{i=1}^n \subset X \subset R^m$ (zbiór X nazywa się *zbiorem planowania*), $f_j : R^m \longrightarrow R$ są znane funkcje mierzalne. Dalej będziemy zakładać, że funkcje te są ciągłe, a zbiór X jest zwartym, tzn. ograniczonym i domkniętym.

Jasne, że estymatory dla θ (np. NLNE $\widehat{\theta}$) zależą od macierzy F, tzn. od punktów $\{x_i\}_{i=1}^n \subset X$. Będziemy mówić o zbiorze tych punktów (one mogą się powtarzać) jako o *planie doświadczeń*. W poprzednich rozdzialach konstruowaliśmy estymatory dla θ przy założeniu, że macierz F jest zadana. *Zadanie optymalnego planowania doświadczeń, związane z estymacjej wektora θ, jest zadaniem wyboru punktów $\{x_i\}_{i=1}^n$ (planu) ze zbioru X w taki sposób, aby estymator $\widehat{\theta}$ byl najlepszym w pewnym sensie.*

Plany dyskretne i ciągłe

Definicja 4.1 *Planem dyskretnym ścisłym nazywamy dyskretną miare probabilistyczna postaci*

$$\xi = \xi^{(n)} = \left\{ \begin{array}{ccc} x_1 & \ldots & x_n \\ 1/n & \ldots & 1/n \end{array} \right\}. \tag{2}$$

Tzn. w punktach $\{x_i\}$ przyjmuje ona wartość $1/n$. Tutaj supp $\xi = \{x_1, \ldots, x_n\}$ nazywa się nośnikiem planu, natomiast wartości $(1/n, \ldots, 1/n)$ – odpowiednimi wagami.

Odpowiada ten plan takiemu przeprowadzeniu n doświadczeń, że w doświadczeniu z numerem i przeprowadzamy pomiar w punkcie x_i.

W tych przypadkach, kiedy spośród punktów plana są powtarzające się, wygodniej jest zapisywać plan dyskretny ścisły w postaci

$$\xi = \xi^{(n)} = \left\{ \begin{array}{ccc} x_1 & \dots & x_r \\ p_1 & \dots & p_r \end{array} \right\}, \quad r \leq n, \tag{3}$$

$$p_i \geq 0, \ \sum_{i=1}^{r} p_i = 1, \ p_i = \frac{n_i}{n}, \ i = 1, 2, \dots, r, \tag{4}$$

gdzie n_i – liczba powtórzeń punktu x_i w planie (2). Zbiór wszystkich planów dyskretnych ścisłych postaci (3), spełniających warunki (4), będziemy poznaczać przez Ω_n.

Jak zobaczymy dalej, rozwiązać zadanie optymalnego planowania doświadczeń dla zbioru planów Ω_n rzadko się udaje. Jednak, jeśli powiększyć zbiór planów dopuszczalnych, to udaje się otrzymać dobre wyniki. Mianowicie, załóżmy, że liczba punktów nośnika r jest ograniczona, natomiast ilość doświadczeń n może być dowolnie duża. Wtedy przychodzimy do następnego pojęcia.

Definicja 4.2 *Planem dyskretnym uogólnionym nazywamy plan postaci (3), dla którego warunki (4) są nieco osłabione, tj.*

$$p_i \geq 0, \ \sum_{i=1}^{r} p_i = 1, \ i = 1, 2, \dots, r.$$

Innymi slowy, planem dyskretnym uogólnionym nazywamy dyskretną miare probabilistyczna postaci

$$\xi = \left\{ \begin{array}{ccc} x_1 & \dots & x_r \\ p_1 & \dots & p_r \end{array} \right\}, \ p_i \geq 0, \ \sum_{i=1}^{r} p_i = 1, \ i = 1, 2, \dots, r. \tag{5}$$

Zauważmy, że jeśli plany postaci (2) lub (3) były ścisłe związane z ilościa obserwacji n, to plan postaci (5) nie jest takim; tutaj można uważać, że liczba obserwacji jest nieskończona.

Zbiór wszystkich planów postaci (5) będziemy poznaczać przez Ω_0. Na zbiorze Ω_0 można określić kombinacje wypuklą dwoch planow. Rzeczywiście, niech

$$\xi' = \left\{ \begin{array}{ccc} x'_1 & \dots & x'_{r_1} \\ p'_1 & \dots & p'_{r_1} \end{array} \right\}, \quad \xi'' = \left\{ \begin{array}{ccc} x''_1 & \dots & x''_{r_2} \\ p''_1 & \dots & p''_{r_2} \end{array} \right\}$$

są dwa plany z Ω_0. Dla dowolnego $\alpha \in [0,1]$ zdefiniujmy plan $(1-\alpha)\xi' + \alpha\xi''$ jako plan postaci (5), dla którego:
i) punkty nośnika są sumą mnogościową zbiorów $\{x'_1, \ldots, x'_{r_1}\}$ i $\{x''_1, \ldots, x''_{r_2}\}$; będziemy ich poznaczać przez $x_1, \ldots, x_r,\ r \leq r_1 + r_2$;
ii) plany ξ' i ξ'' poszerzamy tak oba by miały nośniki $x_1, \ldots, x_r$, a punktom nośnika, które w danym planie pojawiły się na skutek poszerzania, przypisujemy wagi zerowe;
iii) wagi wynikowe $\{p_i\}$ powstają zgodnie ze wzorem $p_i = (1-\alpha)p'_i + \alpha p''_i,\ i = 1, \ldots, r$.

Przykład. Niech $\alpha = 1/2$, a

$$\xi' = \left\{ \begin{array}{cc} -1 & 1 \\ 1/2 & 1/2 \end{array} \right\}, \quad \xi'' = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{array} \right\}.$$

Wtedy $(1-\alpha)\xi' + \alpha\xi''$ ma postać

$$\left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 5/12 & 1/6 & 5/12 \end{array} \right\}.$$

Ponieważ kombinacja wypukła dwoch planów z Ω_0 też jest planem z Ω_0, zbiór ten jest *wypukłym.*

Definicja 4.3 *Planem ciągłym nazywamy dowolną miare probabilistyczna na przestrzeni mierzalnej $(X, \mathcal{B})$. Zbiór wszystkich planów ciągłych będziemy poznaczać przez Ω.*

Macierz informacji i macierz kowariancji

Jak wiemy, dla NLNE $\hat{\theta}$ macierz kowariancji ma postać

$$Cov\ \hat{\theta} = \sigma^2 (F^T F)^{-1}.$$

Macierzą informacji nazywamy macierz M postaci $M = F^T F$. Nazywa się ona tak dla tego, że w przypadku, kiedy rozkłady zmiennych losowych $\{e_i\}$ są normalne, jest to macierz informacji Fishera o wektorze θ zawartą w obserwacjach $\{y_i\}$.

Poznaczmy $f(x_i) = (f_1(x_i), \ldots, f_k(x_i))^T$. Wtedy dla planu postaci (2)

$$M = \sum_{i=1}^{n} f(x_i) f^T(x_i),$$

a dla planu postaci (3)

$$M = \sum_{i=1}^{r} n_i f(x_i) f^T(x_i) = n \sum_{i=1}^{r} p_i f(x_i) f^T(x_i).$$

Dalej wygodniej będzie pracować z unormowaną macierzą informacji. Dla wskazanych planów postaci (2) i (3) będzie to odpowiednio

$$M(\xi^{(n)}) = \frac{1}{n} \sum_{i=1}^{n} f(x_i) f^T(x_i), \quad M(\xi^{(n)}) = \sum_{i=1}^{r} p_i f(x_i) f^T(x_i).$$

Zauważmy, że

$$Cov\ \hat{\theta}(\xi^{(n)}) = \frac{\sigma^2}{n} M^{-1}(\xi^{(n)}).$$

Analogicznie, dla planu dyskretnego uogólnionego postaci (5) zdefiniujmy

$$M(\xi) = \sum_{i=1}^{r} p_i f(x_i) f^T(x_i),$$

a dla planu ciągłego

$$M(\xi) = \int_X f(x) f^T(x) \xi(dx).$$

Dalej będziemy opuszczać slowo "unormowana" o ile nie doprowadzi to do nieporozumień i nazywać macierzą informacji macierz unormowaną.

Własności macierzy informacji

1. Macierz informacji $M(\xi)$ jest symetryczna i nieujemnie określona (bo np. dla $\xi \in \Omega$ mamy $z^T M(\xi) z = \int_X (z^T f(x))^2 \xi(dx) \geq 0$).

2. Jeśli *supp* ξ składa się z $r < k$ punktów, to $\xi \in \Omega_0$ jest planem osobliwym, tzn. takim że $M(\xi)$ jest macierzą osobliwą.

3. Zbiór macierzy informacji jest wypukły, tzn. macierz

   $$M = (1-\alpha) M(\xi') + \alpha M(\xi''), \quad \forall \alpha \in [0,1], \quad \forall \xi', \xi'' \in \Omega(\Omega_0),$$

   też jest macierzą informacji. Wynika to z tego faktu, że $\xi = (1-\alpha)\xi' + \alpha\xi'' \in \Omega(\Omega_0)$ oraz z tego, że M jest macierzą informacji dla planu ξ.

4.* Jeśli X jest zbiorem zwartym, a funkcje $f_j(x)$ są ciągłe, to zbiór macierzy informacji jest zbiorem zwartym.

5.* Dla dowolnego planu $\xi \in \Omega$ istnieje plan $\xi' \in \Omega_0$ taki, że $M(\xi) = M(\xi')$, przy czym *supp* ξ' składa się z nie więcej niż z $k(k+1)/2 + 1$ punktów.

5 Kryteria optymalności. Przykłady.

Kryteria optymalności

Rozważmy zbiór planów Ω_0 oraz niech $\mathcal{M} = \{M(\xi),\ \xi \in \Omega_0\}$ będzie zbiorem osiągalnych macierzy informacji.

Powstaje pytanie: *jak porównywać plany doświadczeń pomiędzy sobą* ? Tzn. jak roztrzygnąc, który z dwoch planów ξ', ξ'' jest lepszy ? Wprowadżmy pojęcie *kryterium optymalności*. Jest to funkcja zadana na zbiorze $\mathcal{M}$ (lub na zbiorze Ω_0) i służacą dla porównania planów doświadczeń.

Najpierw, sprobujmy porównywać plany w sposób następujący. Jasne, że jakość planu może być wyrażona poprzez macierz kowariancji lub macierz informacji. Przyjmijmy, że plan ξ_2 jest lepszym od plana ξ_1, jeśli $Cov\ \widehat{\theta}(\xi_2) \leq Cov\ \widehat{\theta}(\xi_1)$ czy $M(\xi_2) \geq M(\xi_1)$.

Przykład 1. Niech $m = 1,\ k = 2,\ f_1(x) = 1,\ f_2(x) = x,\ X = [-1, 1]$. Wieżmiemy

$$\xi_1 = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{array} \right\}, \quad \xi_2 = \left\{ \begin{array}{cc} -1 & 1 \\ 1/2 & 1/2 \end{array} \right\}.$$

Mamy

$$M(\xi_1) = \frac{1}{3}\sum_{i=1}^{3} \left(\begin{array}{c} 1 \\ x_i \end{array} \right) \left(\begin{array}{cc} 1 & x_i \end{array} \right) = \frac{1}{3}\sum_{i=1}^{3} \left(\begin{array}{cc} 1 & x_i \\ x_i & x_i^2 \end{array} \right)$$

$$= \left(\begin{array}{cc} 1 & \frac{1}{3}(x_1 + x_2 + x_3) \\ \frac{1}{3}(x_1 + x_2 + x_3) & \frac{1}{3}(x_1^2 + x_2^2 + x_3^2) \end{array} \right) = \left(\begin{array}{cc} 1 & 0 \\ 0 & 2/3 \end{array} \right),$$

$$M(\xi_2) = \frac{1}{2}\sum_{i=1}^{2} \left(\begin{array}{c} 1 \\ x_i \end{array} \right) \left(\begin{array}{cc} 1 & x_i \end{array} \right) = \left(\begin{array}{cc} 1 & \frac{1}{2}(x_1 + x_2) \\ \frac{1}{2}(x_1 + x_2) & \frac{1}{2}(x_1^2 + x_2^2) \end{array} \right) = \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right).$$

Ostatecznie, ξ_2 jest lepszy od ξ_1 bo $M(\xi_2) \geq M(\xi_1)$.

Przykład 2. Niech $m = 1,\ k = 3,\ f_1(x) = 1,\ f_2(x) = x,\ f_3(x) = x^2,\ X = [-1, 1]$. Wieżmiemy

$$\xi_1 = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{array} \right\}, \quad \xi_2 = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 1/4 & 1/2 & 1/4 \end{array} \right\}.$$

Mamy

$$M(\xi_1) = \frac{1}{3}\sum_{i=1}^{3} \left(\begin{array}{ccc} 1 & x_i & x_i^2 \\ x_i & x_i^2 & x_i^3 \\ x_i^2 & x_i^3 & x_i^4 \end{array} \right) = \left(\begin{array}{ccc} 1 & 0 & 2/3 \\ 0 & 2/3 & 0 \\ 2/3 & 0 & 2/3 \end{array} \right),$$

$$M(\xi_2) = \frac{1}{3}\sum_{i=1}^{3}\begin{pmatrix} 1 & x_i & x_i^2 \\ x_i & x_i^2 & x_i^3 \\ x_i^2 & x_i^3 & x_i^4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}.$$

Ostatecznie nie można powiedzeć, który plan jest lepszy, bo ani

$$M(\xi_1) - M(\xi_2) = \begin{pmatrix} 0 & 0 & 1/6 \\ 0 & 1/6 & 0 \\ 1/6 & 0 & 1/6 \end{pmatrix},$$

ani

$$M(\xi_2) - M(\xi_1) = \begin{pmatrix} 0 & 0 & -1/6 \\ 0 & -1/6 & 0 \\ -1/6 & 0 & -1/6 \end{pmatrix}$$

nie są nieujemnie określone !

Więc, próba porównywania ze sobą macierzy informacji "w całości" nie bardzo się powiodła, bo nie mogły my porównać dwoch dowolnych planów za pomocą rozważanego kryterium. Dla tego najczęściej posługują się skalarnymi funkcjami macierzy informacji, wybranymi tak, by miały one interpretację statystyczną. W zależności od potrzeb, rozważają różne kryteria optymalności.

Niech $\Psi \ : \ \mathcal{M} \to R^+$.

Definicja 5.1 *Plan ξ^* nazywamy Ψ-optymalnym, jeśli*

$$\xi^* = Arg \ \min_{\xi} \ \Psi(M(\xi)).$$

Funkcję Ψ nazywamy kryteriem optymalności.

Podstawowe własności kryteriów optymalności.

1. Dla planu ξ osobliwego mamy $\Psi(M(\xi)) = \infty$.

2. Monotoniczność : dla wszystkich planów ξ_1, ξ_2 mamy $M(\xi_1) \geq M(\xi_2) \implies \Psi(M(\xi_1)) \leq \Psi(M(\xi_2))$.
   Oznacza to, że im więcej informacji o wektorze θ "niesie w sobie" (posiada) macierz $M(\xi)$, tym jest lepiej.

3. Jednorodność : $\Psi(\beta M(\xi)) = \gamma(\beta)\Psi(M(\xi))$ dla wszystkich $\beta > 0$ i wszystkich planów ξ, gdzie γ jest funkcjej nierosnącej.
   Uwaga. Jednorodność kryterium optymalności pozwala zdefiniować jego również w terminach macierzy kowariancji estymatora $\widehat{\theta}$.

4. Wypukłość : $\forall \alpha \in [0,1], \quad \forall \xi_1, \xi_2 \in \Omega,$

$$\Psi((1-\alpha)M(\xi_1) + \alpha M(\xi_2)) \leq (1-\alpha)\Psi(M(\xi_1)) + \alpha\Psi(M(\xi_2)).$$

Oznacza to, że lepiej mieć do czynienia z planem $(1-\alpha)\xi_1 + \alpha\xi_2$ (niesie on w sobie więcej informacji) niż interpolując plany ξ_1 i ξ_2.
Uwaga. Większość kryteriów optymalności jest nie prosto wypukłe, a ostro wypukłe, tzn. $\forall \alpha \in (0,1), \quad \forall \xi_1, \xi_2 \in \Omega$ z $M(\xi_1) \neq M(\xi_2)$ mamy ostrą nierówność w definicji wypukłości.

Lemat 5.1 *Jeśli funkcja Ψ jest wypukła, to dowolna wypukła kombinacja planów Ψ-optymalnych też jest planem Ψ-optymalnym.*

Jeśli funkcja Ψ jest ostro wypukła, to wszystkie plany Ψ-optymalne mają jednakową macierz informacji.

Dowód. 1. Niech ξ_1^*, ξ_2^* są planami Ψ-optymalnymi i rozważmy plan $\xi^* = (1-\alpha)\xi_1^* + \alpha\xi_2^*, \ \alpha \in [0,1]$. Wtedy z wypukłości funkcji Ψ mamy

$$\Psi(M(\xi^*)) \leq (1-\alpha)\Psi(M(\xi_1^*)) + \alpha\Psi(M(\xi_2^*)) = \min_{\xi} \ \Psi(M(\xi)) \implies$$

$$\Psi(M(\xi^*)) = \min_{\xi} \ \Psi(M(\xi)).$$

2. Niech ξ_1^*, ξ_2^* są planami Ψ-optymalnymi, a odpowiednie macierze informacji nie są równe, tzn. $M(\xi_1^*) \neq M(\xi_2^*)$. Poznaczmy $\xi^* = (1-\alpha)\xi_1^* + \alpha\xi_2^*, \ \alpha \in [0,1]$. Z ostrej wypukłości Ψ mamy

$$\Psi(M(\xi^*)) < (1-\alpha)\Psi(M(\xi_1^*)) + \alpha\Psi(M(\xi_2^*)) = \min_{\xi} \ \Psi(M(\xi))$$

co jest sprzecznością.

Fakt 5.1 *Jeśli kryterium Ψ jest jednorodny i Ψ-optymalny ciągły plan ξ^* jest nieosobliwy, to istnieje plan $\xi \in \Omega_0$ taki, że $M(\xi^*) = M(\xi)$ i supp ξ składa się z nie więcej niż $k(k+1)/2$ punktów.*

Przykłady najczęściej stosowanych kryteriów optymalności

- $\Psi(M) = det \ M^{-1}$ ($\Psi(M) = \infty$ dla M osobliwej) – **kryterium D-optymalności**. Plan ξ^* nazywamy D-optymalnym, jeśli

$$\xi^* = Arg \ \min_{\xi} \ det \ M^{-1}(\xi) = Arg \ \max_{\xi} \ det \ M(\xi).$$

Ponieważ

$$Arg \min_{\xi^{(n)}} det\ M^{-1}(\xi^{(n)}) = Arg \min_{\xi^{(n)}} det\ Cov\ \widehat{\theta}(\xi^{(n)}),$$

minimizujemy tym samym (w pewnym sensie) kowariancję dla NLNE $\widehat{\theta}$. Można również pokazać, że minimizuje się przy tym objętość elipsoidy ufności dla $\widehat{\theta}$.

Sprawdzimy wskazane wyżej własności. Użyjemy dalej znany fakt z Teorii macierzy.

Fakt 5.2 *Niech A i B są $k \times k$-macierzami symetrycznymi oraz B jest macierzą dodatnio okreslona. Wtedy istnieje nieosobliwa macierz R i diagonalna macierz Λ takie, że $A = R^T \Lambda R,\ B = R^T R$.*

1. Oczywiste.

2. Monotoniczność. Wystarzy pokazać, że

$$M_1 \geq M_2 \Longrightarrow det\ M_1 \geq det\ M_2.$$

Jeśli przynajmniej jedna z macierzy M_1, M_2 jest osobliwa, to nierówność jest oczywista. Niech teraz obie macierzy są nieosobliwe, a więc dodatnio określone. Z Faktu 5.2 $M_1 = R^T \Lambda R,\ M_2 = R^T R$, i jeśli $M_1 \geq M_2$, to elementy $\{\lambda_i\}$ stojące na przekątnej macierzy Λ są wszystkie ≥ 1. (**Zadanie domowe.** Udowodnić to.) Więc,

$$det\ M_1 = (det\ R)^2 \prod_{i=1}^{k} \lambda_i \geq (det\ R)^2 = det\ M_2.$$

3. Jednorodność : $\Psi(\beta M) = \beta^{-k} \Psi(M)$.

4. Ostra wypukłość. Udowodnimy, że funkcja $-\ln det\ M$ jest ostro wypukła, z czego będzie wynikało, że i funkcja $\Psi(M) = det\ M^{-1} = e^{-\ln det\ M}$ jest ostro wypukła jako złożenie dwoch funkcji ostro wypukłych. Ostra wypukłość dla $-\ln det\ M$ równoważna jest nierówności

$$det\ ((1-\alpha)M_1 + \alpha M_2) > (det\ M_1)^{1-\alpha} (det\ M_2)^{\alpha},\ \forall \alpha \in (0,1),\ M_1 \neq M_2,$$

lub ze względu na Fakt 5.2,

$$\prod_{i=1}^{k} (\alpha \lambda_i + 1 - \alpha) > \prod_{i=1}^{k} \lambda_i^{\alpha},\ \forall \alpha \in (0,1),\ M_1 \neq M_2,$$

lub

$$\sum_{i=1}^{k} \ln(\alpha\lambda_i + 1 - \alpha) > \alpha \sum_{i=1}^{k} \ln \lambda_i, \ \forall \alpha \in (0,1), \ M_1 \neq M_2.$$

Ostatnia nierówność jest prawdziwa ze względu na ostrą wklęsłość funkcji $\ln x$.

Łatwo sprawdzić, że ma miejsce także następująca własność tego kryterium : *Jeśli ξ^* jest planem D-optymalnym przy estymacji θ, to ten sam plan jest również D-optymalnym przy estymacji $T\theta$, gdzie macierz T jest dowolną nieosobliwą $k \times k$-macierzą.*

- $\Psi(M) = tr\ M^{-1}$ ($\Psi(M) = \infty$ dla M osobliwej) – **kryterium A-optymalności**. Plan ξ^* nazywamy A-optymalnym, jeśli

$$\xi^* = Arg \min_{\xi}\ tr\ M^{-1}(\xi).$$

Ponieważ

$$Arg \min_{\xi^{(n)}}\ tr\ M^{-1}(\xi^{(n)}) = Arg \min_{\xi^{(n)}}\ tr\ Cov\ \hat{\theta}(\xi^{(n)}),$$

minimizujemy tutaj sumarną wariancje estymatorów wspólrzędnych wektora θ. Można również pokazać, że minimizuje się przy tym suma osi elipsoidy ufności dla $\hat{\theta}$.

Sprawdzimy wskazane wyżej własności.

1. Oczywiste.

2. Monotoniczność. Mamy

$$M_1 \geq M_2 \Longrightarrow M_1^{-1} \leq M_2^{-1}.$$

Z Faktu 5.2 wynika, że $M_1 = R^T \Lambda R$, $M_2 = R^T R$, i elementy $\{\lambda_i\}$ macierzy Λ są wszystkie ≥ 1. Teraz

$$tr\ M_2^{-1} - tr\ M_1^{-1} = tr\ R^{-1}(I_k - \Lambda^{-1})(R^T)^{-1} \geq 0,$$

bo macierz która stoi pod tr jest nieujemnie określoną.

3. Jednorodność.

$$\Psi(\beta M) = \beta^{-1}\Psi(M).$$

4. Ostra wypukłość. Oznacza ona, że : $\forall \alpha \in [0,1]$, $\forall M_1, M_2 \in \mathcal{M}$, $M_1 \neq M_2$,

$$tr\ (\alpha M_1 + (1-\alpha)M_2)^{-1} < \alpha tr\ M_1^{-1} + (1-\alpha) tr\ M_2^{-1},$$

lub ze względu na Fakt 5.2,

$$tr\ R^{-1}(\alpha\lambda_i + 1 - \alpha)^{-1}(R^T)^{-1} < tr\ R^{-1}\left(\frac{\alpha}{\lambda_i} + 1 - \alpha\right)(R^T)^{-1}.$$

Ostatnia nierówność jest prawdziwa ze względu na ostrą wypukłość funkcji x^{-1} dla $x > 0$.

- $\Psi(M) = tr\ LM^{-1}$ ($\Psi(M) = \infty$ dla M osobliwej) – **kryterium L-optymalności**, gdzie L jest zadaną symetryczną dodatnio określoną macierzą. Plan ξ^* nazywamy L-optymalnym, jeśli

$$\xi^* = Arg\ \min_{\xi}\ tr\ LM^{-1}(\xi).$$

Minimizujemy tu błąd średniokwadratowy postaci $E(\hat{\theta}-\theta)^T L(\hat{\theta}-\theta)$ dla NLNE $\hat{\theta}$. Można również pokazać, że minimizuje się przy tym suma osi odpowiednio liniowo przekształconej elipsoidy ufności dla $\hat{\theta}$.

- $\Psi(M) = \lambda_{\max}(M^{-1}) = \frac{1}{\lambda_{\min}(M)}$ ($\Psi(M) = \infty$ dla M osobliwej) – **kryterium E-optymalności**. Plan ξ^* nazywamy E-optymalnym, jeśli

$$\xi^* = Arg\ \min_{\xi}\ \lambda_{\max}(M^{-1}(\xi)) = Arg\ \max_{\xi}\ \lambda_{\min}(M(\xi)).$$

Minimizuje się tutaj długość maksymalnej osi elipsoidy ufności dla NLNE $\hat{\theta}$.

- $\Psi(M) = c^T M^{-1} c$, gdzie $c \in R^k$, $c \neq 0$ jest zadane ($\Psi(M) = \infty$ dla M osobliwej) – **kryterium C-optymalności**. Plan ξ^* nazywamy C-optymalnym, jeśli

$$\xi^* = Arg\ \min_{\xi}\ c^T M^{-1}(\xi) c.$$

Minimizuje się tutaj wariancja NLNE parametru $c^T\theta$.

Niech $x \in X$ jest zadanym punktem a $d(x,\xi) = f^T(x)M^{-1}(\xi)f(x)$. Zauważmy, że $\frac{\sigma^2}{n}d(x,\xi)$ jest wariancjej NLNE $\hat{\theta}^T f(x)$ dla funkcji regresji $\theta^T f(x)$. Minimizacja $Arg\ \min_{\xi}\ d(x,\xi)$ jest szczególnym przypadkiem kryterium C-optymalności z $c = f(x)$.

- $\Psi(M) = \max_{x\in X}\ d(x,\xi)$ – **kryterium G-optymalności**. Plan ξ^* nazywamy G-optymalnym, jeśli

$$\xi^* = Arg\ \min_{\xi}\ \max_{x\in X}\ d(x,\xi).$$

Minimizuje się tu maksymalna po $x \in X$ wariancja NLNE funkcji regresji.

6 Twierdzenia równoważności, przykłady ich używania.

Twierdzenia równoważności

Zdefinujmy dwa pojęcia pochodnej związane z macierzami. Wszędzie uważamy, że macierz M jest macierzą symetryczną.

Niech $\Psi = \Psi(M)$.

Definicja 6.1 *Jeśli istnieją ciągłe pochodne cząstkowe $\partial\Psi(M)/\partial M_{ij}$ po elementam macierzy M, to mówią, że funkcja $\Psi(M)$ jest różniczkowalna względem macierzy M i jej pochodnej po macierzy M nazywają macierz złożona z pochodnych cząstkowych, tzn.*

$$\Psi'(M) = \frac{\partial\Psi(M)}{\partial M} = \begin{pmatrix} \frac{\partial\Psi(M)}{\partial M_{11}} & \cdots\cdots & \frac{\partial\Psi(M)}{\partial M_{1k}} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \frac{\partial\Psi(M)}{\partial M_{k1}} & \cdots\cdots & \frac{\partial\Psi(M)}{\partial M_{kk}} \end{pmatrix}.$$

Niech elementy macierzy M zależą od pewnego parametra t.

Definicja 6.2 *Pochodnej od macierzy M po parametru t nazywamy macierz postaci*

$$\frac{\partial M}{\partial t} = \begin{pmatrix} \frac{\partial M_{11}}{\partial t} & \cdots\cdots & \frac{\partial M_{1k}}{\partial t} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \frac{\partial M_{k1}}{\partial t} & \cdots\cdots & \frac{\partial M_{kk}}{\partial t} \end{pmatrix}.$$

Przykłady.

1. $\Psi(M) = tr\ M = \sum_{i=1}^{k} M_{ii} \quad \Longrightarrow \quad \Psi'(M) = I_k.$
2. $\Psi(M) = det\ M \quad \Longrightarrow \quad \Psi'(M) = (det\ M)M^{-1}.$
   Rzeczywiście, wyznacznik macierzy możemy rozłożyć po elementach i-go wiersza
   $$det\ M = \sum_{j=1}^{k} (-1)^{i+j} M_{ij} A_{ij},$$
   gdzie A_{ij} jest dopełnieniem algebraicznym do elementu M_{ij} (nie zależy od elementu M_{ij}). Wtedy
   $$\frac{\partial det\ M}{\partial M_{ij}} = (-1)^{i+j} A_{ij},$$
   skąd mamy zapowiadaną formulę.

3. Policzmy $\partial M^{-1}/\partial M_{ij}$. Mamy

$$MM^{-1} = I_k \implies \frac{\partial M}{\partial t}M^{-1} + M\frac{\partial M^{-1}}{\partial t} = 0 \implies \frac{\partial M^{-1}}{\partial t} = -M^{-1}\frac{\partial M}{\partial t}M^{-1}.$$

Stąd,

$$\frac{\partial M^{-1}}{\partial M_{ij}} = -M^{-1}\frac{\partial M}{\partial M_{ij}}M^{-1} = -M^{-1}E^{ij}M^{-1},$$

gdzie E^{ij} jest $k \times k$-macierzą, która na miejscu (i,j) ma jedynkę, a pozostale elementy są zera. Więc,

$$\frac{\partial M^{-1}}{\partial M_{ij}} = -\|(M^{-1})_{li}(M^{-1})_{js}\|_{l,s}.$$

4. $\Psi(M) = det\ M^{-1} \quad \implies \quad \Psi'(M) = -(det\ M^{-1})M^{-1}$.
Korzystając z drugiego przyklądu mamy

$$\Psi'(M) = ((det\ M)^{-1})' = -(det\ M)^{-2}(det\ M)M^{-1},$$

skąd wynika zapowiadana formula.

5. $\Psi(M) = tr\ M^{-1} \quad \implies \quad \Psi'(M) = -M^{-2}$.
Korzystając z trzeciego przyklądu mamy

$$\frac{\partial \Psi(M)}{\partial M_{ij}} = \sum_{l=1}^{k}\frac{\partial (M^{-1})_{ll}}{\partial M_{ij}} = -\sum_{l=1}^{k}(M^{-1})_{li}(M^{-1})_{jl} = -(M^{-2})_{ji}.$$

Twierdzenie 6.1 *Niech $\Psi(M)$ będzie wypukłą różniczkowalną funkcją (względem macierzy M), a zbiór osiągalnych macierzy informacji $\mathcal{M} = \{M(\xi); \xi \in \Omega\}$ jest zwartym. Następne stwierdzenia są równoważne:*

1. $\Psi(M^*) = \min\limits_{M\in\mathcal{M}} \Psi(M), \quad M^* = M(\xi^*)$;
2. $\inf\limits_{x\in X} f^T(x)\Psi'(M^*)f(x) = tr\ (M^*\Psi'(M^*))$;
3. *supp ξ^* jest skoncentrowany w punktach $z \in X$ dla których*

$$f^T(z)\Psi'(M^*)f(z) = \inf_{x\in X} f^T(x)\Psi'(M^*)f(x).$$

Dowód. Zauważmy, że dla dowolnego planu ciągłego ma miejsce tożsamość

$$\int_X f^T(x)\Psi'(M(\xi))f(x)\xi(dx) = tr\ \left(\int_X f(x)f^T(x)\xi(dx)\Psi'(M(\xi))\right)$$
$$= tr\ (M(\xi)\Psi'(M(\xi))). \qquad (6)$$

Z tej tożsamości wynika równoważność stwierdzeń 2 i 3. Rzeczywiście, ze stwierdzenia 2 i (6) wynika, że

$$\int_X f^T(x)\Psi'(M(\xi^*))f(x)\xi^*(dx) = \inf_{x\in X}\ f^T(x)\Psi'(M(\xi^*))f(x),$$

co powoduje 3. Odwrotne ($3 \implies 2$) też jest oczywiste ze względu na (6).

Udowodnimy teraz, że $1 \implies 2$.

Lemat 6.1 *W warunkach Twierdzenia 6.1, warunkiem koniecznym i dostatecznym na to, aby macierz $M^* \in \mathcal{M}$ dostarczala minimum dla $\Psi(M)$, jest spelnienie nierówności*

$$\left.\frac{\partial\Psi(M^*+\alpha(M-M^*))}{\partial\alpha}\right|_{\alpha=0+} = tr\ (\Psi'(M^*)(M-M^*)) \geq 0 \qquad \forall M \in \mathcal{M}.$$

Dowód Lematu 6.1. Po-pierwsze, rzeczywiście

$$\left.\frac{\partial\Psi(M^*+\alpha(M-M^*))}{\partial\alpha}\right|_{\alpha=0+} = \sum_{i,j=1}^{n} \left.\frac{\partial\Psi(M)}{\partial M_{ij}}\right|_{M=M^*} (M_{ij}-M^*_{ij})$$
$$= tr\ (\Psi'(M^*)(M-M^*)) \qquad \forall M \in \mathcal{M}.$$

Po-drugie, funkcja

$$g(\alpha) = \frac{\Psi(M^*+\alpha(M-M^*)) - \Psi(M^*)}{\alpha} \qquad \forall M \in \mathcal{M}$$

jest niemalejącej funkcjej od $\alpha > 0$. Rzeczywiście, niech $\alpha_2 > \alpha_1 > 0$. Z wypukłości Ψ mamy

$$\Psi(M^*+\alpha_1(M-M^*)) = \Psi\left(\frac{\alpha_1}{\alpha_2}(M^*+\alpha_2(M-M^*)) + \left(1-\frac{\alpha_1}{\alpha_2}\right)M^*\right) \leq$$
$$\frac{\alpha_1}{\alpha_2}\Psi(M^*+\alpha_2(M-M^*)) + \left(1-\frac{\alpha_1}{\alpha_2}\right)\Psi(M^*),$$

skąd

$$\frac{\Psi(M^*+\alpha_1(M-M^*)) - \Psi(M^*)}{\alpha_1} \leq \frac{\Psi(M^*+\alpha_2(M-M^*)) - \Psi(M^*)}{\alpha_2}.$$

($\Longrightarrow$) Jeśli M^* dostarcza minimum funkcji Ψ, to funkcja $g(\alpha),\ \alpha > 0$, jest dodatnia. Teze Lematu 6.1 otrzymujemy przechodząc do granicy $\alpha \to 0+$.

($\Longleftarrow$) Wiemy z założenia, że

$$\lim_{\alpha \to 0+}\ g(\alpha) \geq 0 \qquad \forall M \in \mathcal{M}$$

i że funkcja $g(\alpha)$ jest niemalejąca dla $\alpha > 0$. Stąd wynika, że $g(\alpha) \geq 0\ \forall \alpha > 0$ i $\forall M \in \mathcal{M}$, co oznacza, że w M^* mamy minimum lokalne funkcji $\Psi(M)$. A jak wiadomo, dla funkcji wypukłych wszystkie minima lokalne są minimami globalnymi. $\square$

Wróćmy do dowodu Twierdzenia 6.1. Z Lematu 6.1

$$\inf_{\xi \in \Omega}\ \int_X f^T(x)\Psi'(M(\xi^*))f(x)\xi(dx) \geq tr\ (M(\xi^*)\Psi'(M(\xi^*))). \tag{7}$$

Uwzględniając (6) i równość

$$\inf_{\xi \in \Omega}\ \int_X f^T(x)\Psi'(M(\xi^*))f(x)\xi(dx) = \inf_{x \in X}\ f^T(x)\Psi'(M(\xi^*))f(x),$$

mamy

$$\inf_{x \in X}\ f^T(x)\Psi'(M(\xi^*))f(x) \geq \int_X f^T(x)\Psi'(M(\xi^*))f(x)\xi^*(dx).$$

Ostatnia nierówność może być tylko równością, bo prawa strona w sposób oczywisty jest nie mniejsza od lewej. To znaczy, że 2 jest udowodnione.

W odwrotną strone ($2 \implies 1$) dowód jest oczywisty. Z Lematu 6.1, konieczny i dostateczny warunek na to, aby M^* dostarczala minimum dla $\Psi(M)$, jest spelnienie nierówności (7), tzn. spełnienie nierówności

$$\inf_{\xi \in \Omega}\ \int_X f^T(x)\Psi'(M(\xi^*))f(x)\xi(dx) \geq \inf_{x \in X}\ f^T(x)\Psi'(M(\xi^*))f(x),$$

która jest oczywista. $\square$

Wniosek 6.1 *(Twierdzenie równoważności Kiefera-Wolfowits'a)*
Jeśli zbiór macierzy informacji $\mathcal{M}$ jest zwartym, to następne stwierdzenia są równoważne:

1. *plan ξ^* jest D-optymalnym;*
2. $\max_{x \in X}\ d(x, \xi^*) = k$ *(przypominanie : $d(x, \xi) = f^T(x)M^{-1}(\xi)f(x)$);*

3. supp ξ^* jest skoncentrowany w punktach $z \in X$ dla których

$$d(z, \xi^*) = \max_{x \in X} d(x, \xi^*);$$

4. plan ξ^* jest G-optymalnym.

Dowód. Dla plana D-optymalnego mamy

$$\Psi(M) = det\ M^{-1}, \quad \Psi'(M) = -(det\ M^{-1})M^{-1} \qquad \Longrightarrow$$

$$tr\,(M(\xi)\Psi'(M(\xi))) = -k det\, M^{-1}(\xi),\ f^T(x)\Psi'(M(\xi))f(x) = -(det\, M^{-1}(\xi))d(x, \xi).$$

Stwierdzenia 1–3 Wniosku 6.1 wynikają bespośrednio z Twierdzenia 6.1. Dla dowodu 4 zauważmy, że tożsamość (6) tutaj ma postać

$$\forall \xi \in \Omega \quad \int_X d(x, \xi)\xi(dx) = k,$$

skąd

$$\forall \xi \in \Omega \quad \max_{x \in X} d(x, \xi) \geq k \quad \text{tzn.} \quad \min_{\xi \in \Omega} \max_{x \in X} d(x, \xi) = k. \tag{8}$$

Jak wiadomo, plan $\tilde{\xi}$ jest G-optymalny wtedy i tylko wtedy gdy

$$\forall \xi \in \Omega \quad \max_{x \in X} d(x, \xi) \geq \max_{x \in X} d(x, \tilde{\xi}). \tag{9}$$

Teraz G-optymalność D-optymalnego planu ξ^* wynika z 2 i 3.

Odwrotne też jest prawdziwe, bo jeśli 4, tzn. (9), jest spełnione dla $\tilde{\xi}$, to dla $\tilde{\xi}$ ze względu na (8) musi być także spełnione 2. □

Wniosek 6.2 *Jeśli zbiór macierzy informacji $\mathcal{M}$ jest zwartym, to następne stwierdzenia są równoważne :*

1. plan ξ^* jest A-optymalnym;

2. $\max_{x \in X} f^T(x)M^{-2}(\xi^*)f(x) = tr\ M^{-1}(\xi^*)$;

3. supp ξ^* jest skoncentrowany w punktach $z \in X$ dla których

$$f^T(z)M^{-2}(\xi^*)f(z) = \max_{x \in X} f^T(x)M^{-2}(\xi^*)f(x).$$

Przykłady używania Twierdzeń równoważności

1. Niech $\eta(x,\theta) = \theta$ $(k = 1)$. Wtedy dla każdego plana $\xi \in \Omega$ mamy $M(\xi) = 1$ i każdy plan jest optymalny wg dowolnego kryterium optymalności.

2. (D-optymalny) Niech $\eta(x,\theta) = \theta_1 + \theta_2 x$ $(m = 1,\ k = 2)$, tzn. $f(x) = (1,x)^T,\ X = [-1,1]$. Dla dowolnego $\xi \in \Omega$ mamy

$$M(\xi) = \begin{pmatrix} 1 & \int_{-1}^{1} x\xi(dx) \\ \int_{-1}^{1} x\xi(dx) & \int_{-1}^{1} x^2\xi(dx) \end{pmatrix},$$

$$det\ M(\xi) = \int_{-1}^{1} x^2\xi(dx) - \left(\int_{-1}^{1} x\xi(dx)\right)^2.$$

Jasne, że pierwszy składnik jest ≤ 1, a drugi ≥ 0. Więc, $det\ M(\xi) \leq 1$ i plan

$$\xi^* = \begin{Bmatrix} -1 & 1 \\ 1/2 & 1/2 \end{Bmatrix}$$

jest D-optymalnym. (**Zadanie domowe.** Sprawdzić to, używając Wniosek 6.1.)

3. (D-optymalny) Niech $\eta(x,\theta) = \theta_1 + \theta_2 x + \theta_3 x^2$ $(m = 1,\ k = 3)$, tzn. $f(x) = (1,x,x^2)^T,\ X = [-1,1]$. Za pomocą Wniosku 6.1 sprawdzmy, że plan

$$\xi^* = \begin{Bmatrix} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$$

jest D-optymalnym. Rzeczywiście,

$$M(\xi^*) = \begin{pmatrix} 1 & 0 & 2/3 \\ 0 & 2/3 & 0 \\ 2/3 & 0 & 2/3 \end{pmatrix}, \qquad M^{-1}(\xi^*) = \begin{pmatrix} 3 & 0 & -3 \\ 0 & 3/2 & 0 \\ -3 & 0 & 9/2 \end{pmatrix},$$

$$d(x,\xi^*) = 3 - \frac{9}{2}x^2(1-x^2).$$

Maksimum po x funkcji $d(x,\xi^*)$ jest równy 3 i osiągalny w punktach $\{-1,0,1\}$.

4. (A-optymalny) Niech $\eta(x,\theta) = \theta_1 + \theta_2 x + \theta_3 x^2$ $(m = 1,\ k = 3)$, tzn. $f(x) = (1,x,x^2)^T,\ X = [-1,1]$. Ze względów symetrii załóżmy, że A-optymalny plan ma postać

$$\xi = \begin{Bmatrix} -1 & 0 & 1 \\ p/2 & 1-p & p/2 \end{Bmatrix}$$

i znajdziemy $0 < p < 1$. Dla planu ξ mamy

$$M(\xi) = \begin{pmatrix} 1 & 0 & p \\ 0 & p & 0 \\ p & 0 & p \end{pmatrix}, \; M^{-1}(\xi) = \begin{pmatrix} \frac{1}{1-p} & 0 & -\frac{1}{1-p} \\ 0 & \frac{1}{p} & 0 \\ -\frac{1}{1-p} & 0 & \frac{1}{p(1-p)} \end{pmatrix}, \; tr\; M^{-1}(\xi) = \frac{2}{p(1-p)}.$$

Minimum po p ostatniej funkcji jest równy 8 i osiągalny dla $p = 1/2$. Za pomocą Wniosku 6.2 sprawdzmy, że plan

$$\xi^* = \begin{Bmatrix} -1 & 0 & 1 \\ 1/4 & 1/2 & 1/4 \end{Bmatrix}$$

jest A-optymalnym. Rzeczywiście,

$$M^{-1}(\xi^*) = \begin{pmatrix} 2 & 0 & -2 \\ 0 & 2 & 0 \\ -2 & 0 & 4 \end{pmatrix}, \qquad tr\; M^{-1}(\xi^*) = 8,$$

$$f^T(x) M^{*-2} f(x) = 8 - 20x^2(1-x^2).$$

Maksimum po x ostatniej funkcji jest równy 8 i osiągalny w punktach $\{-1, 0, 1\}$.

5. Zaznaczmy, że dla planów dyskretnych ścisłych twierdzenia równoważności nie muszą być prawdziwe. Pokażemy, że Wniosek 6.1 nie jest prawdziwy w tym przypadku.

Niech $\eta(x, \theta) = \theta_1 + \theta_2 x$ $(m = 1,\; k = 2)$, tzn. $f(x) = (1, x)^T,\; X = [-1, 1]$, a ilość obserwacji $n = 3$. Znajdziemy dyskretny ścisły D-optymalny plan postaci

$$\xi = \begin{Bmatrix} x_1 & x_2 & x_3 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix},$$

gdzie przynajmniej dwa punkty spośród x_i są różne. Mamy

$$M(\xi) = \sum_{i=1}^{3} \frac{1}{3} f(x_i) f^T(x_i) = \frac{1}{3} \begin{pmatrix} 3 & x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 & x_1^2 + x_2^2 + x_3^2 \end{pmatrix},$$

$$det\; M(\xi) = \frac{1}{9} \left((x_2 - x_1)^2 + (x_3 - x_2)^2 + (x_3 - x_1)^2 \right).$$

Ostatnia funkcja osiąga maksimum równe $8/9$ kiedy $x_i \in \{-1, 1\}$. Więc, dyskretne ścisłe D-optymalne plany są postaci

$$\xi_1 = \begin{Bmatrix} -1 & 1 \\ 1/3 & 2/3 \end{Bmatrix}, \qquad \xi_2 = \begin{Bmatrix} -1 & 1 \\ 2/3 & 1/3 \end{Bmatrix}.$$

Rozważmy też dyskretny ścisły plan

$$\xi_3 = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{array} \right\}.$$

Mamy

$$M(\xi_1) = \begin{pmatrix} 1 & 1/3 \\ 1/3 & 1 \end{pmatrix}, \qquad M^{-1}(\xi_1) = \begin{pmatrix} 9/8 & -3/8 \\ -3/8 & 9/8 \end{pmatrix},$$

$$d(x, \xi_1) = 1 + \frac{1}{8}(3x - 1)^2,$$

$$M(\xi_2) = \begin{pmatrix} 1 & -1/3 \\ -1/3 & 1 \end{pmatrix}, \qquad M^{-1}(\xi_2) = \begin{pmatrix} 9/8 & 3/8 \\ 3/8 & 9/8 \end{pmatrix},$$

$$d(x, \xi_2) = 1 + \frac{1}{8}(3x + 1)^2,$$

$$M(\xi_3) = \begin{pmatrix} 1 & 0 \\ 0 & 2/3 \end{pmatrix}, \qquad M^{-1}(\xi_3) = \begin{pmatrix} 1 & 0 \\ 0 & 3/2 \end{pmatrix},$$

$$d(x, \xi_3) = 1 + \frac{3}{2}x^2,$$

$$\max_{x \in [-1,1]} d(x, \xi_1) = \max_{x \in [-1,1]} d(x, \xi_2) = 3, \qquad \max_{x \in [-1,1]} d(x, \xi_3) = \frac{5}{2} < 3.$$

To oznacza, że η_1 i η_2 nie mogą być planami G-optymalnymi.

7 D-optymalne plany dyskretne uogólnione dla regresji wielomianowej.

D-optymalne plany dyskretne uogólnione dla regresji wielomianowej

Niech $\eta(x,\theta) = \theta_1 + \theta_2 x + \ldots + \theta_k x^{k-1}$, tzn. $f(x) = (1, x, \ldots, x^{k-1})^T$, $X = [-1, 1]$. Funkcja $d(x,\xi)$ ma postać

$$d(x,\xi) = f^T(x) M^{-1}(\xi) f(x) = \sum_{i,j=1}^{k} m_{ij}(\xi) x^{i+j-2},$$

gdzie $m_{ij}(\xi)$ są elementami macierzy $M^{-1}(\xi)$. Jest to wielomian stopnia $(2k-2)$, który ma w przedziale $[-1,1]$ nie więcej niż $2k-3+2 = 2k-1$ lokalnych ekstremów. Jasne, że lokalne minima sosiadują z lokalnymi maksymami. To oznacza, że lokalnych maksymów jest $k-1$ lub k. Z Wniosku 6.1 wynika, że liczba punktów nośnika plana D-optymalnego jest równa k. W tym przypadku spośród nich są punkty -1 i 1. Więc, szukamy planu dyskretnego uogólnionego postaci

$$\xi = \left\{ \begin{array}{ccccc} -1 & x_2 & \ldots\ldots & x_{k-1} & 1 \\ p_1 & p_2 & \ldots\ldots & p_{k-1} & p_k \end{array} \right\}, \qquad x_2 < \ldots < x_{k-1},$$

który maksymalizuje wartość

$$det\ M(\xi) = det\ (\sum_{i=1}^{k} p_i f(x_i) f^T(x_i)) = (det\ F)^2 \prod_{i=1}^{k} p_i. \tag{10}$$

Jasne, że problem maksymalizacji (10) rozpada się (jest to bardzo ważne !) na dwie niezależne części : maksymalizacja $\prod_{i=1}^{k} p_i$ poprzez wybieranie wag $\{p_i\}$ i maksymalizacja $|det\ F|$ poprzez wybieranie punktów $\{x_i\}$. Pierwsze jest proste, bo z nierówności pomiędzy średnim geometrycznym i średnim arytmetycznym wynika, że

$$\prod_{i=1}^{k} p_i \leq \left(\frac{\sum_{i=1}^{k} p_i}{k} \right)^k = \frac{1}{k^k},$$

przy czym równość jest osiągalna wtedy i tylko wtedy, gdy wszystkie p_i są równe, tzn. $p_i = 1/k,\ i = 1, 2, \ldots, k$.

Można też policzyć, że

$$det\ F = \prod_{i,j=1,i>j}^{k} (x_i - x_j). \tag{11}$$

Zadanie domowe. Sprawdzić (11) w przypadku $k = 3$.

Punkty $\{x_2, \ldots, x_{k-1}\}$ (przypominamy, że $x_1 = -1$, $x_k = 1$), którzy maksymizują (11), są rozwiązaniami równania (otrzymujemy biorąc pochodne po x_i od prawej części (11))

$$\frac{1}{x_i - x_1} + \cdots + \frac{1}{x_i - x_{i-1}} + \frac{1}{x_i - x_{i+1}} + \cdots + \frac{1}{x_i - x_k} = 0,\ i = 2, \ldots, k-1,$$

czyli

$$\frac{1}{x_i + 1} + \frac{1}{x_i - 1} + \frac{1}{2}\frac{\phi''(x_i)}{\phi'(x_i)} = 0,\ i = 2, \ldots, k-1, \tag{12}$$

gdzie $\phi(x) = (x - x_2) \ldots (x - x_{k-1})$.

Zadanie domowe. Sprawdzić (12) w przypadku $k = 4,\ k = 5$.

Więc,

$$(x_i^2 - 1)\phi''(x_i) + 4x_i\phi'(x_i) = 0,\ i = 2, \ldots, k-1.$$

Z ostatniej równości wynika, że wielomian

$$(x^2 - 1)\phi''(x) + 4x\phi'(x)$$

jest tego samego stopnia $(k-2)$, co i $\phi(x)$, i posiada pierwiastki w tych samych punktach $\{x_2, \ldots, x_{k-1}\}$. To znaczy, że

$$(x^2 - 1)\phi''(x) + 4x\phi'(x) = c\phi(x)$$

dla pewnej stałej c. Przyrównując koeficjenty przy x^{k-2} z lewej i prawej strony znajdujemy $c = (k+1)(k-2)$. Więc,

$$(x^2 - 1)\phi''(x) + 4x\phi'(x) - (k+1)(k-2)\phi(x) = 0. \tag{13}$$

Zadanie domowe. Sprawdzić (13) w przypadku $k = 3,\ k = 4$.

Fakt 7.1 *W zbiorze wszystkich wielomianów równanie (13) posiada tylko jedno (z dokladnością do współczynnika) rozwiązanie, mianowicie* $\phi(x) = L'_{k-1}(x)$, *gdzie*

$$L_n(x) = \frac{1}{2^n n!}\frac{\partial^n (x^2 - 1)^n}{\partial x^n}, \qquad n \geq 1$$

jest wielomianem Legendre'a stopnia n *(*$L_0(x) = 1$*).*

Wynik. D-optymalnym planem dyskretnym uogólnionym dla regresji wielomianowej stopnia $(k-1)$ na przedziale $[-1,1]$ jest plan postaci

$$\xi^* = \left\{ \begin{array}{ccccc} -1 & x_2 & \ldots\ldots & x_{k-1} & 1 \\ 1/k & 1/k & \ldots\ldots & 1/k & 1/k \end{array} \right\},$$

gdzie $x_2 < \ldots < x_{k-1}$ są pierwiastkami pochodnej od wielomianu Legendre'a stopnia $(k-1)$ (lub równoważne : punktami ekstremalnymi wielomianu Legendre'a stopnia $(k-1)$).

Przykład. Niech $\eta(x,\theta) = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 x^3$ $(m = 1,\ k = 4)$. D-optymalnym planem dyskretnym uogólnionym jest plan postaci

$$\xi^* = \left\{ \begin{array}{cccc} -1 & x_2 & x_3 & 1 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{array} \right\},$$

przy czym $x_2,\ x_3$ są pierwiastkami wielomianu

$$L_3'(x) = \left(\frac{5}{2}x^3 - \frac{3}{2}x\right)' = \frac{15}{2}x^2 - \frac{3}{2}.$$

Więc, $x_2 = -1/\sqrt{5},\ x_3 = 1/\sqrt{5}$.

8 Metody numeryczne konstrukcji planów optymalnych.

Konstrukcja optymalnych planów dyskretnych uogólnionych

Z własności 5* macierzy informacji (patrz Rozdzial 4 i też Fakt 5.1 z Rozdzialu 5) wynika, że zajmując się minimizacjej funkcji $\Psi(M(\xi)),\ \xi \in \Omega$, wystarczy rozważać tylko plany dyskretne uogólnione, tzn. z Ω_0.

Bardzo rzadko można znaleźć plan Ψ-optymalny bespośrednio, dla tego posługują się metodami numerycznymi. Istnieje dużo różnych podejść do problemu konstrukcji planów Ψ-optymalnych. Będziemy tutaj rozważać podejście polegające na minimizacji $\Psi(M(\xi))$ poprzez wybór odpowiednich punktów nośnika $\{x_i\} \subset X$ i wag $p_i > 0,\ \sum_i p_i = 1$.

Główna zasada algorytmu jest następująca. Będziemy konstruować kolejne przybliżenia do planu optymalnego zmnieszając za każdym razem wartość funkcji $\Psi(M(\xi))$. Niech na kroku s mamy plan ξ_s. Następne przybliżenie ξ_{s+1} będziemy szukać w postaci $\xi_{s+1} = (1-\alpha)\xi_s + \alpha\xi = \xi_s + \alpha(\xi - \xi_s)$ wybierając w odpowiedni sposób $\xi \in \Omega_0$ i $\alpha \in [0,1]$. Mamy (patrz Lemat 6.1)

$$\left.\frac{\partial\Psi(M(\xi_{s+1}))}{\partial\alpha}\right|_{\alpha=0+} = tr\ (\Psi'(M(\xi_s))M(\xi)) - tr\ (\Psi'(M(\xi_s))M(\xi_s)) =$$

$$\int_X f^T(x)\Psi'(M(\xi_s))f(x)\xi(dx) - tr\ (\Psi'(M(\xi_s))M(\xi_s)) \geq \tag{14}$$

$$\min_{x\in X}\ f^T(x)\Psi'(M(\xi_s))f(x) - tr\ (\Psi'(M(\xi_s))M(\xi_s)).$$

Plan ξ, dla którego ma miejsce równość w nierówności (14), co daje najmniejszą wartość dla $\left.\frac{\partial\Psi(M(\xi_{s+1}))}{\partial\alpha}\right|_{\alpha=0+}$, to np. plan $\xi(x_{(s)})$ skoncentrowany w jednym punkcie

$$x_{(s)} = Arg\ \min_{x\in X}\ f^T(x)\Psi'(M(\eta_s))f(x) \tag{15}$$

z wagą 1 (jeśli takich punktów jest więcej, to możemy wziąc dowolny z nich). Jeśli prawa część (14) jest ujemna, to dla małych wartości $\alpha > 0$ największe zmniejszenie funkcji Ψ startując z ξ_s odbywa się własnie w kierunku $\xi(x_{(s)})$.

Następny algorytm minimizacji Ψ jest w zasadzie algorytmem gradientnego typu.

Algorytm 1. (*Procedura Fedorova-Wynna*)

1. Wybieramy początkowy plan nieosobliwy ξ_0.
2. Niech mamy ξ_s. Odnajdujemy $x_{(s)}$ wg wzoru (15).
3. Sprawdzamy czy
$$f^T(x_{(s)})\Psi'(M(\xi_s))f(x_{(s)}) < tr\ (\Psi'(M(\xi_s))M(\xi_s)).$$
Jeśli tak, to idziemy do 4. Jeśli nie, to plan ξ_s już jest szukany.
4. Budujemy $\xi_{s+1} = (1-\alpha_s)\xi_s + \alpha_s\xi(x_{(s)})$. Wartość α_s można wybierać w różny sposób, np.
$$\alpha_s = Arg \min_{0\le\alpha\le 1}\ \Psi(M((1-\alpha)\xi_s + \alpha\xi(x_{(s)}))).$$
5. Sprawdzamy czy $\Psi(M(\xi_s)) - \Psi(M(\xi_{s+1})) \le \varepsilon$, gdzie $\varepsilon > 0$ jest zadaną z góry dostatecznie małą liczbą. Jeśli tak, to uważamy plan ξ_{s+1} za szukany. Jeśli nie, to idziemy do 2 z $s = s+1$.

Lemat 8.1 *Niech warunki Twierdzenia równoważności są spełnione. Wtedy dla każdego nieosobliwego planu* ξ

$$tr\ (\Psi'(M(\xi))M(\xi)) - \min_{x\in X}\ f^T(x)\Psi'(M(\xi))f(x) \ge \Psi(M(\xi)) - \min_{\xi}\ \Psi(M(\xi)).$$

Dowód. Rozważmy $\xi_\alpha = (1-\alpha)\xi + \alpha\xi^*$, gdzie plan ξ^* jest Ψ-optymalny. Z wypukłości Ψ mamy

$$\Psi(M(\xi_\alpha)) \le (1-\alpha)\Psi(M(\xi)) + \alpha\Psi(M(\xi^*)),$$

skąd

$$\frac{\Psi(M(\xi_\alpha)) - \Psi(M(\xi))}{\alpha} \le \Psi(M(\xi^*)) - \Psi(M(\xi))$$

i

$$\left.\frac{\partial\Psi(M(\xi_\alpha))}{\partial\alpha}\right|_{\alpha=0+} \le \Psi(M(\xi^*)) - \Psi(M(\xi)).$$

Z drugiej strony, mamy nierówność (14), tzn.

$$\left.\frac{\partial\Psi(M(\xi_\alpha))}{\partial\alpha}\right|_{\alpha=0+} \ge \min_{x\in X}\ f^T(x)\Psi'(M(\xi))f(x) - tr\ (\Psi'(M(\xi))M(\xi)).$$

Dwie ostatnie nierówności dają teze lematu. □

Lemat 8.1 pozwala na sformulowanie wygodnego warunku zakonczenia pracy algorytmu. Rzeczywiście, jeśli chcemy uzyskać

$$\Psi(M(\xi_s)) - \min_{\xi} \Psi(M(\xi)) \leq \varepsilon,$$

to dla tego wystarczy, aby

$$\min_{x \in X} f^T(x)\Psi'(M(\xi_s))f(x) - tr\ (\Psi'(M(\xi_s))M(\xi_s)) \geq -\varepsilon.$$

Można udowodnić, że jeśli zbiory X i $\mathcal{M}$ są zwartymi, a Ψ jest funkcją wypukłą i różniczkowalną, to Algorytm 1 jest zbieżny, tzn.

$$\Psi(M(\xi_s)) \to \min_{\xi} \Psi(M(\xi)),\ s \to \infty.$$

Pokażemy też jak można powiększyć szybkość zbieżności Algorytmu 1 modyfikując go. Zastanowmy się, co będzie jeśli $x_{(s)}$ zdefiniowane wzorem (15) jest takie, że $x_{(s)} \in supp\ \xi_s$? Wtedy nośniki planów ξ_s i ξ_{s+1} są jednakowe i dla zmniejszenia Ψ wystarczy tylko zmienić wagi planu ξ_s. Jak łatwo policzyć, dopuszczalne wartości dla α wtedy są $-\frac{p(x_{(s)})}{1-p(x_{(s)})} \leq \alpha \leq 1$, gdzie $p(x_{(s)})$ jest wagą punktu $x_{(s)}$ w planie ξ_s. A więc w przedstawieniu $\xi_{s+1} = (1-\alpha)\xi_s + \alpha\xi$ wartość α może być ujemną ! Wybór wartości ujemnej dla α oznacza zmniejszenie wagi punktu $x_{(s)}$ w planie.

Algorytm 2.

1. Wybieramy początkowy plan nieosobliwy ξ_0.

2. Niech mamy ξ_s. Wybieramy $x_{(s)}$ wg zasady : $x_{(s)} = x^-_{(s)}$, jeśli

$$f^T(x^-_{(s)})\Psi'(M(\xi_s))f(x^-_{(s)}) - tr\ (\Psi'(M(\xi_s))M(\xi_s)) \leq$$

$$tr\ (\Psi'(M(\xi_s))M(\xi_s)) - f^T(x^+_{(s)})\Psi'(M(\xi_s))f(x^+_{(s)}),$$

i $x_{(s)} = x^+_{(s)}$ w pozostalych wypadkach, gdzie
$x^-_{(s)} = Arg \min_{x \in X} f^T(x)\Psi'(M(\xi_s))f(x)$ jest "najlepszym punktem" (to co było w poprzednim algorytmie), zaś
$x^+_{(s)} = Arg \max_{x \in supp\xi_s} f^T(x)\Psi'(M(\xi_s))f(x)$ jest "najgorszym punktem" spośród znajdujących się w $supp\ \xi_s$.

3. Sprawdzamy czy wartość

$$\min\{f^T(x^-_{(s)})\Psi'(M(\xi_s))f(x^-_{(s)}) - tr\ (\Psi'(M(\xi_s))M(\xi_s)),$$

$$tr\ (\Psi'(M(\xi_s))M(\xi_s)) - f^T(x^+_{(s)})\Psi'(M(\xi_s))f(x^+_{(s)})\}$$

jest ujemna. Jeśli tak, to idziemy do 4. Jeśli nie, to plan ξ_s już jest szukany.

4. Budujemy $\xi_{s+1} = (1-\alpha_s)\xi_s + \alpha_s\xi(x_{(s)})$, gdzie α_s wyznacza się jak w punkcie 4 z Algorytmu 1, jeśli $x_{(s)} = x^-_{(s)}$, i wg wzoru

$$\alpha_s = Arg \min_{-\frac{p(x_{(s)})}{1-p(x_{(s)})} \le \alpha \le 0} \Psi(M((1-\alpha)\xi_s + \alpha\xi(x_{(s)}))),$$

jeśli $x_{(s)} = x^+_{(s)}$.

5. Sprawdzamy czy $\Psi(M(\xi_s)) - \Psi(M(\xi_{s+1})) \le \varepsilon$, gdzie $\varepsilon > 0$ jest zadaną z góry dostatecznie małą liczbą. Jeśli tak, to uważamy plan ξ_{s+1} za szukany. Jeśli nie, to idziemy do 2 z $s = s+1$.

Następny przykład pokazuje różnice w działaniu Algorytmów 1 i 2.

Przykład. Niech $\eta(x,\theta) = \theta_1 + \theta_2 x,\ X = [-1,1]$. Zastosujmy powyższe algorytmy dla kryterium D-optymalności i początkowego planu

$$\xi_0 = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{array} \right\}.$$

Mamy

$$M(\xi_0) = \begin{pmatrix} 1 & 0 \\ 0 & 2/3 \end{pmatrix}, \qquad det\ M(\xi_0) = 2/3.$$

Zastosujmy jeden krok Algorytmu 1. Ponieważ

$$f^T(x)\Psi'(M(\xi_0))f(x) = -\frac{3}{2}\begin{pmatrix} 1 & x \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 3/2 \end{pmatrix}\begin{pmatrix} 1 \\ x \end{pmatrix} = -\frac{3}{2} - \frac{9}{4}x^2 \implies$$

$$x_{(0)} = -1 \text{ (lub } 1).$$

Mamy

$$tr\ (\Psi'(M(\xi_0))M(\xi_0)) = -2det\ M^{-1}(\xi_0) = -3 > -3/2 - 9/4$$

wieć budujemy plan

$$\xi_1 = (1-\alpha)\xi_0 + \alpha\xi(x_{(0)}) = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ (1+2\alpha)/3 & (1-\alpha)/3 & (1-\alpha)/3 \end{array} \right\}.$$

Dla niego

$$M(\xi_1) = \left(\begin{array}{cc} 1 & -\alpha \\ -\alpha & (2+\alpha)/3 \end{array} \right), \qquad det\ M(\xi_1) = (2+\alpha)/3 - \alpha^2.$$

Ponieważ $\alpha_0 = Arg \max\limits_{0\le\alpha\le 1}\ det\ M(\xi_1) = 1/6$, to ostatecznie

$$\xi_1 = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 4/9 & 5/18 & 5/18 \end{array} \right\}, \quad M(\xi_1) = \left(\begin{array}{cc} 1 & -1/6 \\ -1/6 & 13/18 \end{array} \right), \quad det\ M(\xi_1) = 25/36.$$

Teraz zastosujmy jeden krok Algorytmu 2. Mamy

$$x^-_{(0)} = -1,\ f^T(x^-_{(0)})\Psi'(M(\xi_0))f(x^-_{(0)}) - tr\,(\Psi'(M(\xi_0))M(\xi_0)) = -15/4+3 = -3/4,$$

$$x^+_{(0)} = 0,\ tr\,(\Psi'(M(\xi_0))M(\xi_0)) - f^T(x^+_{(0)})\Psi'(M(\xi_0))f(x^+_{(0)})$$
$$= -3 + 3/2 = -3/2 < -3/4.$$

Więc, $x_{(0)} = x^+_{(0)}$ i budujemy plan

$$\xi_1 = (1-\alpha)\xi_0 + \alpha\xi(x_{(0)}) = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ (1-\alpha)/3 & (1+2\alpha)/3 & (1-\alpha)/3 \end{array} \right\}.$$

Dla niego

$$M(\xi_1) = \left(\begin{array}{cc} 1 & 0 \\ 0 & 2(1-\alpha)/3 \end{array} \right), \qquad det\ M(\xi_1) = 2(1-\alpha)/3.$$

Ponieważ $\alpha_0 = Arg \max\limits_{-1/2\le\alpha\le 0}\ det\ M(\xi_1) = -1/2$, to ostatecznie

$$\xi_1 = \left\{ \begin{array}{cc} -1 & 1 \\ 1/2 & 1/2 \end{array} \right\}$$

- jest to już plan D-optymalny, jak wiemy.

Konstrukcja optymalnych planów dyskretnych ścisłych

Niech mamy przeprowadzić n doświadczeń i niech plan

$$\xi^* = \left\{ \begin{array}{ccc} x_1^* & \ldots\ldots & x_r^* \\ p_1^* & \ldots\ldots & p_r^* \end{array} \right\}$$

jest Ψ-optymalnym dyskretnym uogólnionym. Naturalnym przybliżeniem dla ξ^* w klasie planów dyskretnych ścisłych jest plan, który można uzyskać z ξ^* poprzez "zaokrąglenie" wag. Np. bierzemy jako przybliżenie plan postaci

$$\xi^{(n)} = \left\{ \begin{array}{ccc} x_1^* & \ldots\ldots & x_r^* \\ \frac{n_1+\alpha_1}{n} & \ldots\ldots & \frac{n_r+\alpha_r}{n} \end{array} \right\},$$

gdzie $\{n_i\}$ są liczbami naturalnymi takimi, że $n_i \leq np_i^*$ (mogą one być wybrane zgodnie ze wzorem $n_i = [np_i^*]$), a liczby $\{\alpha_i\}$ przyjmują wartość 0 lub 1 tak, aby $\sum_{i=1}^r \alpha_i = n - \sum_{i=1}^r n_i$. Można ich wybrać w sposób następujący. Niech wartości $\{np_i^* - n_i\}$ są uporządkowane w porządku malejącym. Wtedy jeśli $(np_j^* - n_j)$ stoi na jednym z pierwszych $(n - \sum_{i=1}^r n_i)$ miejsc, to przyjmujemy $\alpha_j = 1$, w przeciwnym wypadku $\alpha_j = 0$.

Lemat 8.2 *Niech funkcja Ψ jest monotoniczna i jednorodna. Wtedy*

$$\Psi(M(\xi^{(n)})) - \Psi(M(\xi^*)) \leq \left(\frac{\gamma(a)}{\gamma(n)} - 1 \right) \Psi(M(\xi^*)),$$

gdzie $a = \min\limits_i \frac{[np_i^*]}{p_i^*}$ (funkcja γ - z definicji jednorodności kryterium optymalności (patrz Rozdzial 5)).

Dowód. Z definicji jednorodności

$$\gamma(n)\Psi(M(\xi^{(n)})) = \Psi(nM(\xi^{(n)})) = \Psi\left(\sum_{i=1}^r (n_i + \alpha_i) f(x_i^*) f^T(x_i^*) \right),$$

$$\gamma(a)\Psi(M(\xi^*)) = \Psi(aM(\xi^*)) = \Psi\left(\sum_{i=1}^r a p_i^* f(x_i^*) f^T(x_i^*) \right).$$

Ponieważ $n_i + \alpha_i \geq ap_i^*$ dla każdego i, to

$$\sum_{i=1}^r (n_i + \alpha_i) f(x_i^*) f^T(x_i^*) \geq \sum_{i=1}^r a p_i^* f(x_i^*) f^T(x_i^*)$$

i potrzebną nierówność mamy teraz z monotoniczności kryterium Ψ. $\square$

Uwaga. Przybliżenie $\xi^{(n)}$ do ξ^* polepsza się ze wzrostem n ponieważ $\gamma(a)/\gamma(n) \to 1,\ n \to \infty$.

Przykład. Zastosujmy opisaną procedure "zaokrąglenia" do planu A-optymalnego w przypadku $\eta(x,\theta) = \theta_1 + \theta_2 x,\ X = [0,1]$, który jest postaci

$$\xi^* = \left\{ \begin{array}{cc} 0 & 1 \\ 2-\sqrt{2} & \sqrt{2}-1 \end{array} \right\}, \quad tr\ M^{-1}(\xi^*) = \sqrt{2} - 1.$$

Mamy $r = 2$ i niech najpierw $n = 10$. Wtedy $n_1 = 5,\ n_2 = 4,\ \alpha_1 = 1,\ \alpha_2 = 0$, bo $2 - \sqrt{2} = 0.585...,\ \sqrt{2} - 1 = 0.414....$ Odpowiedni plan dyskretny ścisły ma postać

$$\xi^{(10)} = \left\{ \begin{array}{cc} 0 & 1 \\ 0.6 & 0.4 \end{array} \right\}.$$

Ponieważ $a = \min\{\frac{5}{0.585}, \frac{4}{0.414}\} = 8.5$, to zgodnie z Lematem 8.2

$$tr\ M^{-1}(\xi^{(10)}) - tr\ M^{-1}(\xi^*) \leq \left(\frac{10}{8.5} - 1\right)(\sqrt{2} - 1) \approx 0.07,$$

a więc przybliżenie już jest w miare dobre. Popatrzmy co się zmieni, jeśli $n = 100$. Wtedy $n_1 = 58,\ n_2 = 41,\ \alpha_1 = 1,\ \alpha_2 = 0$. Odpowiedni plan dyskretny ścisły ma postać

$$\xi^{(100)} = \left\{ \begin{array}{cc} 0 & 1 \\ 0.59 & 0.41 \end{array} \right\}.$$

Tutaj $a = \min\{\frac{58}{0.585}, \frac{41}{0.414}\} = 99.03$ i zgodnie z Lematem 8.2

$$tr\ M^{-1}(\xi^{(100)}) - tr\ M^{-1}(\xi^*) \leq \left(\frac{100}{99.03} - 1\right)(\sqrt{2} - 1) \approx 0.004.$$

To oznacza, że przybliżenie jest znacznie lepsze : zwiększenie n w 10 razy dalo polepszenie przybliżenia w 18 razy.

Doświadczenia wykazały, że dość dobrze spisuje się też następny prosty algorytm, którego idea jest zbliżona do tej realizowanej w Algorytmie 2.

Algorytm 3.

1. Wybieramy początkowy plan nieosobliwy postaci (punkty nośnika mogą się powtarzać)
$$\xi_0^{(n)} = \left\{ \begin{array}{ccc} x_1^{(0)} & \ldots\ldots & x_n^{(0)} \\ 1/n & \ldots\ldots & 1/n \end{array} \right\}.$$

2. Niech mamy $\xi_s^{(n)}$. Odnajdujemy $x_{(s)}^-$ wg wzoru
$$x_{(s)}^- = Arg \min_{x \in X} f^T(x)\Psi'(M(\xi_s^{(n)}))f(x).$$

3. Budujemy plan pomocniczy
$$\xi_s^{(n+1)} = \left(1 - \frac{1}{n+1}\right)\xi_s^{(n)} + \frac{1}{n+1}\xi(x_{(s)}^-) =$$
$$\left\{ \begin{array}{cccc} x_1^{(s)} & \ldots\ldots & x_n^{(s)} & x_{(s)}^- \\ 1/(n+1) & \ldots\ldots & 1/(n+1) & 1/(n+1) \end{array} \right\}.$$

4. Odnajdujemy $x_{(s)}^+$ wg wzoru
$$x_{(s)}^+ = Arg \max_{x \in supp\ \xi_s^{(n+1)}} f^T(x)\Psi'(M(\xi_s^{(n+1)}))f(x).$$
Jeśli $x_{(s)}^+ = x_{(s)}^-$, to plan $\xi_s^{(n)}$ już jest szukany. W przeciwnym wypadku idziemy do 5.

5. Budujemy plan
$$\xi_{s+1}^{(n)} = \left(1 + \frac{1}{n}\right)\xi_s^{(n+1)} - \frac{1}{n}\xi(x_{(s)}^+) = \left\{ \begin{array}{ccc} x_1^{(s+1)} & \ldots\ldots & x_n^{(s+1)} \\ 1/n & \ldots\ldots & 1/n \end{array} \right\}.$$

6. Sprawdzamy czy $\Psi(M(\xi_s^{(n)})) - \Psi(M(\xi_{s+1}^{(n)})) \leq \varepsilon$, gdzie $\varepsilon > 0$ jest zadaną z góry dostatecznie małą liczbą. Jeśli tak, to uważamy plan $\xi_{s+1}^{(n)}$ za szukany. Jeśli nie, to idziemy do 2 z $s = s + 1$.

Przykład. Niech $\eta(x, \theta) = \theta_1 + \theta_2 x,\ X = [-1, 1]$. Zastosujmy Algorytm 3 w przypadku $n = 3$ dla kryterium D-optymalności (patrz Przykład 5 z Rozdzialu 6) do początkowego planu
$$\xi_0^{(3)} = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{array} \right\}.$$

Ponieważ

$$M(\xi_0^{(3)}) = \begin{pmatrix} 1 & 0 \\ 0 & 2/3 \end{pmatrix}, \quad det\ M(\xi_0^{(3)}) = 2/3,$$

to

$$f^T(x)\Psi'(M(\xi_0^{(3)}))f(x) = -\frac{3}{2}\begin{pmatrix} 1 & x \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 3/2 \end{pmatrix}\begin{pmatrix} 1 \\ x \end{pmatrix} = -\frac{3}{2} - \frac{9}{4}x^2 \implies$$

$x_{(0)}^- = -1$ (albo $= 1$). Budujemy plan

$$\xi_0^{(4)} = \left(1 - \frac{1}{4}\right)\xi_0^{(3)} + \frac{1}{4}\xi(x_{(0)}^-) =$$

$$\begin{Bmatrix} -1 & -1 & 0 & 1 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{Bmatrix}.$$

Ponieważ

$$M(\xi_0^{(4)}) = \begin{pmatrix} 1 & -1/4 \\ -1/4 & 3/4 \end{pmatrix}, \quad det\ M(\xi_0^{(4)}) = 11/16,$$

to

$$f^T(x)\Psi'(M(\xi_0^{(4)}))f(x) = -\frac{64}{121}\begin{pmatrix} 1 & x \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 1 & 4 \end{pmatrix}\begin{pmatrix} 1 \\ x \end{pmatrix} = -\frac{64}{121}(4x^2 + 2x + 3)$$

$$\implies x_{(0)}^+ = 0.$$

Budujemy plan

$$\eta_1^{(3)} = \left(1 + \frac{1}{3}\right)\xi_0^{(4)} - \frac{1}{3}\xi(x_{(s)}^+) = \begin{Bmatrix} -1 & -1 & 1 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix} = \begin{Bmatrix} -1 & 1 \\ 2/3 & 1/3 \end{Bmatrix}.$$

Jest to juz plan D-optymalny, jak wiemy.

Zadanie domowe. Sprawdzić, że na następnym kroku Algorytmu 3 dostaniemy $x_{(1)}^- = x_{(1)}^+ = 1$.

9 Doświadczenia czynnikowe.

Jest to najstarsza i najczęściej używana w praktyce dziedzina teorii planowania doświadczeń. Zameszczamy tu tylko jej zarys.

Rozważmy następujący problem. Niech wyniki doświadczenia zależa od m *czynników* $x_1,\ x_2,\ \ldots,\ x_m$, przy czym czynnik x_i może przyjąc tylko skończoną ilość $s_i \geq 2$ wartości, $i = 1, \ldots, m$. Nazywają sie te wartości *poziomami czynnika* i zakładamy, że możemy ich wybierać przed prowadzeniem doświadczeń. Często mają one raczej *charakter jakościowy.* Naszym celem jest badanie wpływu czynników na wyniki doświadczeń. Takie doświadczenia nazywają się *czynnikowymi.*

Zakładają, że schemat w ramach którego przeprowadzamy doświadczenia jest schematem regresji liniowej z funkcją regresji postaci :

$$\eta(x,\theta) = \theta_0 + \sum_{i=1}^{m} \theta_i x_i \quad - \quad \text{regresja liniowa,}$$

$$\eta(x,\theta) = \theta_0 + \sum_{i=1}^{m} \theta_i x_i + \sum_{i<j} \theta_{ij} x_i x_j \quad - \quad \text{regresja niepełna kwadratowa.}$$

Tutaj parametr θ_0 nazywa się *wspólną średnią*, parametr θ_i – *głownym efektem* i-go czynnika, parametr θ_{ij} – *efektem współdzialań (interakcji)* i-go i j-go czynników.

Planem czynnikowym nazywamy kombinacje poziomów czynników dla których przeprowadzamy doświadczenia. Plan czynnikowy nazywamy *pełnym*, jeśli zgodnie z tym planem przeprowadzamy doświadczenia dla wszystkich możliwych kombinacji poziomów czynników. Odpowiednie doświadczenia też nazywamy pełnymi. Pełny plan czynnikowy wymaga przeprowadzenia $s_1 s_2 \ldots s_m$ doświadczeń i mówią, że plan ten jest typu $s_1 s_2 \ldots s_m$. Jeśli $s_1 = \ldots = s_m = s$, to pełny plan czynnikowy poznacza się jako s^m. Plan czynnikowy nazywamy *ułamkowym*, jeśli dla jego realizacji przeprowadzamy mniej niż $s_1 s_2 \ldots s_m$ doświadczeń (tzn. nie dla wszystkich możliwych kombinacji czynników). Plany ułamkowe mają większe praktyczne zastosowanie niż pełne z tego powodu że wymagają znacznie mniej doświadczeń.

Przykład. Niech $\eta(x,\theta) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$, czynniki x_1 i x_2 są dwopoziomowe, przyjmujące wartości ± 1. Pełny plan czynnikowy tutaj jest planem

typu 2^2 i zwykle zapisuje się w postaci tabeli:

x_0	x_1	x_2
1	1	1
1	-1	1
1	1	-1
1	-1	-1

Wiec, realizując pelny plan czynnikowy typu 2^2 musimy przeprowadzić 4 doświadczenia. Jeśli wyrzucimy niektóre wiersze z tabeli, to otrzymamy plan ułamkowy. Np. wyrzucając dwa ostatnie wiersze z tabeli otrzymamy plan ulamkowy typu 2^{2-1}, zgodnie z którym musimy przeprowadzić dwa doświadczenia.

Zwykle celem przeprowadzenia doświadczeń jest oszacowanie nieznanych parametrów modelu, czyli parametrów $\theta_0,\ \theta_i,\ \theta_{ij}$. Zakładając, że błędy pomiarów mają zerowe wartości oczekiwane i są one nieskorelowane o jednakowej wariancji, otzrymujemy estymator MNK zgodnie z Twierdzeniem Gaussa-Markowa postaci

$$\widehat{\theta} = (F^T F)^{-1} F^T y, \tag{16}$$

gdzie jak zwykle $y = (y_1, \ldots, y_n)^T,\ \theta = (\theta_0, \theta_1, \ldots, \theta_m,\ \theta_{11},\ \ldots)^T$, natomiast macierz F - macierz planu każdy wiersz której jest odpowiednią kombinacją wartości czynników dla których przeprowadzamy doświadczenia.

Więc, tabela podana w powyższym przykładzie jest pożyteczna jeszcze z tego powodu, że daje wartości elementów macierzy planu F.

W praktyce najpopularnejsze są (i tylko ich będziemy dalej rozważać) plany dwopoziomowe, kiedy czynniki przyjmują wartości tylko na dwoch poziomach (zwykle to ± 1). Motywacja:
a) jeśli czynnik przyjmuje wartości na s poziomach, to zawsze można go zamienić na s czynników dwopoziomowych rozważając każdy poziom jako nowy czynnik przyjmujący wartości: $+1$ (czynnik obecny), -1 (czynnik nieobecny);
b) w praktyce przy prowadzeniu doświadczenia typowej jest sytuacja kiedy czynniki mogą lub być obecnymi $(+1)$ lub też nie (-1), a więc są dwopoziomowe;
c) jeśli czynnik przyjmuje wartości na wielu poziomach lub wręcz na calym przedzialie, to często jest sens rozważać tylko dwa skrajne poziomy.

Rozważmy teraz przypadek $m = 3$ i najpierw model regresji liniowej tzn.

$$\eta(x, \theta) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3.$$

Pełny plan czynnikowy typu 2^3 zapisuje się w postaci:

x_0	x_1	x_2	x_3
1	1	1	1
1	-1	1	-1
1	1	-1	-1
1	-1	-1	1
1	1	1	-1
1	-1	1	1
1	1	-1	1
1	-1	-1	-1

Macierz planu F ma postać

$$F = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & -1 & -1 \end{pmatrix}.$$

Jest to macierz wymiaru 8×4 i pełnego rzędu 4. Jak łatwo policzyć,

$$F^T F = \begin{pmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix},$$

a więc wiersze macierzy F są ortogonalne. Plany dla których macierze planu mają ortogonalne wiersze też nazywają się ortogonalnymi. Są one bardzo wygodne z obliczeniowego punktu widzenia.

Stosując formule (16) możemy teraz oszacować w sposób nieobciążony wszystkie efekty główne.

Pytanie 1. Mając do oszacowania 4 nieznane współczynniki, jaka jest najmniejsza ilość doświadczeń po przeprowadzeniu której moglibyśmy oszacować je w sposób nieobciążony ? Oczywiście, 4. Więc, wystarzy tutaj zastosować

plan ułamkowy typu $2^3 - 1$ np. postaci:

x_0	x_1	x_2	x_3
1	1	1	1
1	-1	1	-1
1	1	-1	-1
1	-1	-1	1

(17)

lub postaci:

x_0	x_1	x_2	x_3
1	1	1	-1
1	-1	1	1
1	1	-1	1
1	-1	-1	-1

(18)

Te plany nazywają się też *półpowtórzeniami*.

Pytanie 2. Mając przeprowadzić 8 doświadczeń i stosując pełny plan czynnikowy typu 2^3, ilu najwięcej nieznanych współczynników moglibyśmy oszacować w sposób nieobciążony ? Oczywiście, 8. Więc, moglibyśmy użyc niepełnego modelu kwadratowego

$$\eta(x, \theta) = \theta_0 + \sum_{i=1}^{3} \theta_i x_i + \sum_{i<j} \theta_{ij} x_i x_j$$

gdzie uwzględnione są wszystkie interakcje podwójne (zauważmy, że niema sensu tutaj wlączać do modelu składniki typu $\theta_{ii} x_i^2$, bo $x_i^2 \equiv 1$), lub wręcz modelu

$$\eta(x, \theta) = \theta_0 + \sum_{i=1}^{3} \theta_i x_i + \sum_{i<j} \theta_{ij} x_i x_j + \theta_{123} x_1 x_2 x_3 \tag{19}$$

gdzie uwzględniona również została interakcja potrójna.

Wróczmy jeszcze do planów ułamkowych typu (17) i (18). Jak łatwo zauważyć oba one są ortogonalne. Można też zauważyć, że ostatnia kolumna planu (17) ma własność $x_3 = x_1 x_2$, natomist ostania kolumna planu (18) ma własność $x_3 = -x_1 x_2$. Właśnie zwykłe w taki sposób tworzą plany ułamkowe: najpierw biorą plan pelny dla mniejszej iłosci czynników (pierwsze trzy kolumny w (17) lub (18)), a potem następne kolumny tworzą przez mnożenie poprzednich kolumn. Taka konstrukcja jest dostatecznie prosta i pozwala na zachowanie ortogonalności.

Rozważmy teraz model (19) i plan ułamkowy 2^{3-1} postaci (patrz (17)):

x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
1	1	1	1	1	1	1	1
1	−1	1	−1	−1	1	−1	1
1	1	−1	−1	−1	−1	1	1
1	−1	−1	1	1	−1	−1	1

(20)

Plan (20) został skonstruowany w sposób następujący: dla czynników x_1 i x_2 były wzięte wszystkie możliwe wartości, natomiast $x_3 = x_1x_2$, a wszystkie pozostale kolumny obliczone zostały na podstawie pierwszych.

Pytanie 3. Czy możemy tutaj oszacować w sposób nieobciążony wszystkie 8 nieznanych współczynników ? Oczywiście, nie, bo mamy tylko 4 doświadczenia. A co możemy oszacować stosując plan (20) ?

Zauważmy, że pierwsza kolumna jest to samo co ósma, druga równa się siódmej, trzecia – szóstej, czwarta – piątej. A więc stosując ten plan nie rozróżniamy: x_0 i $x_1x_2x_3$, x_1 i x_2x_3, x_2 i x_1x_3, x_3 i x_1x_2, tzn.

$$x_0 = x_{123}, \quad x_1 = x_{23}, \quad x_2 = x_1x_3, \quad x_3 = x_1x_2. \tag{21}$$

Oznacza to że możemy na podstawie tego planu oszacować w sposób nieobciążony tylko 4 współczynniki mieszane:

$$\theta_0 + \theta_{123}, \quad \theta_1 + \theta_{23}, \quad \theta_2 + \theta_{13}, \quad \theta_3 + \theta_{12}.$$

Jak łatwo zrozumieć, wszystkie relacje (21) możemy uzyskać mając tylko ostatnią, tzn. $x_3 = x_1x_2$. Nazywa się ona dla tego *relacja generująca*. Znając relacje generujące dla planu ułamkowego postaci 2^{m-p} (tych relacji jest dokładnie p) możemy odtworzyć ten plan.

Zadanie domowe. Odtworzyć plan ułamkowy typu 2^{3-1} mając relację generującą postaci $x_3 = -x_1x_2$ i podać współczynniki mieszane które w tym przypadku możemy oszacować w sposób nieobciążony.

Oprócz pojęcia relacji generującej jest jeszcze pojęcie *relacji określającej*. Jest to równość, w której po lewej stronie stoi iloczyn pewnych czynników, z plusem lub z minusem, a po prawej – jedynka. W przypadku planu (20) mamy jedyną relację określają- ca $x_1x_2x_3 = 1$, którą uzyskamy mnożąc obie części relacji generującej na x_3. W ogóle relacji określających jest co najmniej tyle samo ile generujących. Wszystkie relacje określające możemy uzyskać w sposób następujący: z każdej relacji generującej możemy uzyskać bezpośrednio jedną relacje określającą poprzez mnożenie obu części relacji generującej na odpowiedni czynnik. Dalej mnożąc już uzyskane relacje

określające same na siebie możemy uzyskać nowe (o ile nie będą się one powtarzać).

Relację określające są pożyteczne ponieważ pozwalają szybko odpowiedzieć na pytanie z jakimi współczynnikami będzie pomieszany wybrany współczynnik przy estymacji w sposób nieobciążony na podstawie zadanego planu ułamkowego. Rzeczywiście, np. dla znależenia współczynników z którymi będzie mieszany θ_1 trzeba: pomnożyć lewą i prawą stronę każdej relacji określającej na x_1; odpowiednie współczynniki odczytujemy z lewych stron otrzymanych równości biorąc efekty współdzialań odpowiednich czynników.

Przykład. Wypisać relacje określające i podać współczynniki z którymi będą mieszane efekty główne w przypadku planu ułamkowego typu 2^{5-2} i relacji generujących $x_4 = x_1x_2x_3,\ x_5 = x_1x_2$.

Relacje określające to: $x_1x_2x_3x_4 = 1,\ x_1x_2x_5 = 1,\ x_3x_4x_5 = 1$.

Przy korzystaniu z zadanego planu efekty główne będą mieszane tak:

$$\theta_1 + \theta_{25} + \theta_{234} + \theta_{1345}, \quad \theta_2 + \theta_{15} + \theta_{134} + \theta_{2345}, \quad \theta_3 + \theta_{45} + \theta_{124} + \theta_{1235},$$

$$\theta_4 + \theta_{35} + \theta_{123} + \theta_{1245}, \quad \theta_5 + \theta_{12} + \theta_{34} + \theta_{12345}.$$

Analogicznie postępują i w przypadku $m > 3$. W praktyce wybierając plan ułamkowy korzystają się ze wszystkich wiadomości na temat procesu badanego mając na celu wyodrębnienie tych efektów głównych i interakcji wpływ których na wyniki doświadczeń jest istotny. Przy tym robią tak że te wspólczynniki, które są istotne, mieszają ze wspólczynnikami (np. interakcjami wysokiego rzędu) które są nieistotne lub wręcz nieobecne.

10 Wybrane pytania analizy wariancji.

Analiza jednoczynnikowa

Rozważmy następujący problem. Niech mamy N populacji, przy czym i-ta populacja ma rozkład $N(\theta_i, \sigma^2)$. Niech mamy niezależne próby $(y_{i1}, \ldots, y_{in_i})$ objętości n_i z populacji i, $i = 1, \ldots, N$. Więc, możemy zapisać

$$y_{ij} = \theta_i + \xi_{ij},\ j = 1, \ldots, n_i;\ i = 1, \ldots, N, \tag{22}$$

gdzie $\{\xi_{ij}\}$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie $N(0, \sigma^2)$.

Ogólnie mówiąc można powiedzieć, że analiza wariancji jest podrozdziałem analizy regresji, mająca troche inne cele i metody. Zmienne niezależne tutaj też nazywają sie *czynnikami*. Analiza zależności sprowadza się do rozbicia wariancji modeli na składniki (patrz dalej). W pzrypadku rozważanym mówią, że mamy do czynienia z jednym czynnikiem, który ma N *poziomów* $\theta_1, \ldots, \theta_N$.

Wprowadzmy poznaczenia :

$$y = (y_{11}, \ldots, y_{1n_1}, \ldots, y_{N1}, \ldots, y_{Nn_N})^T, \quad \theta = (\theta_1, \ldots, \theta_N)^T,$$

$$\xi = (\xi_{11}, \ldots, \xi_{1n_1}, \ldots, \xi_{N1}, \ldots, \xi_{Nn_N})^T,$$

$$F = \begin{pmatrix} 1 & 0 & 0 & \ldots\ldots & 0 \\ \ldots & \ldots & \ldots & \ldots\ldots & \ldots \\ 1 & 0 & 0 & \ldots\ldots & 0 \\ 0 & 1 & 0 & \ldots\ldots & 0 \\ \ldots & \ldots & \ldots & \ldots\ldots & \ldots \\ 0 & 1 & 0 & \ldots\ldots & 0 \\ \ldots & \ldots & \ldots & \ldots\ldots & \ldots \\ 0 & 0 & 0 & \ldots\ldots & 1 \\ \ldots & \ldots & \ldots & \ldots\ldots & \ldots \\ 0 & 0 & 0 & \ldots\ldots & 1 \end{pmatrix}$$

- $n \times N$-macierz, $n = \sum_{i=1}^{N} n_i$, $r(F) = N$. Tym samym mamy klasyczny model regresji liniowej $(y, F\theta, \sigma^2 I_n)$. MNK estymatory dla parametrów $\{\theta_i\}$ zgodnie

z Twierdzeniem Gaussa-Markova mają postać

$$\hat{\theta}_i = (F^T F)^{-1} F^T y = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} = \bar{y}_i.$$

Dalej będziemy się zajmowali sprawdzaniem hipotezy H_0: $\theta_1 = \ldots = \theta_N$. Będzie nam potrzebne dalej pewne twierdzenie (poznaczenia jak w Rozdzialach 2 i 3).

Lemat 10.1 *Niech T jest $q \times k$ macierzą, $r(T) = q \leq k$. Wtedy w klasycznym modelu regresji liniowej $(y, F\theta, \sigma^2 I_n)$*

$$(\hat{\tau} - \tau_0)^T V^{-1} (\hat{\tau} - \tau_0) = R_1^2 - R_0^2,$$

gdzie $\hat{\tau} = T\hat{\theta}$, $V = \sigma^{-2} T Cov\ \hat{\theta} T^T$, $\hat{\theta}$ jest MNK estymatorem, $\tau_0 \in R^q$ jest zadane,

$$R_0^2 = \min_{\theta} (y - F\theta)^T (y - F\theta) = (y - F\hat{\theta})^T (y - F\hat{\theta}),$$

$$R_1^2 = \min_{\theta : T\theta = \tau_0} (y - F\theta)^T (y - F\theta).$$

Dowód. Zastosujmy metode Lagrange'a do szukania minimum $(y - F\theta)^T (y - F\theta)$ względem θ, $T\theta = \tau_0$. Funkcja Lagrange'a ma postać

$$L(\theta, \lambda) = (y - F\theta)^T (y - F\theta) + 2\lambda^T (T\theta - \tau_0).$$

Biorąc pochodną po θ_l, $l = 1, \ldots, N$ i przyrównując ją do zera otrzymamy równanie wektorowe $F^T F\theta + T^T \lambda = F^T y$. Dołączając do niego warunek $T\theta = \tau_0$ i uwzględniając, że $\hat{\theta} = (F^T F)^{-1} F^T y$, $V = T(F^T F)^{-1} T^T$, mamy rozwiązania

$$\left\{ \begin{array}{c} \theta = \hat{\theta} - (F^T F)^{-1} T^T \lambda \\ T\theta = \tau_0 \end{array} \right. \Longrightarrow \left\{ \begin{array}{c} \theta = \hat{\theta} - (F^T F)^{-1} T^T \lambda \\ T\hat{\theta} - T(F^T F)^{-1} T^T \lambda = \tau_0 \end{array} \right. \Longrightarrow$$

$$\left\{ \begin{array}{c} \theta_* = \hat{\theta} - (F^T F)^{-1} T^T V^{-1} (\hat{\tau} - \tau_0) \\ \lambda_* = V^{-1} (\hat{\tau} - \tau_0) \end{array} \right. . \qquad (23)$$

Podstawiając θ_* w R_1^2 otrzymujemy

$$R_1^2 = R_0^2 + (\hat{\tau} - \tau_0)^T V^{-1} (\hat{\tau} - \tau_0). \quad \square$$

Uwaga. Stwierdzenie Lematu 10.1 jest prawdziwe także w przypadku modeli regresji liniowej niepełnego rzędu, jeśli założyć, że macierz T dopuszcza przedstawienie $T = QF$ (patrz Rozdzial 3).

Zauważmy, że jeśli $\theta_1 = \ldots = \theta_N$, to mamy model regresji liniowej $(y, F_0\theta, \sigma^2 I_n)$ niepełnego rzędu z $n \times N$-macierzą

$$F_0 = \begin{pmatrix} 1 & 0 & \ldots\ldots & 0 \\ \ldots & \ldots & \ldots\ldots & \ldots \\ 1 & 0 & \ldots\ldots & 0 \end{pmatrix},$$

$r(F_0) = 1$, $\theta = \theta_1 \mathbf{1}$, $\mathbf{1} = (1, \ldots, 1)^T \in R^N$. Wtedy, jak łatwo zrozumieć, jedynym rozwiązaniem układu równań normalnych będzie

$$\widehat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} n_i \bar{y}_i = \frac{1}{n} \sum_{i,j} y_{ij} = \bar{y}. \tag{24}$$

Ten sam wynik można uzyskać z formuly (23) dla $(N-1) \times N$-macierzy

$$T = \begin{pmatrix} 1 & 0 & \ldots\ldots & 0 & -1 \\ 0 & 1 & \ldots\ldots & 0 & -1 \\ \ldots & \ldots & \ldots\ldots & \ldots & \ldots \\ 0 & 0 & \ldots\ldots & 1 & -1 \end{pmatrix}, \quad \tau_0 = 0.$$

Zadanie domowe. Uzyskać w ten sposób formule (24) z formuly (23) dla $N = 3$.

Zgodnie z wynikami Rozdzialów 2 i 3 (patrz Wniosek 2.1, Fakty 3.1, 3.2),

$$\frac{1}{\sigma^2} R_0^2 \sim \chi^2_{n-N}, \qquad \frac{1}{\sigma^2}(R_1^2 - R_0^2) \sim \chi^2_{N-1},$$

$$\frac{R_1^2 - R_0^2}{\sigma^2(N-1)} \Big/ \frac{R_0^2}{\sigma^2(n-N)} = \frac{(n-N)(R_1^2 - R_0^2)}{(N-1)R_0^2} \sim F(N-1, n-N).$$

Zauważmy, że

$$R_0^2 = (y - F\widehat{\theta})^T (y - F\widehat{\theta}) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2,$$

$$R_1^2 = (y - F\theta_*)^T (y - F\theta_*) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2,$$

$$R_1^2 - R_0^2 = \sum_{i=1}^{N} \sum_{j=1}^{n_i} [(y_{ij} - \bar{y})^2 - (y_{ij} - \bar{y}_i)^2] = \sum_{i=1}^{N} \sum_{j=1}^{n_i} (\bar{y}_i - \bar{y})(2y_{ij} - \bar{y}_i - \bar{y}) =$$

$$\sum_{i=1}^{N} (\bar{y}_i - \bar{y})(2n_i\bar{y}_i - n_i\bar{y}_i - n_i\bar{y}) = \sum_{i=1}^{N} n_i (\bar{y}_i - \bar{y})^2.$$

Widać, że całkowita wariancja R_1^2 w modelu (całkowity rozrzut względem wspólnej średniej) jest równa sumie dwoch niezależnych składników : R_0^2 – sumarny rozrzut wewnątrz każdej populacji i $(R_1^2 - R_0^2)$ – sumarny rozrzut między populacjami.

Często zamiast zapisu modelu w postaci (22) wykorzystują inny zapis. Mianowicie, poznaczymy przez $\mu = N^{-1}\sum_{i=1}^{N}\theta_i$ wspólną średnią. Wartość $\alpha_i = \theta_i - \mu$ nazwiemy *efektem* i-go poziomu czynnika. Zauważmy, że

$$\sum_{i=1}^{N}\alpha_i = 0. \tag{25}$$

Zamiast (22) możemy zapisać

$$y_{ij} = \mu + \alpha_i + \xi_{ij},\ j = 1, \ldots, n_i;\ i = 1, \ldots, N, \tag{26}$$

gdzie $\{\xi_{ij}\}$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie $N(0, \sigma^2)$. W tym przypadku, $\theta = (\mu, \alpha_1, \ldots, \alpha_N)^T$,

$$F = \begin{pmatrix} 1 & 1 & 0 & 0 & \ldots\ldots & 0 \\ \ldots & \ldots & \ldots & \ldots\ldots & \ldots & \\ 1 & 1 & 0 & 0 & \ldots\ldots & 0 \\ 1 & 0 & 1 & 0 & \ldots\ldots & 0 \\ \ldots & \ldots & \ldots & \ldots\ldots & \ldots & \\ 1 & 0 & 1 & 0 & \ldots\ldots & 0 \\ \ldots & \ldots & \ldots & \ldots\ldots & \ldots & \\ 1 & 0 & 0 & 0 & \ldots\ldots & 1 \\ \ldots & \ldots & \ldots & \ldots\ldots & \ldots & \\ 1 & 0 & 0 & 0 & \ldots\ldots & 1 \end{pmatrix}$$

- $n \times (N+1)$-macierz, $n = \sum_{i=1}^{N} n_i,\ r(F) = N$. Więc, mamy do czynienia z modelem niepełnego rzędu. Ale w rzeczywistości, nieznanych parametrów jest na jeden mniej z powodu (25). Zgodnie z Twierdzeniem 3.1, estymatorami dla μ i α_i są

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_i = \bar{y}, \quad \hat{\alpha}_i = \hat{\theta}_i - \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_i = \bar{y}_i - \bar{y}, \quad i \geq 1.$$

Zauważmy, że

$$R_0^2 = \sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij} - \hat{\mu} - \hat{\alpha}_i)^2, \quad R_1^2 = \sum_{i=1}^{N}\sum_{j=1}^{n_i}(y_{ij} - \hat{\mu})^2,$$

$$R_1^2 - R_0^2 = \sum_{i=1}^{N} n_i \hat{\alpha}_i^2.$$

Natomiast hipoteza $H_0 : \theta_1 = \ldots = \theta_N$ dla zapisu (22) oznacza hipoteze $H_0 : \alpha_1 = \ldots = \alpha_N = 0$ zapisu (26).

Jeśli zrezygnować z warunku (25), to już nie możemy wszystkie parametry oszacować w sposób jednoznaczny i nieobciążony (patrz Rozdzial 3). Przypomnimy, że w przypadku modeli niepełnego rzędu ważnym było pojęcie *estymowalności* funkcji liniowej od parametrów (patrz Rozdzial 3). Funkcja $t^T\theta$ jest estymowalna wtedy i tylko wtedy kiedy $t = F^T b$ dla pewnego $b \in R^n$. To oznacza, że estymowalne są tylko funkcje liniowe postaci

$$b^T F\theta = \sum_{i,j} b_{ij} E y_{ij} = \sum_{i,j} b_{ij}(\mu+\alpha_i) = \sum_{i=1}^{N} c_i(\mu+\alpha_i) = c_1\alpha_1+\cdots+c_N\alpha_N+\sum_{i=1}^{N} c_i\mu,$$

gdzie $c_i = \sum_j b_{ij}$. Stąd wynika, że np. $\sum_{i=1}^N c_i\alpha_i$ jest estymowalne wtedy i tylko wtedy kiedy $\{c_i\}$ są takie, że $\sum_{i=1}^N c_i = 0$. Np. $\alpha_i - \alpha_1$ dla każdego $i > 1$ jest estymowalna.

Analiza dwoczynnikowa

Niech mamy dwa czynniki : pierwszy ma N poziomów, drugi – M poziomów (np. rośliny i działki). Dla każdego poziomu i pierwszego czynnika i dla każdego poziomu j drugiego czynnika mamy n_{ij} obserwacji $(y_{ij1}, \ldots, y_{ijn_{ij}})$ z populacji o rozkładzie $N(\theta_{ij}, \sigma^2)$. Więc,

$$y_{ijl} = \theta_{ij} + \xi_{ijl},\ l = 1, \ldots, n_{ij};\ i = 1, \ldots, N;\ j = 1, \ldots, M, \qquad (27)$$

gdzie $\{\xi_{ijl}\}$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie $N(0, \sigma^2)$.

W praktyce często się korzystają z innego zapisu tego samego modelu. Poznaczymy : $\bar{\theta}_i^{(1)} = M^{-1}\sum_{j=1}^M \theta_{ij}$ – średnia wartość i-go poziomu pierwszego czynnika, $\bar{\theta}_j^{(2)} = N^{-1}\sum_{i=1}^N \theta_{ij}$ – średnia wartość j-go poziomu drugiego czynnika, $\mu = N^{-1}M^{-1}\sum_{i,j}\theta_{ij}$ – wspólna średnia. Zdefiniujmy efekt i-go poziomu pierwszego czynnika jako $\alpha_i = \bar{\theta}_i^{(1)} - \mu$, efekt j-go poziomu drugiego czynnika jako $\beta_j = \bar{\theta}_j^{(2)} - \mu$. Zauważmy, że $\sum_{i=1}^N \alpha_i = 0,\ \sum_{j=1}^M \beta_j = 0$. Efekt i-go poziomu pierwszego czynnika można też zapisać w postaci $\alpha_i = \bar{\theta}_i^{(1)} - \mu =$

$M^{-1}\sum_{j=1}^{M}(\theta_{ij}-\bar{\theta}_j^{(2)})$. Wartość

$$\gamma_{ij}=\theta_{ij}-\bar{\theta}_j^{(2)}-\alpha_i=\theta_{ij}-\bar{\theta}_j^{(2)}-\bar{\theta}_i^{(1)}+\mu \tag{28}$$

nazywa się *współdziałaniem* i-go poziomu pierwszego czynnika i j-go poziomu drugiego czynnika. Zaznaczmy, że $\sum_{j=1}^{M}\gamma_{ij}=0,\ \sum_{i=1}^{N}\gamma_{ij}=0$. Podsta- wiając $\bar{\theta}_j^{(2)}=\mu+\beta_j,\ \bar{\theta}_i^{(1)}=\mu+\alpha_i$ w (28), mamy $\theta_{ij}=\mu+\alpha_i+\beta_j+\gamma_{ij}$ i w końcu przychodzimy do zapisu

$$y_{ijl}=\mu+\alpha_i+\beta_j+\gamma_{ij}+\xi_{ijl},\ l=1,\ldots,n_{ij};\ i=1,\ldots,N;\ j=1,\ldots,M, \tag{29}$$

gdzie $\{\xi_{ijl}\}$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie $N(0,\sigma^2)$.

Rozpatrzmy dalej prostszy przypadek, kiedy $n_{ij}=K=const$. Wtedy MNK estymatory dla $\{\theta_{ij}\}$ z modelu (27) mają postać $\hat{\theta}_{ij}=K^{-1}\sum_{l=1}^{K}y_{ijl}=\bar{y}_{ij}$. Dalej korzystając z Twierdzenia 3.1 uzyskujemy NLNE dla nieznanych parametrów w (29) :

$$\hat{\mu}=\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\hat{\theta}_{ij}=\frac{1}{NMK}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{l=1}^{K}y_{ijl}=\bar{y},$$

$$\hat{\alpha}_i=\frac{1}{M}\sum_{j=1}^{M}\hat{\theta}_{ij}-\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\hat{\theta}_{ij}=\frac{1}{MK}\sum_{j=1}^{M}\sum_{l=1}^{K}y_{ijl}-\frac{1}{NMK}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{l=1}^{K}y_{ijl}=\bar{y}_i^{(1)}-\bar{y},$$

$$\hat{\beta}_j=\frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_{ij}-\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\hat{\theta}_{ij}=\frac{1}{NK}\sum_{i=1}^{N}\sum_{l=1}^{K}y_{ijl}-\frac{1}{NMK}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{l=1}^{K}y_{ijl}=\bar{y}_j^{(2)}-\bar{y},$$

$$\hat{\gamma}_{ij}=\hat{\theta}_{ij}-\frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_{ij}-\frac{1}{M}\sum_{j=1}^{M}\hat{\theta}_{ij}-\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\hat{\theta}_{ij}=$$

$$\frac{1}{K}\sum_{l=1}^{K}y_{ijl}-\frac{1}{MK}\sum_{j=1}^{M}\sum_{l=1}^{K}y_{ijl}-\frac{1}{NK}\sum_{i=1}^{N}\sum_{l=1}^{K}y_{ijl}+\frac{1}{NMK}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{l=1}^{K}y_{ijl}=$$

$$\bar{y}_{ij}-\bar{y}_i^{(1)}-\bar{y}_j^{(2)}+\bar{y}.$$

Załóżmy, że chcemy sprawdzić hipoteze $H_0:\ \alpha_1=\ldots=\alpha_N=0$. Analogicznie, jak w przypadku analizy jednoczynnikowej, można policzyć, że

$$R_0^2=\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{l=1}^{K}\left(y_{ijl}-\hat{\mu}-\hat{\alpha}_i-\hat{\beta}_j-\hat{\gamma}_{ij}\right)^2=\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{l=1}^{K}\left(y_{ijl}-\bar{y}_{ij}\right)^2$$

jest sumarnym rozrzutem wariancji wewnątrz każdej populacji,

$$R_1^2 = \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{l=1}^{K}\left(y_{ijl} - \widehat{\mu} - \widehat{\beta}_j - \widehat{\gamma}_{ij}\right)^2 = \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{l=1}^{K}(y_{ijl} - \bar{y}_{ij})^2 + MK\sum_{i=1}^{N}\left(\bar{y}_i^{(1)} - \bar{y}\right)^2$$

jest sumarnym rozrzutem wariancji między populacjami spowodowanym pierwszym czynnikiem,

$$R_1^2 - R_0^2 = MK\sum_{i=1}^{N}\hat{\alpha}_i^2 = MK\sum_{i=1}^{N}\left(\bar{y}_i^{(1)} - \bar{y}\right)^2.$$

Teraz

$$\frac{1}{\sigma^2}R_0^2 \sim \chi^2_{NM(K-1)}, \qquad \frac{1}{\sigma^2}(R_1^2 - R_0^2) \sim \chi^2_{N-1},$$

$$\frac{R_1^2 - R_0^2}{\sigma^2(N-1)} \Big/ \frac{R_0^2}{\sigma^2 NM(K-1)} = \frac{NM(K-1)(R_1^2 - R_0^2)}{(N-1)R_0^2} \sim F(N-1, NM(K-1)).$$

Analogicznie można wypisać formuly dla sprawdzania hipotezy $H_0 : \beta_1 = \ldots = \beta_M = 0$ lub hipotezy $H_0 : \gamma_{ij} = 0,\ i = 1, \ldots, N, j = 1, \ldots, M.$

11 Pytania do teoretycznej części egzaminu

1. Problem znależenia zależności pomiędzy zmiennymi i klasyczny model regresji liniowej.

2. Problem estymacji nieznanych parametrów w modelu regresji liniowej i metoda najmniejszych kwadratów.

3. Sens metody najmniejszych kwadratów jako minimalizacji pewnej odległości w przestrzeni R^n. Formulowanie Twierdzenia Gaussa-Markova. Macierz rzutu ortogonalnego na podprzestrzeń.

4. Twierdzenie Gaussa-Markova (dowód) i Wniosek z niego.

5. Pojęcie nieobciążoności estymatora. Nieobciążoność statystyki $s^2 = \frac{1}{n-k}\varepsilon^T\varepsilon$ jako estymatora dla σ^2.

6. Definicja ENW i Twierdzenie o ENW dla θ i σ^2. Obciążoność. ENW dla θ i σ^2 w przypadku $\eta(x,\theta) = \theta$.

7. Procedura sprawdzania prostej hipotezy $H_0 : \theta_i = a$ w ramach modelu klasycznego (σ^2 jest znane i nieznane).

8. Rozkład statystyki $\varepsilon^T\varepsilon$.

9. Rozkład statystyki testowej $\frac{\widehat{\theta}_i - a}{s\sqrt{((F^TF)^{-1})_{ii}}}$.

10. Estymatory współczynników regresji dla modelu pełnego rzędu w przypadku $Cov\ y = \sigma^2 W,\ W > 0$.

11. Estymatory MNK współczynników regresji dla modelu niepełnego rzędu.

12. Pojęcie estymowalności funkcji liniowej od współczynników regresji. Twierdzenie o NLNE dla $t^T\theta$.

13. Procedura sprawdzania prostej hipotezy $H_0 : T\theta = \tau_0$ w ramach modelu regresji liniowej.

14. Rozkład statystyki testowej $\frac{(n-p)(\widetilde{\tau}-\tau_0)^T V^{-1}(\widetilde{\tau}-\tau_0)}{q\varepsilon^T W^{-1}\varepsilon}$.

15. Problem planowania doświadczeń i pojęcie planu eksperymentu.

16. Plany dyskretne. Wypukłość zbioru planów dyskretnych uogólnionych.

17. Plany ciągłe. Macierz informacji i macierz kowariancji.

18. Własności macierzy informacji.

19. Kryteria optymalności. Porównywanie macierzy informacji pomiędzy sobą "w całości".

20. Pojęcie planu Ψ-optymalnego i podstawowe własności kryteriów optymalności. Lemat o funkcjach wypukłych.

21. Kryterium D-optymalności.

22. Kryterium A-optymalności.

23. Definicja najczęściej stosowanych kryteriów optymalności (D-, A-, L-, E-, C-, G-) i ich sens statystyczny.

24. Pojęcie różniczkowalności funkcji $\Psi(M)$ po macierzy M i macierzy $M(t)$ po parametru t. Przykłady liczenia pochodnych.

25. Formulowanie Twierdzenia równoważności. Formulowanie i dowód pomocniczego Lematu.

26. Formulowanie i dowód Twierdzenia równoważności (bez dowodu pomocniczego Lematu).

27. Formulowanie i dowód Twierdzenia równoważności Kiefera-Wolfowits'a (dla kryterium D-optymalności). Formulowanie Twierdzenia równoważności dla kryterium A-optymalności.

28. Przykłady używania Twierdzeń równoważności.

29. D-optymalne plany dyskretne uogólnione dla regresji wielomianowej (bez dowodu optymalności wskazanych punktów nośnika).

30. Znależenie optymalnych punktów nośnika D-optymalnego planu dyskretnego uogólnionego dla regresji wielomianowej.

31. Znależenie D-optymalnego planu dyskretnego uogólnionego dla regresji wielomianowej (szkić). Zastosowanie do przypadku $\eta(x,\theta) = \theta_1 + \theta_2 x^2 + \theta_3 x^2 + \theta_4 x^3,\ X = [-1,1]$.

32. O metodach numerycznych konstrukcji planów optymalnych. Główna idea Algorytmu 1.

33. Algorytm 1. Lemat pozwalający na sformulowanie wygodnego warunku zakonczenia pracy algorytmu.

34. Algorytm 2.

35. Konstrukcja optymalnych planów dyskretnych ścisłych. Procedura "zaokrąglenia" wag. Lemat o szacowaniu jakości przybliżenia.

36. Algorytm 3.

37. Doświadczenia czynnikowe, plany czynnikowe pełne i ułamkowe.

38. Plany typu 2^3 i 2^{3-1} a szacowanie nieznanych współczynników.

39. Relację generujące. Konstrukcja planów ułamkowych typu 2^{m-p}.

40. Mieszanie efektów głównych i efektów współdzialań. Relację określające.

41. Analiza wariancji jednoczynnikowa. NLNE nieznanych wartości oczekiwanych populacji.

42. Lemat o statystykie $(\widehat{\tau}-\tau_0)^T V^{-1}(\widehat{\tau}-\tau_0)$ w ramach modelu klasycznego regresji liniowej.

43. Wartości $R_0^2,\ R_1^2,\ R_1^2 - R_0^2$ w przypadku sprawdzania hipotezy $H_0 : \theta_1 = \ldots = \theta_N$. Ich rozkłady i rozkład statystyki testowej. Calkowita wariancja modelu i jej składniki.

44. Pojęcie efektu i-go poziomu czynnika i inny zapis modelu. NLNE dla μ i $\alpha_i,\ i \geq 1$. Estymowalność funkcji liniowych od parametrów.

45. Analiza wariancji dwoczynnikowa. Dwa zapisy modelu.

46. NLNE nieznanych parametrów dla obu zapisów modelu przy $n_{ij} = K$. Wartości $R_0^2,\ R_1^2,\ R_1^2 - R_0^2$ w przypadku sprawdzania hipotezy $H_0 : \alpha_1 = \ldots = \alpha_N = 0$. Ich rozkłady i rozkład statystyki testowej.

12 Zadania do praktycznej części egzaminu

1. Dopasować prostą do punktów $(-1/2,-1),\ (0,0),\ (1,1)$ za pomocą MNK.

2. Dopasować prostą do punktów $(-1/2,-1),\ (0,0),\ (1,1/2),\ (3/2,3/2)$ za pomocą MNK.

3. Znaleźć estymatory MNK współczynników regresji $\eta(x,\theta)=\theta_1+\theta_2 x$ dla $n=3,\ x_1=0,\ x_2=1,\ x_3=2$.

4. Znaleźć estymatory MNK współczynników regresji $\eta(x,\theta)=\theta_1+\theta_2 x+\theta_3 x^2$ dla $n=3,\ x_1=-1,\ x_2=0,\ x_3=1$.

5. Policzyć ENW dla σ^2 w przypadku $\eta(x,\theta)=\theta_1+\theta_2 x$ dla $n=3$ i punktów (x_i,y_i) postaci $(-1,0),\ (0,1),\ (1,3)$.

6. Policzyć ENW dla σ^2 w przypadku $\eta(x,\theta)=\theta_1+\theta_2 x$ dla $n=4$ i punktów (x_i,y_i) postaci $(-1/2,-1),\ (0,0),\ (1,1/2),\ (3/2,3/2)$.

7. Sprawdzić prostą hipoteze $\theta_2=1$ w przypadku nieznanego σ^2 i $\eta(x,\theta)=\theta_1+\theta_2 x$ dla $n=3$ i punktów (x_i,y_i) postaci $(-1/2,-1),\ (0,0),\ (1,1)$ $(1-\alpha=0.95)$.

8. Sprawdzić prostą hipoteze $\theta_1=0$ w przypadku nieznanego σ^2 i $\eta(x,\theta)=\theta_1+\theta_2 x$ dla $n=4$ i punktów (x_i,y_i) postaci $(-1/2,-1),\ (0,0),\ (1,1/2),\ (3/2,3/2)$ $(1-\alpha=0.95)$.

9. Policzyć estymatory zwykłe i uogólnione MNK dla wspólczynników regresji $\eta(x,\theta)=\theta_1+\theta_2 x$ dla $n=3$ i $x_1=0,\ x_2=1,\ x_3=2$,

$$Cov\ y=\sigma^2\begin{pmatrix}1&0&0\\0&2&0\\0&0&1\end{pmatrix}.$$

10. Policzyć estymatory zwykłe i uogólnione MNK dla wspólczynników regresji $\eta(x,\theta)=\theta_1+\theta_2 x$ dla $n=3$ i $x_1=-1,\ x_2=0,\ x_3=1$,

$$Cov\ y=\sigma^2\begin{pmatrix}1&0&0\\0&2&0\\0&0&1\end{pmatrix}.$$

11. Znaleźć wszystkie rozwiązania układu równań normalnych dla modelu regresji liniowej $(y,F\theta,\sigma^2 I_3)$ niepełnego rzędu z

$$F=\begin{pmatrix}1&-1&0\\0&1&-1\\-1&0&1\end{pmatrix}.$$

12. Znaleźć wszystkie rozwiązania układu równań normalnych dla modelu regresji liniowej $(y, F\theta, \sigma^2 I_3)$ niepełnego rzędu z

$$F = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

13. Sprawdzić, czy funkcja liniowa $\tau = \theta_1 - \theta_2$ jest estymowalna dla modelu regresji liniowej $(y, F\theta, \sigma^2 I_3)$ z

$$F = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix}.$$

14. Sprawdzić, czy funkcja liniowa $\tau = \theta_1 + \theta_2$ jest estymowalna dla modelu regresji liniowej $(y, F\theta, \sigma^2 I_3)$ z

$$F = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

15. Znaleźć NLNE dla $\tau = \theta_1 - \theta_2$ w modelu regresji liniowej $(y, F\theta, \sigma^2 I_3)$ z

$$F = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix}.$$

16. Znaleźć NLNE dla $\tau = \theta_1 - \theta_2 + \theta_3$ w modelu regresji liniowej $(y, F\theta, \sigma^2 I_3)$ z

$$F = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

17. Sprawdzić prostą hipoteze liniową $\theta_1 + \theta_2 = 1$ w przypadku $\eta(x, \theta) = \theta_1 + \theta_2 x$ dla $n = 4$ i punktów (x_i, y_i) postaci $(-1/2, -1)$, $(0, 0)$, $(1, 1/2)$, $(3/2, 3/2)$ $(1 - \alpha = 0.95)$.

18. Sprawdzić prostą hipoteze liniową $\theta_1 - \theta_2 = 1$ w modelu regresji liniowej $(y, F\theta, \sigma^2 I_3)$ z $y = (0, 1, 2)^T$,

$$F = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix} \quad (1 - \alpha = 0.95).$$

19. Policzyć macierze informacji $M(\xi_1)$ i $M(\xi_2)$ dla planów

$$\xi_1 = \left\{ \begin{array}{ccc} 0 & 1 & 2 \\ 1/3 & 1/3 & 1/3 \end{array} \right\}, \qquad \xi_2 = \left\{ \begin{array}{cc} -2 & 2 \\ 1/3 & 2/3 \end{array} \right\}$$

w przypadku funkcji regresji $\eta(x, \theta) = \theta_1 + \theta_2 x$. Czy można powiedzieć, że $M(\xi_2) \geq M(\xi_1)$ lub $M(\xi_1) \geq M(\xi_2)$?

20. Policzyć macierze informacji $M(\xi_1)$ i $M(\xi_2)$ dla planów

$$\xi_1 = \left\{ \begin{array}{ccc} -1 & 1/2 & 1 \\ 1/3 & 1/3 & 1/3 \end{array} \right\}, \qquad \xi_2 = \left\{ \begin{array}{ccc} -1 & -3/4 & 1 \\ 1/4 & 1/2 & 1/4 \end{array} \right\}$$

w przypadku funkcji regresji $\eta(x, \theta) = \theta_1 + \theta_2 x$. Czy można powiedzieć, że $M(\xi_2) \geq M(\xi_1)$ lub $M(\xi_1) \geq M(\xi_2)$?

21. Znaleźć kombinację wypukłą z $\alpha = 2/3$ dla dwoch planów z zadania 19.

22. Znaleźć kombinację wypukłą z $\alpha = 1/4$ dla dwoch planów z zadania 20.

23. Niech kryterium optymalności jest zdefiniowany wzorem $\Psi(M) = det\ AM^{-1}A^T$, gdzie A jest $m \times k$-macierzą, $m \leq k,\ r(A) = m$. Czy dla niego są spełnione własności jednorodności i monotoniczności ?

24. Niech kryterium optymalności jest zdefiniowany wzorem $\Psi(M) = tr\ LM^{-1}$, gdzie L jest $k \times k$-macierzą, $r(L) = k$. Czy dla niego są spełnione własności jednorodności i monotoniczności ?

25. Policzyć A-optymalny i E-optymalny dyskretne uogólnione plany dla przypadku $\eta(x, \theta) = \theta_1 + \theta_2 x,\ X = [-1, 1]$.

26. Policzyć L-optymalny dyskretny uogólniony plan dla przypadku $\eta(x, \theta) = \theta_1 + \theta_2 x,\ X = [-1, 1]$,

$$L = \left(\begin{array}{cc} 1 & 1 \\ 1 & 2 \end{array} \right).$$

27. Policzyć C-optymalny dyskretny uogólniony plan dla przypadku $\eta(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2,\ X = [-1, 1],\ c = (1, 2, 3)^T$.

28. Policzyć G-optymalny dyskretny uogólniony plan dla przypadku $\eta(x, \theta) = \theta_1 + \theta_2 x,\ X = [-1, 1]$.

29. Policzyć G-optymalny dyskretny uogólniony plan dla przypadku $\eta(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2,\ X = [-1, 1]$.

30. Porównać plany

$$\xi_1 = \left\{ \begin{array}{cc} -1 & 1 \\ 1/2 & 1/2 \end{array} \right\}, \qquad \xi_2 = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{array} \right\}$$

według kryteriumów A-, E-optymalności w przypadku $\eta(x,\theta)=\theta_1+\theta_2 x,\ X=[-1,1]$.

31. Porównać plany

$$\xi_1=\left\{\begin{array}{ccc}-1 & 0 & 1\\ 1/4 & 1/2 & 1/4\end{array}\right\},\qquad \xi_2=\left\{\begin{array}{ccc}-1 & 0 & 1\\ 1/3 & 1/3 & 1/3\end{array}\right\}$$

według kryteriumów G-, C-optymalności w przypadku $\eta(x,\theta)=\theta_1+\theta_2 x,\ X=[-1,1],\ c=(2,1)^T$.

32. Spośród planów dyskretnych uogólnionych postaci

$$\left\{\begin{array}{cc}-1 & 1\\ p & 1-p\end{array}\right\},\quad 0<p<1,$$

znaleźć najlepszy według kryterium E-optymalności dla przypadku $\eta(x,\theta)=\theta_1+\theta_2 x,\ X=[-1,1]$.

33. Spośród planów dyskretnych postaci

$$\left\{\begin{array}{ccc}-1 & 0 & x\\ 1/3 & 1/3 & 1/3\end{array}\right\},\quad -1\le x\le 1,$$

znaleźć najlepszy według kryterium A-optymalności dla przypadku $\eta(x,\theta)=\theta_1+\theta_2 x,\ X=[-1,1],\ n=3$.

34. Zrealizować jeden krok Algorytmu 1 dla kryterium A-optymalności w przypadku $\eta(x,\theta)=\theta_1+\theta_2 x,\ X=[-1,1]$ i planu początkowego

$$\xi_0=\left\{\begin{array}{cc}-1/2 & 1/2\\ 1/2 & 1/2\end{array}\right\}.$$

35. Zrealizować jeden krok Algorytmu 2 dla kryterium A-optymalności w przypadku $\eta(x,\theta)=\theta_1+\theta_2 x,\ X=[-1,1]$ i planu początkowego

$$\xi_0=\left\{\begin{array}{cc}-1/2 & 1/2\\ 1/2 & 1/2\end{array}\right\}.$$

36. Porównać Algorytmy 1 i 2 dla kryterium D-optymalności w przypadku $\eta(x,\theta)=\theta_1+\theta_2 x,\ X=[-1,1]$ i planu początkowego

$$\xi_0=\left\{\begin{array}{ccc}-1/2 & 0 & 1/2\\ 1/4 & 1/2 & 1/4\end{array}\right\}.$$

37. Zastosować procedure "zaokrąglenia" wag do planu L-optymalnego postaci

$$\xi^* = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 2.5 - \sqrt{5} & 2\sqrt{5} - 4 & 2.5 - \sqrt{5} \end{array} \right\}$$

w przypadku $n = 100$ i oszacować jakość przybliżenia, gdzie

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix},$$

$\eta(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2,\ X = [-1, 1]$.

38. Zrealizować jeden krok Algorytmu 3 dla kryterium D-optymalności w przypadku $\eta(x, \theta) = \theta_1 + \theta_2 x,\ X = [-1, 1]$ i planu początkowego

$$\xi_0^{(3)} = \left\{ \begin{array}{ccc} -1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{array} \right\}.$$

39. Podać pełny plan czynnikowy typu 2^4 i przykład planu ulamkowego ortogonalnego postaci 2^{4-2}.

40. Wypisać relacje określające i podać współczynniki z którymi będą mieszane efekty główne w przypadku planu ułamkowego typu 2^{4-1} i relacji generującej $x_3 = -x_1 x_2$.

41. Wypisać relacje określające i podać współczynniki z którymi będą mieszane efekty główne w przypadku planu ułamkowego typu 2^{4-2} i relacji generujących $x_3 = x_1 x_2,\ x_4 = -x_1 x_2$.

42. Analiza wariancji jednoczynnikowa : $N = 2,\ n_1 = 4,\ n_2 = 3,\ \sigma = 1,\ y_{11} = 0.51,\ y_{12} = -0.33,\ y_{13} = -0.76,\ y_{14} = 0.10,\ y_{21} = 1.15,\ y_{22} = -0.13,\ y_{23} = -0.36$. Wypisać NLNE dla nieznanych parametrów θ_1 i θ_2 i sprawdzić hipotezę $\theta_1 = \theta_2$.

43. Analiza wariancji jednoczynnikowa : $N = 2,\ n_1 = n_2 = 4,\ \sigma = 1,\ y_{11} = -0.12,\ y_{12} = -0.68,\ y_{13} = -1.13,\ y_{14} = 0.21,\ y_{21} = 0.54,\ y_{22} = -0.38,\ y_{23} = 0.22,\ y_{24} = -0.10$. Wypisać NLNE dla nieznanych parametrów $\mu,\ \alpha_1,\ \alpha_2$ i sprawdzić hipotezę $\alpha_1 = \alpha_2 = 0$.

44. Analiza wariancji dwoczynnikowa : $N = 2,\ M = 2,\ n_{ij} = K = 4,\ \sigma = 1,\ y_{111} = 0.68,\ y_{112} = 1.33,\ y_{113} = -0.11,\ y_{114} = 0.30,\ y_{121} = 0.13,\ y_{122} = -0.36,\ y_{123} = -0.74,\ y_{124} = 0.21,\ y_{211} = -0.38,\ y_{212} = 1.04,\ y_{213} = -0.27,\ y_{214} = 0.05,\ y_{221} = -1.44,\ y_{222} = -0.36,\ y_{223} = 0.11,\ y_{224} = -0.35$. Wypisać NLNE dla nieznanych parametrów $\mu,\ \alpha_1,\ \alpha_2$ i sprawdzić hipotezę $\alpha_1 = \alpha_2 = 0$.

45. Analiza wariancji dwoczynnikowa : $N = 2,\ M = 2,\ n_{ij} = K = 4,\ \sigma = 1,\ y_{111} = 0.68,\ y_{112} = 1.33,\ y_{113} = -0.11,\ y_{114} = 0.30,\ y_{121} = 0.13,\ y_{122} = -0.36,\ y_{123} = -0.74,\ y_{124} = 0.21,\ y_{211} = -0.38,\ y_{212} = 1.04,\ y_{213} = -0.27,\ y_{214} = 0.05,\ y_{221} = -1.44,\ y_{222} = -0.36,\ y_{223} = 0.11,\ y_{224} = -0.35.$ Wypisać NLNE dla nieznanych parametrów $\mu,\ \beta_1,\ \beta_2$ i sprawdzić hipotezę $\beta_1 = \beta_2 = 0.$