Modele liniowe - teoria

dr Janusz Przewocki

Uniwersytet Gdański

Estymatory MLE

Metoda MLE (ang. Maximum Likelihood Estimation) polega na znalezieniu takiej wartości parametrów, przy której nasze dane są najbardziej prawdopodobne.

Przypomnijmy, że nasze dane modelujemy rozkładem:

$$\vec{y} \sim N(\vec{\theta}, \sigma^2 I)$$

W związku z tym, szukamy takiego wektora $\vec{\theta}$, dla którego wartość gęstości prawdopodobieństwa jest maksymalna.

Estymatory MLE

Metoda MLE (ang. Maximum Likelihood Estimation) polega na znalezieniu takiej wartości parametrów, przy której nasze dane są najbardziej prawdopodobne.
Przypomnijmy, że nasze dane modelujemy rozkładem:

$$\vec{y} \sim N(\vec{\theta}, \sigma^2 I)$$

W związku z tym, szukamy takiego wektora $\vec{\theta}$, dla którego wartość gęstości prawdopodobieństwa jest maksymalna.

Funkcja wiarygodności

Ponieważ modelujemy dane wielowymiarowym rozkładem normalnym, funkcja wiarygodności przyjmuje formę:

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

Maksimum funkcji wiarygodności

Posługując się metodą mnożników Lagrange, oblicz że wartość $\log L$ maksymalizuje się, gdy spełnione są równania

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\vec{y} - \vec{\theta})^T(\vec{y} - \vec{\theta}) = 0$$

$$\frac{1}{\sigma^2}(\vec{y} - \vec{\theta}) - (I_n - H)\vec{\lambda} = 0$$

$$(I_n - H)\vec{\theta} = 0 \quad (1)$$

Maksimum funkcji wiarygodności

Z pierwszego równania wyznaczamy

$$\sigma^2 = \frac{1}{n}\|\vec{y} - \vec{\theta}\|^2$$

Z drugiego i trzeciego (mnożąc drugie równanie przez H) otrzymujemy

$$\vec{\theta} = H\vec{y}$$

Problem. Udowodnij powyższe stwierdzenia.

Estymator MLE w regresji

Widzimy zatem

- Najlepszym estymatorem $\vec{\theta}$ w sensie MLE jest rzut ortogonalny na przestrzeń Ω
- Najlepszym estymatorem σ^2 w sensie MLE jest $\hat{\sigma}^2 = \frac{1}{n}\|\vec{y} - \hat{\vec{\theta}}\|^2$
- Estymator $\hat{\sigma}^2$ jest obciążony! Ale jest on zgodny...
- ...i efektywny (w tym sensie, że dąży do ograniczenia Cramera-Rao przy $n \to \infty$)

Estymator MLE w regresji

Widzimy zatem

- Najlepszym estymatorem $\vec{\theta}$ w sensie MLE jest rzut ortogonalny na przestrzeń Ω
- Najlepszym estymatorem σ^2 w sensie MLE jest $\hat{\sigma}^2 = \frac{1}{n}\|\vec{y} - \hat{\vec{\theta}}\|^2$
- Estymator $\hat{\sigma}^2$ jest obciążony! Ale jest on zgodny...
- ...i efektywny (w tym sensie, że dąży do ograniczenia Cramera-Rao przy $n \to \infty$)

Estymator MLE w regresji

Widzimy zatem

- Najlepszym estymatorem $\vec{\theta}$ w sensie MLE jest rzut ortogonalny na przestrzeń Ω
- Najlepszym estymatorem σ^2 w sensie MLE jest $\hat{\sigma}^2 = \frac{1}{n}\|\vec{y} - \hat{\vec{\theta}}\|^2$
- Estymator $\hat{\sigma}^2$ jest obciążony! Ale jest on zgodny...
- ...i efektywny (w tym sensie, że dąży do ograniczenia Cramera-Rao przy $n \to \infty$)

Estymator MLE w regresji

Widzimy zatem

- Najlepszym estymatorem $\vec{\theta}$ w sensie MLE jest rzut ortogonalny na przestrzeń Ω
- Najlepszym estymatorem σ^2 w sensie MLE jest $\hat{\sigma}^2 = \frac{1}{n}\|\vec{y} - \hat{\vec{\theta}}\|^2$
- Estymator $\hat{\sigma}^2$ jest obciążony! Ale jest on zgodny...
- ...i efektywny (w tym sensie, że dąży do ograniczenia Cramera-Rao przy $n \to \infty$)

Problem

Problem. Oblicz wartość maksimum funkcji wiarygodności dla estymatorów $\hat{\theta}$ oraz $\hat{\sigma}^2$

Maksimum funkcji wiarygodności

Maksimum funkcji wiarygodności wyraża się wzorem:

$$L(\vec{\tilde{\theta}}, \hat{\sigma}) = (2\pi\hat{\sigma}^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}$$

Hipotezy liniowe

Niech $\omega \subset \Omega$ będzie podprzestrzenią liniową. Będziemy rozważać następujący zestaw hipotez

$$H_0 : \theta \in \omega$$

$$H_A : \theta \notin \omega$$

W naszych rozważaniach zakładamy, że $\vec{\mathbf{y}} \sim N(\theta, \sigma^2 I)$

Hipotezy liniowe (oznaczenia)

Literą A będziemy oznaczać rzut ortogonalny na przestrzeń Ω.
Natomiast litera B oznacza rzut ortogonalny na przestrzeń ω.
Mamy dwa estymatory odchylenia standardowego:

$$\hat{\sigma}^2 = \frac{1}{n}\|\vec{y} - A\vec{y}\|^2$$

$$\tilde{\sigma}^2 = \frac{1}{n}\|\vec{y} - B\vec{y}\|^2$$

Hipotezy liniowe (oznaczenia)

Literą A będziemy oznaczać rzut ortogonalny na przestrzeń Ω.
Natomiast litera B oznacza rzut ortogonalny na przestrzeń ω.
Mamy dwa estymatory odchylenia standardowego:

$$\hat{\sigma}^2 = \frac{1}{n}\|\vec{y} - A\vec{y}\|^2$$

$$\tilde{\sigma}^2 = \frac{1}{n}\|\vec{y} - B\vec{y}\|^2$$

Test ilorazu wiarygodności

Statystyką, której będziemy używać w celu badania hipotezy $H_0 : \theta \in \omega$ będzie *iloraz funkcji wiarygodności* (likelihood quotient test):

$$Q = \frac{\sup_{\theta' \in \omega} L(\theta', \sigma^2)}{\sup_{\theta' \in \Omega} L(\theta', \sigma^2)}$$

Test ilorazu wiarygodności

Uwzględniając maksimum funkcji wiarygodności, które obliczyliśmy wcześniej widzimy, że

$$\mathbf{Q} = \left(\frac{\hat{\sigma}^2}{\tilde{\sigma}^2}\right)^{\frac{n}{2}} = \left(\frac{\|\vec{\mathrm{y}} - A\vec{\mathrm{y}}\|}{\|\vec{\mathrm{y}} - B\vec{\mathrm{y}}\|}\right)^n$$

Test F

Tę statystykę przekształcamy do takiej postaci, która ma wygodniejszy rozkład

$$\mathbf{F} = (\mathbf{Q}^{-\frac{2}{n}} - 1)\frac{n-p}{p-q} = \frac{\vec{\mathbf{y}}^T(A-B)\vec{\mathbf{y}}}{\vec{\mathbf{y}}^T(I-A)\vec{\mathbf{y}}}\frac{n-p}{p-q},$$

gdzie $p = \dim(\Omega), q = \dim(\omega)$

Zadanie. Udowodnij powyższy wzór.

Zadanie. Udowodnij, że w liczniku statystyki F występuje różnica sum kwadratów wartości przewidywanych przez model A i przez model B. Czyli

$$\|A\vec{y}\|^2 = (A\vec{y})^T(A\vec{y}), \quad \|B\vec{y}\|^2 = (B\vec{y})^T(B\vec{y})$$

Test F

Tę statystykę przekształcamy do takiej postaci, która ma wygodniejszy rozkład

$$\mathbf{F} = (\mathbf{Q}^{-\frac{2}{n}} - 1)\frac{n-p}{p-q} = \frac{\vec{\mathbf{y}}^T(A-B)\vec{\mathbf{y}}}{\vec{\mathbf{y}}^T(I-A)\vec{\mathbf{y}}}\frac{n-p}{p-q},$$

gdzie $p = \dim(\Omega), q = \dim(\omega)$

Zadanie. Udowodnij powyższy wzór.

Zadanie. Udowodnij, że w liczniku statystyki F występuje różnica sum kwadratów wartości przewidywanych przez model A i przez model B. Czyli

$$\|A\vec{y}\|^2 = (A\vec{y})^T(A\vec{y}), \quad \|B\vec{y}\|^2 = (B\vec{y})^T(B\vec{y})$$

Test F

Tę statystykę przekształcamy do takiej postaci, która ma wygodniejszy rozkład

$$\mathbf{F} = (\mathbf{Q}^{-\frac{2}{n}} - 1)\frac{n-p}{p-q} = \frac{\vec{\mathbf{y}}^T(A-B)\vec{\mathbf{y}}}{\vec{\mathbf{y}}^T(I-A)\vec{\mathbf{y}}}\frac{n-p}{p-q},$$

gdzie $p = \dim(\Omega), q = \dim(\omega)$

Zadanie. Udowodnij powyższy wzór.

Zadanie. Udowodnij, że w liczniku statystyki F występuje różnica sum kwadratów wartości przewidywanych przez model A i przez model B. Czyli

$$\|A\vec{y}\|^2 = (A\vec{y})^T(A\vec{y}), \quad \|B\vec{y}\|^2 = (B\vec{y})^T(B\vec{y})$$

Twierdzenie Cochrana

Twierdzenie

(Cochran; Rasch: Theorem 4.6) Jeżeli $\vec{\mathbf{y}}$ *ma rozkład* $N(\theta, I_n)$, *wówczas dodatnio półokreślone formy* $\vec{\mathbf{y}}^T A_i \vec{\mathbf{y}}$ $(i = 1, 2, ...k)$ *rządu* n_i *są niezależne i mają rozkłady* $\chi^2(n_i, \lambda_i)$, *z parametrem niecentralności* $\lambda_i = (\theta)^T A_i(\theta)$ *wtedy i tylko wtedy, gdy spełnione są co najmniej dwa z poniższych warunków*

1. A_i *są idempotentne*
2. $\sum_{i=1}^{k} A_i$ *jest idempotentna*
3. $A_i A_j = 0$, *dla dowolnych* $i \neq j$.

Wniosek

Wniosek

Jeżeli $\vec{\mathbf{y}}$ *ma rozkład* $N(\theta, I_n)$ *oraz* $\vec{\mathbf{y}}^T\vec{\mathbf{y}} = \sum_{i=1}^{k} \vec{\mathbf{y}}^T A_i \vec{\mathbf{y}}$*, wówczas formy kwadratowe* $\vec{\mathbf{y}}^T A_i \vec{\mathbf{y}}$ *są niezależne i mają rozkład* $\chi^2(n_i, \lambda_i)$*, z parametrem niecentralności* $\lambda_i = (\theta)^T A_i(\theta)$ *wtedy i tylko wtedy, gdy jest spełniony co najmniej jeden z warunków:*

1. A_i *są idempotentne*
2. $\sum_{i=1}^{k} \mathrm{rank}(A_i) = \mathrm{rank}(\sum_{i=1}^{k} A_i) = n$,
3. $A_i A_j = 0$*, dla dowolnych* $i \neq j$.

Statystyka F

Twierdzenie

Niech $\vec{\mathbf{y}}$ ma rozkład $N(\theta, \sigma^2 I_n)$ oraz A i B są macierzami rzutu ortogonalnego, odpowiednio, na podprzestrzenie Ω i ω (gdzie $\mathrm{rank}(A) = p$ i $\mathrm{rank}(B) = q$). Wówczas **F** *ma rozkład $F(p-q, n-p, \lambda)$ z parametrem niecentralności $\lambda = \frac{1}{\sigma^2}\theta^T(A-B)\theta$.*

Dowód. Formy kwadratowe określone przez macierze $I-A$, $A-B$ oraz B spełniają założenia powyższego wniosku. Stąd widać, że $\frac{1}{\sigma^2}\vec{\mathbf{y}}^T(I-A)\vec{\mathbf{y}}$ ma rozkład $\chi^2(n-p)$, natomiast $\frac{1}{\sigma^2}\vec{\mathbf{y}}^T(A-B)\vec{\mathbf{y}}$ ma rozkład $\chi^2(p-q, \lambda)$, gdzie $\lambda = \frac{1}{\sigma^2}\theta^T(A-B)\theta$.

□

Tablica analizy wariancji

Składowa wariancji	Suma kwadratów SS	Stopnie swobody df	Błąd średniokwadratowy $MS = \frac{SS}{df}$	E(MS)	F
Całkowita	$\vec{\mathbf{y}}^T\vec{\mathbf{y}}$	n			$F = \frac{n-p}{p-q}\frac{\vec{\mathbf{y}}^T(A-B)\vec{\mathbf{y}}}{\vec{\mathbf{y}}^T(I-A)\vec{\mathbf{y}}}$
Hipoteza alternatywna $\theta \notin \omega$	$\vec{\mathbf{y}}^T(A-B)\vec{\mathbf{y}}$	$p-q$	$\frac{1}{p-q}\vec{\mathbf{y}}^T(A-B)\vec{\mathbf{y}}$	$\sigma^2\left(1+\frac{\lambda}{q}\right)$	
Reszty	$\vec{\mathbf{y}}^T(I-A)\vec{\mathbf{y}}$	$n-p$	$\frac{1}{n-p}\vec{\mathbf{y}}^T(I-A)\vec{\mathbf{y}}$	σ^2	
Hipoteza zerowa $\theta \in \omega$	$\vec{\mathbf{y}}^T B\vec{\mathbf{y}}$	q			

Ogólny wzór modelu jest analogiczny jak przy regresji:

$$\vec{\mathbf{y}} = X\vec{\beta} + \vec{\mathbf{e}}$$

Jednak tym razem X ma rząd $r < p$.

Jednoczynnikowa analiza wariancji

Załóżmy, że mamy k populacji, z każdej losujemy próbę n_j-elementową. Zmienną y w poszczególnych populacjach modelujemy:

$$y_{ij} = \mu + a_j + e_{ij}.$$

Zapiszmy wektor współczynników:

$$\beta^T = [\mu, a_1, ..., a_k]$$

Wówczas macierz X ma postać:

$$X = \left(1_n, \bigoplus_{j=1}^{k} 1_{n_j} \right)$$

gdzie $n = \sum_j n_j$.

Analiza wariancji - estymacja współczynników

Układ równań normalnych

$$X^T X\vec{\beta} = X^T \vec{y}$$

Niech $G = (X^T X)^\dagger$. Wówczas:

$$\vec{\beta} = GX^T \vec{y}$$

Analiza wariancji - estymacja współczynników (c.d.)

Zawsze możemy dobrać macierz B, tak aby

$$\operatorname{rank}\begin{pmatrix} X \\ B \end{pmatrix} = k + 1$$

oraz

$$B\beta = 0$$

Fakt. Macierz $(X^T X + B^T B)^{-1}$ jest pseudoodwrotna względem $X^T X$.

Rozważymy teraz przypadek $k = 2$, ponadto $n_1 = n_2 = m$.
Zadanie. Udowodnij, że

$$X = \begin{bmatrix} n & m & m \\ m & n & m \\ m & m & n \end{bmatrix}$$

Mamy następujący wybór macierzy B

$$B = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}$$

co daje $a_1 + a_2 = 0$
Zadanie. Udowodnij, że

$$X^T X = \begin{bmatrix} n & m & m \\ m & m & 0 \\ m & 0 & m \end{bmatrix}, \quad B^T B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$

Oraz

$$X^T X + B^T B = \begin{bmatrix} 2m & m & m \\ m & m+1 & 1 \\ m & 1 & m+1 \end{bmatrix}$$

Zadanie. Udowodnij, że

$$(X^T X + B^T B)^{-1} = \frac{1}{4m} \begin{bmatrix} m+2 & -m & -m \\ -m & m+2 & m-2 \\ -m & m-2 & m+2 \end{bmatrix}$$

Oraz

$$\begin{bmatrix} \hat{\mu} \\ \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = \begin{bmatrix} \bar{y}_{..} \\ \frac{1}{2}(\bar{y}_{1.} - \bar{y}_{2.}) \\ \frac{1}{2}(\bar{y}_{2.} - \bar{y}_{1.}) \end{bmatrix}$$

Inny wybór macierzy B

Rozważymy teraz przykład bardziej adekwatny do tego co dzieje sie w języku R. Tutaj *Intercept* oznacza parametr μ, natomiast jeden z poziomów czynnika jest *bazowy*, w tym sensie, że odpowiadający mu współczynnik a jest równy zero. Omówmy przykład, gdzie bazowy jest współczynnik poziomu 1. Wówczas

$$B = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

Zadanie. Udowodnij, że

$$(X^TX+B^TB)^{-1} = \frac{1}{m}\begin{bmatrix} m+1 & -m & -m-1 \\ -m & m & m \\ -m-1 & m & m+2 \end{bmatrix} \quad \text{oraz} \quad \begin{bmatrix} \hat{\mu} \\ \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = \begin{bmatrix} \bar{y}_{1.} \\ 0 \\ y_{2.} - y_{1.} \end{bmatrix}$$

Hipotezy estymowalne

- Różne wybory macierzy pseudoodwrotnej dają różne estymacje współczynników.
- W związku z tym, bardziej interesujące są estymacje $\mu + a_j$.
- Wartości, które nie zależą od wyboru B zwane są *funkcjami estymowalnymi*, a hipotezy na temat tych funkcji *hipotezami estymowalnymi*.

Hipotezy estymowalne

- Różne wybory macierzy pseudoodwrotnej dają różne estymacje współczynników.
- W związku z tym, bardziej interesujące są estymacje $\mu + a_j$.
- Wartości, które nie zależą od wyboru B zwane są *funkcjami estymowalnymi*, a hipotezy na temat tych funkcji *hipotezami estymowalnymi*.

Hipotezy estymowalne

- Różne wybory macierzy pseudoodwrotnej dają różne estymacje współczynników.
- W związku z tym, bardziej interesujące są estymacje $\mu + a_j$.
- Wartości, które nie zależą od wyboru B zwane są *funkcjami estymowalnymi*, a hipotezy na temat tych funkcji *hipotezami estymowalnymi*.

Hipotezy estymowalne

Własności funkcji estymowalnych

- Funkcje liniowe $\mathrm{E}\vec{\mathbf{y}}$ są estymowalne
- Kombinacje linowe funkcji estymowalnych są estymowalne
- Jeżeli $\vec{q}^T\vec{\beta}$ jest estymowalna, to jej BLUE jest $\vec{q}^T\hat{\vec{\beta}}$

Dwuczynnikowa analiza wariancji

W dwuczynnikowej analizie mamy dwa czynniki A i B o liczbie poziomów odpowiednio k_a i k_b. Równanie można zapisać

$$\mathbf{y}_{ijk} = \eta_{ij} + \mathbf{e}_{ijk} \qquad \text{dla } i = 1, ..., k_a, \quad j = 1, ..., k_b, \quad k = 1, ..., n_{ij}$$

Współczynniki η_{ij} możemy zapisać

$$\eta_{ij} = \mu + a_i + b_j + (a, b)_{i,j}$$

gdzie a_i jest głównym efektem i-tego poziomu czynnika A, współczynnik b_j jest głównym efektem j-tego poziomu czynnika B, natomiast $(a, b)_{ij}$ jest efektem interakcji odpowiednich poziomów.

W regresji nieliniowej $\vec{\theta}$ jest funkcją $\vec{x}$ i $\vec{\beta}$, która jest nieliniowa ze względu na $\vec{\beta}$. A zatem

$$\vec{\mathbf{y}} = f(\vec{x}^{(1)}, \vec{x}^{(2)}, ..., \vec{x}^{(k)}, \vec{\beta}) + \vec{\mathbf{e}}$$

Regresja nieliniowa

Przykład. Rozważymy następującą funkcję:

$$f(x, \vec{\beta}) = \beta_1 e^{\beta_2 x}$$

Pytanie. Jak model nieliniowy oparty na powyższej funkcji różni się od modelu:

$$\log y = \beta_1 + \beta_2 x + e$$

gdzie $e \sim N(0, \sigma^2)$.

Przykład. Rozważymy następującą funkcję:

$$f(x, \vec{\beta}) = \beta_1 e^{\beta_2 x}$$

Pytanie. Jak model nieliniowy oparty na powyższej funkcji różni się od modelu:

$$\log y = \beta_1 + \beta_2 x + e$$

gdzie $e \sim N(0, \sigma^2)$.

Uogólnione modele liniowe

- Uogólnione modele liniowe przypominają regresję nieliniową w tym sensie, iż wartość oczekiwana modelowana jest funkcją nieliniową.
- Uogólnione modele są na tyle uniwersalne by dopuszczać różne struktury błędów.

W poniższej tabeli przedstawiono przykładowe zastosowania modeli uogólnionych

zmienna objaśniana	**rodzaj modelu**	**struktura błędu**
binarna	regresja logistyczna	rozkład Bernoulliego
proporcja	regresja logistyczna	rozkład dwumianowy
liczność	regresja log-liniowa	rozkład Poissona

Uogólnione modele liniowe

- Uogólnione modele liniowe przypominają regresję nieliniową w tym sensie, iż wartość oczekiwana modelowana jest funkcją nieliniową.
- Uogólnione modele są na tyle uniwersalne by dopuszczać różne struktury błędów.

W poniższej tabeli przedstawiono przykładowe zastosowania modeli uogólnionych

zmienna objaśniana	rodzaj modelu	struktura błędu
binarna	regresja logistyczna	rozkład Bernoulliego
proporcja	regresja logistyczna	rozkład dwumianowy
liczność	regresja log-liniowa	rozkład Poissona

- Uogólnione modele liniowe przypominają regresję nieliniową w tym sensie, iż wartość oczekiwana modelowana jest funkcją nieliniową.
- Uogólnione modele są na tyle uniwersalne by dopuszczać różne struktury błędów.

W poniższej tabeli przedstawiono przykładowe zastosowania modeli uogólnionych

zmienna objaśniana	**rodzaj modelu**	**struktura błędu**
binarna	regresja logistyczna	rozkład Bernoulliego
proporcja	regresja logistyczna	rozkład dwumianowy
liczność	regresja log-liniowa	rozkład Poissona

Regresja logistyczna

W uogólnionym modelu liniowym predyktor liniowy i wartość oczekiwana powiązane są tzw. funkcją wiążącą

$$\sum_i \beta_i x_i = g(\mathrm{E}(\mathbf{y}|x_1, x_2, ...x_k))$$

W przypadku regresji logistycznej mamy:

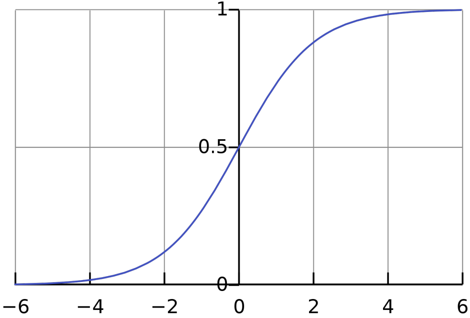
$$g(\theta) = \log \frac{\theta}{1-\theta}$$

Zadanie. Udowodnij, że funkcją odwrotną do

$$g(\theta) = \log \frac{\theta}{1-\theta}$$

Jest funkcja logistyczna

$$f(x) = \frac{e^x}{e^x + 1}$$



Regresja logistyczna

- Z typowym zastosowaniem regresji logistycznej mamy do czynienia, gdy zmienna y jest proporcją.
- Pamiętajmy, że standardowo liczbę sukcesów modelujemy *rozkładem dwumianowym*.
- Próbujemy estymować średnią wartość $\theta = \mathrm{E}(\mathbf{y}|x_1, x_2, ...x_k)$ zmiennej y w zależności od wartości predyktorów.

Trzeba wziąć pod uwagę, iż wariancja w przypadku rozkładu dwumianowego jest uzależniona od średniej:

$$\mathrm{var}(\mathbf{y_i}|x_1, x_2, ...x_k) = \theta(1-\theta)/m_i$$

gdzie m_i oznacza liczbę prób.

Regresja logistyczna

- Z typowym zastosowaniem regresji logistycznej mamy do czynienia, gdy zmienna y jest proporcją.
- Pamiętajmy, że standardowo liczbę sukcesów modelujemy *rozkładem dwumianowym*.
- Próbujemy estymować średnią wartość $\theta = \mathrm{E}(\mathbf{y}|x_1, x_2, ...x_k)$ zmiennej y w zależności od wartości predyktorów.

Trzeba wziąć pod uwagę, iż wariancja w przypadku rozkładu dwumianowego jest uzależniona od średniej:

$$\mathrm{var}(\mathbf{y_i}|x_1, x_2, ...x_k) = \theta(1-\theta)/m_i$$

gdzie m_i oznacza liczbę prób.

- Z typowym zastosowaniem regresji logistycznej mamy do czynienia, gdy zmienna y jest proporcją.
- Pamiętajmy, że standardowo liczbę sukcesów modelujemy *rozkładem dwumianowym*.
- Próbujemy estymować średnią wartość $\theta = \mathrm{E}(\mathbf{y}|x_1, x_2, ...x_k)$ zmiennej y w zależności od wartości predyktorów.

Trzeba wziąć pod uwagę, iż wariancja w przypadku rozkładu dwumianowego jest uzależniona od średniej:

$$\mathrm{var}(\mathbf{y_i}|x_1, x_2, ...x_k) = \theta(1-\theta)/m_i$$

gdzie m_i oznacza liczbę prób.

Regresja logistyczna

- Z typowym zastosowaniem regresji logistycznej mamy do czynienia, gdy zmienna y jest proporcją.
- Pamiętajmy, że standardowo liczbę sukcesów modelujemy *rozkładem dwumianowym*.
- Próbujemy estymować średnią wartość $\theta = \mathrm{E}(\mathbf{y}|x_1, x_2, ...x_k)$ zmiennej y w zależności od wartości predyktorów.

Trzeba wziąć pod uwagę, iż wariancja w przypadku rozkładu dwumianowego jest uzależniona od średniej:

$$\mathrm{var}(\mathbf{y_i}|x_1, x_2, ...x_k) = \theta(1-\theta)/m_i$$

gdzie m_i oznacza liczbę prób.