Poprzednim razem

- Statystyka (w dwóch znaczeniach)
- Estymatory punktowe

$$T_n(E(X)) \equiv \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$T_n(\sigma(X)) \equiv S(X) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}$$

- Estymatory przedziałowe
- Przedział ufności, poziom ufności
- Dla E(X) przy znanym σ(X)

$$P\left(\overline{X} - \frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}} \le E(X) \le \overline{X} + \frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}}\right) = \gamma$$

Estymacja przedziałowa wartości oczekiwanej

- **Przypadek 2** – nie znamy σ(X)
- Twierdzenie: Statystyka

$$t \equiv \frac{\overline{X} - E(X)}{S(\overline{X})} = \frac{\left(\overline{X} - E(X)\right)\sqrt{n}}{S(X)}$$

$$S(\overline{X}) = \sqrt{\frac{1}{n}}S(X) = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \overline{X})^2}$$

ma rozkład t-Studenta o n-1 stopniach swobody. Tak jak w poprzednim przypadku jest to rozkład symetryczny względem x=0 i analogicznie:

$$P\left(\overline{X} - \frac{1}{\sqrt{n}}S(X)t_{\frac{1+\gamma}{2}} \le E(X) \le \overline{X} + \frac{1}{\sqrt{n}}S(X)t_{\frac{1+\gamma}{2}}\right) = \gamma$$

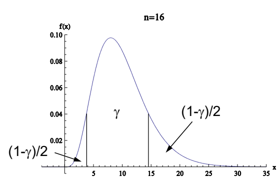
$t_{\frac{1-\gamma}{2}}$ *i* $t_{\frac{1+\gamma}{2}}$ to kwantyle rozkładu t-Studenta o n-1 stopniach swobody

Estymacja przedziałowa wariancji

- Twierdzenie: Statystyka

$$Y \equiv \frac{(n-1)S^2(X)}{\sigma^2(X)} \qquad S(X) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2}$$

- ma rozkład χ^2 (Chi-kwadrat) o n-1 stopniach swobody. Ten rozkład nie jest symetryczny względem x=0.



Estymacja przedziałowa wariancji

$$P\left(\chi^{2,n-1}_{\frac{1-\gamma}{2}} \le Y \le \chi^{2,n-1}_{\frac{1+\gamma}{2}}\right) = \gamma$$

$$P\left(\chi^{2,n-1}_{\frac{1-\gamma}{2}} \le \frac{(n-1)S^2(X)}{\sigma^2(X)} \le \chi^{2,n-1}_{\frac{1+\gamma}{2}}\right) = \gamma$$

$$P\left(\frac{(n-1)S^2(X)}{\chi^{2,n-1}_{\frac{1+\gamma}{2}}} \le \sigma^2(X) \le \frac{(n-1)S^2(X)}{\chi^{2,n-1}_{\frac{1-\gamma}{2}}}\right) = \gamma$$

$$P\left(\sqrt{\frac{(n-1)S^2(X)}{\chi^{2,n-1}_{\frac{1+\gamma}{2}}}} \le \sigma(X) \le \sqrt{\frac{(n-1)S^2(X)}{\chi^{2,n-1}_{\frac{1-\gamma}{2}}}}\right) = \gamma$$

$$\boxed{T_n^L(var(X)) = \frac{(n-1)S^2(X)}{\chi^{2,n-1}_{\frac{1+\gamma}{2}}} \qquad T_n^P(var(X)) = \frac{(n-1)S^2(X)}{\chi^{2,n-1}_{\frac{1-\gamma}{2}}}}$$

$\chi^{2,n-1}_{\frac{1-\gamma}{2}}$ *i* $\chi^{2,n-1}_{\frac{1+\gamma}{2}}$

to kwantyle rozkładu Chi-kwadrat o n-1 stopniach swobody

Znajdowanie estymatorów – metoda największej wiarygodności

- Idea: To co zaobserwowaliśmy było najbardziej prawdopodobne spośród wszystkich możliwych prób.
- Prawdopodobieństwo wylosowania próby $x_1, x_2, \ldots x_n$

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2, \ldots, \theta_p)dx_i$$

- **Funkcja (największej) wiarygodności**

$$\tilde{L}(\theta_1, \theta_2, \ldots, \theta_p) = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2, \ldots, \theta_p)$$

- **Metoda największej wiarygodności** polega na szukaniu wartości parametrów $\Theta_1, \Theta_2, \ldots, \Theta_p$, dla których funkcja największej wiarygodności osiąga maksimum.
  Tak otrzymane wartości przyjmujemy za estymatory parametrów $\Theta_1, \Theta_2, \ldots, \Theta_p$.

Metoda największej wiarygodności

- Uwaga: często wygodniej szukać maksimum funkcji

  $$L(\theta_1, \theta_2, \ldots, \theta_p) = ln\left(\tilde{L}(\theta_1, \theta_2, \ldots, \theta_p)\right)$$

  $L(\Theta_1, \Theta_2, \ldots, \Theta_p)$ nazywamy **(logarytmiczną) funkcją (największej) wiarygodności**.

- Estymatory otrzymane metodą największej wiarygodności są zgodne, asymptotycznie nieobciążone i asymptotycznie najbardziej efektywne.
- Przykład: Estymatory parametrów rozkładu normalnego.

Metoda największej wiarygodności - przykład: estymatory parametrów rozkładu normalnego

- Szukamy estymatorów parametrów θ_1 i θ_2 rozkładu $f_X(x) = \frac{1}{\sqrt{2\pi}\theta_2} e^{\frac{-(x-\theta_1)^2}{2\theta_2^2}}$
- Mając do dyspozycji próbę $(x_1,x_2,\ldots,x_n)$ budujemy logarytmiczną funkcję największej wiarygodności L

$$L(\theta_1,\theta_2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\theta_2} e^{-\frac{(x_i-\theta_1)^2}{2\theta_2^2}} = \frac{1}{(2\pi)^{\frac{n}{2}}\theta_2^n} e^{-\frac{1}{2\theta_2^2}\sum_{i=1}^{n}(x_i-\theta_1)^2}$$

$$L = \ln\left((2\pi)^{-\frac{n}{2}}\theta_2^{-n} e^{-\frac{1}{2\theta_2^2}\sum_{i=1}^{n}(x_i-\theta_1)^2}\right) = -\frac{n}{2}\ln(2\pi) - n\ln(\theta_2) - \frac{1}{2\theta_2^2}\sum_{i=1}^{n}(x_i-\theta_1)^2$$

- I szukamy jej maksimum $\frac{\partial L}{\partial \theta_1} = 0 \quad \frac{\partial L}{\partial \theta_2} = 0$

$$\frac{\partial L}{\partial \theta_1} = -\frac{1}{2\theta_2^2}\sum_{i=1}^{n} 2(x_i-\theta_1)(-1) = \frac{1}{\theta_2^2}\sum_{i=1}^{n}(x_i-\theta_1)$$

$$\frac{\partial L}{\partial \theta_2} = -n\frac{1}{\theta_2} - \frac{(-1)}{2\theta_2^4} 2\theta_2 \sum_{i=1}^{n}(x_i-\theta_1)^2 = \frac{-n}{\theta_2} + \frac{1}{\theta_2^3}\sum_{i=1}^{n}(x_i-\theta_1)^2$$

(proszę sprawdzić drugie pochodne)

Metoda największej wiarygodności - przykład: estymatory parametrów rozkładu normalnego

- Porównujemy pochodne do zera; wiemy, że $\theta_2 \neq 0$

$$\frac{\partial L}{\partial \theta_1} = 0 \;\rightarrow\; \sum_{i=1}^{n}(x_i-\theta_1) = 0 \;\rightarrow\; (\sum_{i=1}^{n} x_i) - n\theta_1 = 0 \;\rightarrow\; \theta_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{\partial L}{\partial \theta_2} = 0 \;\rightarrow$$

$$-\frac{n}{\theta_2} + \frac{1}{\theta_2^3}\sum_{i=1}^{n}(x_i-\theta_1)^2 = 0 \quad /\cdot\theta_2 \;\rightarrow\; \frac{1}{\theta_2^2}\sum_{i=1}^{n}(x_i-\theta_1)^2 = n \;\rightarrow\; \frac{1}{n}\sum_{i=1}^{n}(x_i-\theta_1)^2 = \theta_2^2$$

- Ostatecznie przyjmujemy

$$T_n(\theta_1) = \frac{1}{n}\sum_{i=1}^{n} x_i \equiv \bar{x}$$

$$T_n(\theta_2^2) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\theta_1)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - T_n(\theta_1))^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$T_n(\theta_2) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

- Są to te same estymatory jakie wynikają z metody momentów; estymator wariancji jest obciążony.

Znajdowanie estymatorów – metoda najmniejszych kwadratów

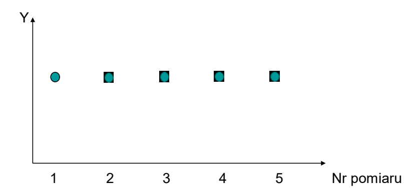
- Idea: Szukamy estymatora parametrów $\Theta_1, \Theta_2, \ldots, \Theta_p$ występujących w równaniu
  $g(y_1,y_2,\ldots,y_n;\ \Theta_1,\Theta_2,\ldots,\Theta_p)=0$,
  gdzie $(y_1,y_2,\ldots,y_n)$ to wynik n-elementowej próby.
- **Metoda najmniejszych kwadratów** polega na szukaniu wartości parametrów $\Theta_1,\Theta_2,\ldots,\Theta_p$, dla których funkcja

$$\sum_{i=1}^{n} w_i\left(g(y_i;\theta_1,\theta_2,\ldots,\theta_p)\right)^2$$

osiąga minimum. Współczynniki liczbowe w_i określają wagę jaką przykładamy do kolejnych wartości y_i, mogą być to na przykład odwrotności kwadratów błędów pomiaru y_i.
Tak otrzymane wartości przyjmujemy za estymatory parametrów $\Theta_1,\Theta_2,\ldots,\Theta_p$.

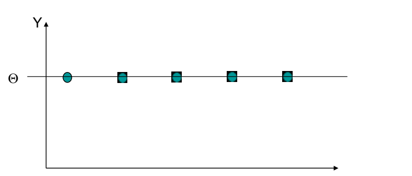
Metoda najmniejszych kwadratów

- Np. zmierzono n razy zmienną Y uzyskując wartości $y_1,y_2,\ldots,y_n$.
  Ile wynosi prawdziwa wartość zmiennej Y (oznaczamy ją przez Θ)?
  Rozpatrzmy $g(Y,\theta) = y_i - \theta$
  Gdyby pomiar był dokładny to $g(Y,\Theta)=0$.



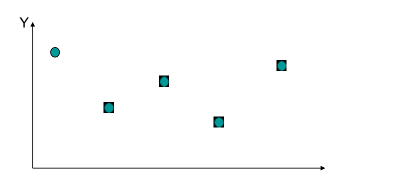
Metoda najmniejszych kwadratów

- Np. zmierzono n razy zmienną Y uzyskując wartości $y_1,y_2,\ldots,y_n$.
  Ile wynosi prawdziwa wartość zmiennej Y (oznaczamy ją przez Θ)?
  Rozpatrzmy $g(Y,\theta) = y_i - \theta$
  Gdyby pomiar był dokładny to $g(Y,\Theta)=0$.



Metoda najmniejszych kwadratów

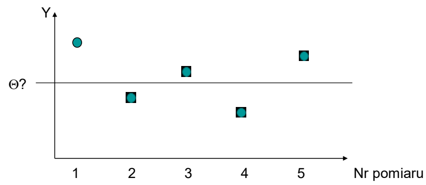
- Np. zmierzono n razy zmienną Y uzyskując wartości $y_1,y_2,\ldots,y_n$.
  Ile wynosi prawdziwa wartość zmiennej Y (oznaczamy ją przez Θ)?
  Rozpatrzmy $g(Y,\theta) = y_i - \theta$
  Gdyby pomiar był dokładny to $g(Y,\Theta)=0$.



Metoda najmniejszych kwadratów

- Np. zmierzono n razy zmienną Y uzyskując wartości $y_1, y_2, \ldots, y_n$.
  Ile wynosi prawdziwa wartość zmiennej Y (oznaczamy ją przez Θ)?
  Rozpatrzmy $g(Y,\theta) = y_i - \theta$

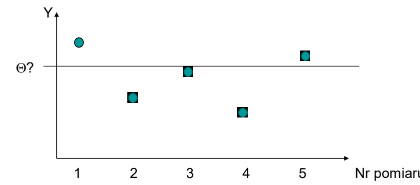
  Gdyby pomiar był dokładny to g(Y,Θ)=0.



Metoda najmniejszych kwadratów

- Np. zmierzono n razy zmienną Y uzyskując wartości $y_1, y_2, \ldots, y_n$.
  Ile wynosi prawdziwa wartość zmiennej Y (oznaczamy ją przez Θ)?
  Rozpatrzmy $g(Y,\theta) = y_i - \theta$

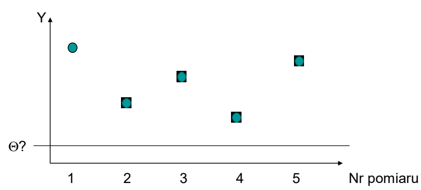
  Gdyby pomiar był dokładny to g(Y,Θ)=0.



Metoda najmniejszych kwadratów

- Np. zmierzono n razy zmienną Y uzyskując wartości $y_1, y_2, \ldots, y_n$.
  Ile wynosi prawdziwa wartość zmiennej Y (oznaczamy ją przez Θ)?
  Rozpatrzmy $g(Y,\theta) = y_i - \theta$

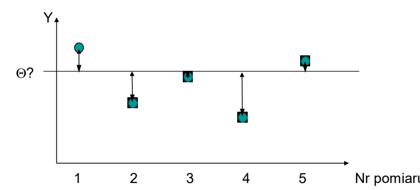
  Gdyby pomiar był dokładny to g(Y,Θ)=0.



Metoda najmniejszych kwadratów

- Np. zmierzono n razy zmienną Y uzyskując wartości $y_1, y_2, \ldots, y_n$.
  Ile wynosi prawdziwa wartość zmiennej Y (oznaczamy ją przez Θ)?
  Rozpatrzmy $g(Y,\theta) = y_i - \theta$

  Gdyby pomiar był dokładny to g(Y,Θ)=0.



Metoda najmniejszych kwadratów

- Np. zmierzono n razy zmienną Y uzyskując wartości $y_1, y_2, \ldots, y_n$.
  Ile wynosi prawdziwa wartość zmiennej Y (oznaczamy ją przez Θ)?
  Rozpatrzmy $g(Y,\theta) = y_i - \theta$

  Gdyby pomiar był dokładny to g(Y,Θ)=0.
  W praktyce minimalizujemy

$$\sum_{i=1}^{n}(y_i-\theta)^2 \quad \rightarrow \quad \frac{d}{d\theta}\left(\sum_{i=1}^{n}(y_i-\theta)^2\right) = -2\sum_{i=1}^{n}(y_i-\theta)$$

$$-2\sum_{i=1}^{n}(y_i-\theta)=0 \qquad /:(-2)$$

$$\sum_{i=1}^{n}y_i - \sum_{i=1}^{n}\theta = 0$$

$$\sum_{i=1}^{n}y_i = n\theta \quad \rightarrow \quad T_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}y_i$$

Metoda najmniejszych kwadratów

- W ogólnym przypadku estymatory pochodzące z metody najmniejszych kwadratów nie mają optymalnych własności (nawet asymptotycznie).

  Wyjątki:
- Wyniki pomiaru y_i mają rozkład normalny i są nieskorelowane:
  wtedy metoda najmniejszych kwadratów jest równoważna metodzie największej wiarygodności.
- Szukane parametry są liniowymi współczynnikami w funkcji regresji.

Funkcja regresji

- **Funkcją regresji I rodzaju** zmiennej Y względem zmiennej X nazywamy warunkową wartość oczekiwaną E(Y|X) traktowaną jako funkcję zmiennej X.

$$E(Y \mid X = x) = \int_{-\infty}^{+\infty} y f_{Y|X}(y \mid x) dy$$

$$f_{Y|X}(y \mid x) = f_{X,Y}(x,y) / f_X(x)$$

- Tw: Wartość oczekiwana kwadratu odchyleń zmiennej losowej Y od dowolnej funkcji u(X) jest minimalna gdy u(X)=E(Y|X)

$$E[(Y-u(X))^2] \geq E[(Y-E(Y \mid X))^2]$$

- Dow:

$$E[(Y-u(X))^2] = \iint dx\, dy\, f_{X,Y}(x,y)\,(y-u(x))^2 =$$

$$= \iint dx\, dy\, f_X(x) f_{Y|X}(y \mid x)\,(y-u(x))^2 =$$

$$= \int dx\, f_X(x) \int dy\, f_{Y|X}(y \mid x)\,(y-u(x))^2 =$$

Funkcja regresji

$$\int dy\, f_{Y|X}(y \mid x)\,(y-u(x))^2 =$$

$$= \int dy\, f_{Y|X}(y \mid x)\,((y-E(Y \mid X)+E(Y \mid X)-u(x))^2 =$$

$$= \int dy\, f_{Y|X}(y \mid x)\Big[(y-E(Y \mid X))^2 +$$

$$+2(y-E(Y \mid X))(E(Y \mid X)-u(x))+(E(Y \mid X)-u(x))^2\Big] =$$

$$= \int dy\, f_{Y|X}(y \mid x)\Big[(y-E(Y \mid X))^2+(E(Y \mid X)-u(x))^2\Big]+$$

$$+2(E(Y \mid X)-u(x))\int dy\, f_{Y|X}(y \mid x)(y-E(Y \mid X)) =$$

$$= \int dy\, f_{Y|X}(y \mid x)\Big[(y-E(Y \mid X))^2+(E(Y \mid X)-u(x))^2\Big]+$$

$$+2(E(Y \mid X)-u(x))\Big[\int dy\, f_{Y|X}(y \mid x) y - E(Y \mid X)\int dy\, f_{Y|X}(y \mid x)\Big]$$

$$= \int dy\, f_{Y|X}(y \mid x)\Big[(y-E(Y \mid X))^2+(E(Y \mid X)-u(x))^2\Big]+$$

$$+2(E(Y \mid X)-u(x))\big[E(Y \mid X)-E(Y \mid X)\cdot 1\big] =$$

$$= \int dy\, f_{Y|X}(y \mid x)\Big[(y-E(Y \mid X))^2+(E(Y \mid X)-u(x))^2\Big]$$

Funkcja regresji

- Zatem

$$E[(Y-u(X))^2] = \int dx\, f_X(x) \int dy\, f_{Y|X}(y \mid x)\,(y-u(x))^2 =$$

$$= \int dx\, f_X(x) \int dy\, f_{Y|X}(y \mid x)\Big[(y-E(Y \mid X))^2+(E(Y \mid X)-u(x))^2\Big] =$$

$$= \iint dx\, dy\, f_X(x)\, f_{Y|X}(y \mid x)\Big[(y-E(Y \mid X))^2\Big]+$$

$$+ \iint dx\, dy\, f_X(x)\, f_{Y|X}(y \mid x)\Big[(E(Y \mid X)-u(x))^2\Big] \geq$$

$$\geq \iint dx\, dy\, f_X(x)\, f_{Y|X}(y \mid x)\Big[(y-E(Y \mid X))^2\Big] =$$

$$= \iint dx\, dy\, f_{X,Y}(x,y)\Big[(y-E(Y \mid X))^2\Big] = E\Big[(Y-E(Y \mid X))^2\Big]$$

- Pozwala to szukać funkcji u(x), zależnej od pewnych parametrów, aproksymującej funkcję regresji I rodzaju E(Y|X) przez minimalizowanie kwadratów odchyleń u(x) od y: $\sum_{i=1}^{n}(y_i - u(x_i))^2$

Liniowa funkcja regresji II rodzaju

- **Liniowa funkcja regresji II rodzaju** przybliża liniowo E(Y|X)

  $E(Y \mid X) \approx u(x) = ax + b$

- **Regresja krzywoliniowa** to nieliniowa funkcja u(x) przybliżająca E(Y|X), której parametry znajdujemy korzystając z metody najmniejszych kwadratów.
- Przypadek I:

  Parametry $\Theta_1, \Theta_2, \dots, \Theta_p$ wchodzą do u(x) liniowo, np. jako współczynniki wielomianów, wtedy dostajemy układ równań liniowych.

  Esymatory $\Theta_1, \Theta_2, \dots, \Theta_p$ pochodzące z metody najmniejszych kwadratów są nieobciążone, najbardziej efektywne i są liniowymi funkcjami y_i. Te własności nie zależą od rozkładu zmiennej Y i są spełnione nawet dla niewielkich prób.
- Przypadek II:

  Parametry u(x) wchodzą do niej nieliniowo, np. jako współczynniki wielomianów, wtedy dostajemy układ równań nieliniowych.

Regresja krzywoliniowa - przykład

- Syt. W wyniku pomiaru otrzymaliśmy n par punktów (X_i, Y_i) i=1,…,n. Zakładamy, że pomiary są niezależne, a ich dokładność może być różna (tzn. odchylenia standardowe Y_i mogą być różne). Funkcję regresji przybliżamy funkcją $y = \sum_{j=1}^{m} \theta_j f_j(x)$, gdzie $f_j(x)$ to dowolne funkcje np. wielomiany.
- Zatem chcemy minimalizować

$$Q^2 = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}(y_i - \sum_{j=1}^{m} \theta_j f_j(x_i))^2$$

$$\frac{\partial Q^2}{\partial \theta_k} = \sum_{i=1}^{n} \frac{2}{\sigma_i^2}(y_i - \sum_{j=1}^{m} \theta_j f_j(x_i))(-f_k(x_i))$$

$$\sum_{i=1}^{n} \frac{2}{\sigma_i^2}(y_i - \sum_{j=1}^{m} \theta_j f_j(x_i))(-f_k(x_i)) = 0 \leftarrow$$

Układ m równań do rozwiązania (k=1,…,m)

Regresja krzywoliniowa - przykład

- Zapis macierzowy

$$B_{ij} = \delta_{i,j} \frac{1}{(\sigma_i)^2} = \begin{pmatrix} (\sigma_1)^{-2} & 0 & \cdots & 0 \\ 0 & (\sigma_2)^{-2} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & (\sigma_n)^{-2} \end{pmatrix} \qquad A_{ij} = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_m(x_1) \\ f_1(x_2) & f_2(x_2) & & \vdots \\ \vdots & & \ddots & \vdots \\ f_1(x_n) & \cdots & \cdots & f_m(x_n) \end{pmatrix}$$

$$Q^2 = (\vec{Y} - A\vec{\theta})^T B(\vec{Y} - A\vec{\theta}) \qquad \vec{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \qquad \vec{\theta} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_m \end{pmatrix}$$

$$\frac{\partial Q^2}{\partial \theta_k} = \Big[-2A^T B(\vec{Y} - A\vec{\theta})\Big]_k$$

$$\frac{\partial Q^2}{\partial \theta_k} = 0 \quad \rightarrow \quad A^T B(\vec{Y} - A\vec{\theta}) = 0$$

$$A^T B\vec{Y} = A^T BA\vec{\theta} \quad /(A^T BA)^{-1}$$

$$\boxed{\vec{\theta} = (A^T BA)^{-1} A^T B\vec{Y}}$$

Regresja krzywoliniowa – przykład liniowy

$\vec{\theta} = (A^T BA)^{-1} A^T B\vec{Y}$

- Wniosek: liniowy związek Y i Θ.
- Współczynniki Θ łatwe do wyliczenia (odwracanie macierzy)
- Przykład: regresja liniowa ($m=2$, $f_1(x_i)=x_i$, $f_2(x_i)=1$), dodatkowo zakładamy $\sigma_1=\sigma_2=\ldots=\sigma_n\equiv\sigma_y$.

$Y = \theta_1 f_1(X) + \theta_2 f_2(X) = \theta_1 X + \theta_2$

$$\rightarrow \boxed{\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix} = \frac{1}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\begin{pmatrix} n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i \end{pmatrix}}$$

- Jest to **regresja zwyczajna**.
- Istnieją też inne odmiany regresji: **klasyczna** (nic nie wiemy o σ_i), **ważona** (różne σ_i), **efektywna** (uwzględnia także niepewność x_i).