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)

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

1

Estymacja punktowa

- Nieobciążony
  Dow: niech wszystkie X_i pochodzą z populacji o $E(X)=X_0$

$$E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n} E(X) = \frac{1}{n}\sum_{i=1}^{n} X_0 = \frac{1}{n} n X_0 = X_0$$

- Zgodny

$$P\left(\left|\overline{X} - E(X)\right| > \varepsilon\right) = P\left(\left|\overline{X} - E(\overline{X})\right| > \varepsilon\right) \le var(\overline{X})/\varepsilon^2 = var(X)/(n\varepsilon^2)$$

$$\lim_{n\to\infty} P\left(\left|\overline{X} - E(X)\right| > \varepsilon\right) = \lim_{n\to\infty}\left(var(X)/(n\varepsilon^2)\right) = 0$$

- Najbardziej efektywny

$$\forall a > 0: \quad P(|X - \mu| \ge a) \le \frac{\sigma^2}{a^2}$$

Nierówność Czebyszewa-Bienayme

2

Estymacja punktowa

$$var(\overline{X}) = E\left[\left(\overline{X} - E(\overline{X})\right)^2\right] = E\left(\overline{X}^2\right) - \left(E(\overline{X})\right)^2 = \frac{1}{n^2}E\left(\left(\sum_{i=1}^{n} x_i\right)^2\right) - X_0^2 =$$

$$= \frac{1}{n^2}E\left(\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{n} x_j\right)\right) - X_0^2 = \frac{1}{n^2}E\left(\sum_{i=1,\ldots,n\ j\ne i} x_i x_j + \sum_{i=1}^{n} x_i^2\right) - X_0^2 =$$

$$= \frac{1}{n^2}E\left(\sum_{i=1,\ldots,n\ j\ne i} x_i x_j\right) + \frac{1}{n^2}E\left(\sum_{i=1}^{n} x_i^2\right) - X_0^2 = \frac{1}{n^2}\sum_{i=1,\ldots,n\ j\ne i} E(x_i)E(x_j) + \frac{n}{n^2}E\left(X^2\right) - X_0^2 =$$

$$= \frac{n(n-1)}{n^2}(E(X))^2 + \frac{1}{n}E\left(X^2\right) - X_0^2 = \frac{1}{n}\left[(n-1)(E(X))^2 + E\left(X^2\right) - n(E(X))^2\right] =$$

$$= \frac{1}{n}\left[E\left(X^2\right) - (E(X))^2\right] = \frac{1}{n}var(X)$$

- Czyli $\boxed{var(\overline{X}) = \frac{1}{n}var(X) \quad \Rightarrow \quad \sigma(\overline{X}) = \frac{1}{\sqrt{n}}\sigma(X)}$
- Ostatecznie przyjmujemy $X_0 = \overline{X} \pm \sigma(\overline{X})$

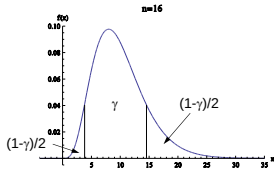
3

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.



4

Estymacja przedziałowa wariancji

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

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

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

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

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

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

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

5

Estymacja przedziałowa odchylenia standardowego

Z estymacji przedziałowej wariancji otrzymujemy

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

Czyli przedział ufności dla odchylenia standardowego dany jest przez:

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

6

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 $\Theta_1,\Theta_2,\ldots,\Theta_p$.

7

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.

8

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

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

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

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

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

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

(proszę sprawdzić drugie pochodne)

9

Metoda największej wiarygodności - przykład: estymatory parametrów rozkładu normalnego

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

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

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

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

- 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.

10

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

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

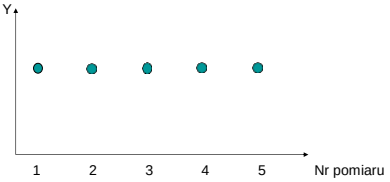
osiąga minimum. Współczynniki liczbowe w_i określają wagę jaką przykładamy do kolejnych wartości y_i, mogą być to na przykład odwrotności kwadratów błędów pomiaru y_i.
Tak otrzymane wartości przyjmujemy za estymatory parametrów $\Theta_1,\Theta_2,\ldots,\Theta_p$.

11

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 $g(Y,\Theta)=0$.

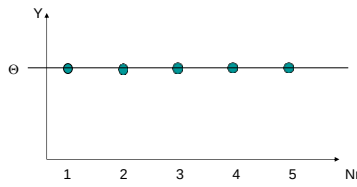


12

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.

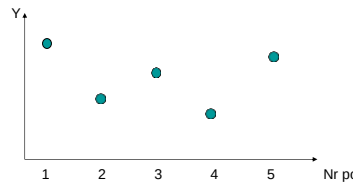


13

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.

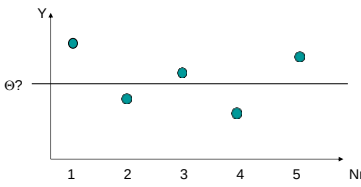


14

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.

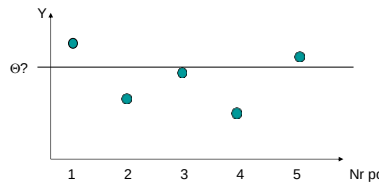


15

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.

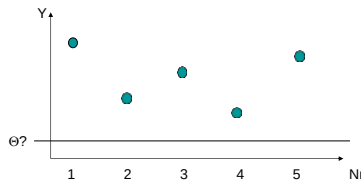


16

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.

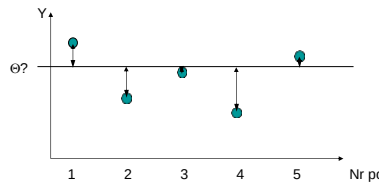


17

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.



18

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

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

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

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

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

19

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.

20

Funkcja regresji

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

$$\begin{cases} E(Y \mid X = x) = \int_{-\infty}^{+\infty} y f_{Y|X}(y \mid x)dy \\ f_{Y|X}(y \mid x) = f_{X,Y}(x,y) / f_X(x) \end{cases}$$

- 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:

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

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

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

21

Funkcja regresji

$$\int dy\, f_{Y|X}(y \mid x)(y-u(x))^2 =$$
$$= \int dy\, f_{Y|X}(y \mid x)((y-E(Y \mid X)+E(Y \mid X)-u(x))^2 =$$
$$= \int dy\, f_{Y|X}(y \mid x)\left[(y-E(Y \mid X))^2 + \right.$$
$$\left. + 2(y-E(Y \mid X))(E(Y \mid X)-u(x)) + (E(Y \mid X)-u(x))^2\right] =$$
$$= \int dy\, f_{Y|X}(y \mid x)\left[(y-E(Y \mid X))^2 + (E(Y \mid X)-u(x))^2\right] +$$
$$+ 2(E(Y \mid X)-u(x)) \int dy\, f_{Y|X}(y \mid x)(y-E(Y \mid X)) =$$
$$= \int dy\, f_{Y|X}(y \mid x)\left[(y-E(Y \mid X))^2 + (E(Y \mid X)-u(x))^2\right] +$$
$$+ 2(E(Y \mid X)-u(x))\left[\int dy\, f_{Y|X}(y \mid x)y - E(Y \mid X)\int dy\, f_{Y|X}(y \mid x)\right]$$
$$= \int dy\, f_{Y|X}(y \mid x)\left[(y-E(Y \mid X))^2 + (E(Y \mid X)-u(x))^2\right] +$$
$$+ 2(E(Y \mid X)-u(x))\left[E(Y \mid X) - E(Y \mid X)\cdot 1\right] =$$
$$= \int dy\, f_{Y|X}(y \mid x)\left[(y-E(Y \mid X))^2 + (E(Y \mid X)-u(x))^2\right]$$

22

Funkcja regresji

- Zatem

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

- 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: $\sum_{i=1}^{n}(y_i-u(x_i))^2$

23

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

24