Literatura

Faraway J.: Linear models with R, Taylor & Francis

Faraway J.: Practical Regression and Anova using R

Rencher A.C., Schaalje G.B.: Linear models in statistic, Wiley

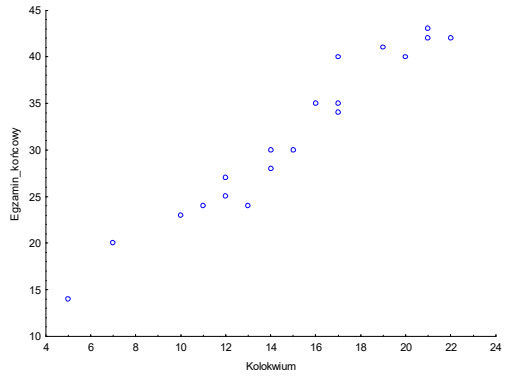
Kutner M.H, Nachtsheim C.J., Neter J., Li W.: Applied Linear Statistical Models, Wiley

Analiza zależności dwóch zmiennych ilościowych

W typowych sytuacjach, nawet jeśli celem jest analiza zachowania się pewnej losowej wielkości Y, nie ograniczamy się jedynie do obserwacji tej zmiennej, ale zbieramy również informacje towarzyszące, które mogą mieć znaczenie w analizie interesującej nas wielkości. Rozważmy na początku sytuację w której oprócz interesującej nas zmiennej Y obserwujemy jedną zmienną towarzyszącą X.

Przykład 1. Rozważmy rezultaty kolokwium (skala 0÷25 punktów) i egzaminu końcowego (skala 0÷50 punktów) z rachunku prawdopodobieństwa i statystyki (Koronacki,Mielniczuk str. 260) dla 19 studentów pewnej szkoły technicznej. Poniżej przedstawimy wykres par (x,y), gdzie x oznacza wynik kolokwium a y wynik egzaminu końcowego dla danego studenta.

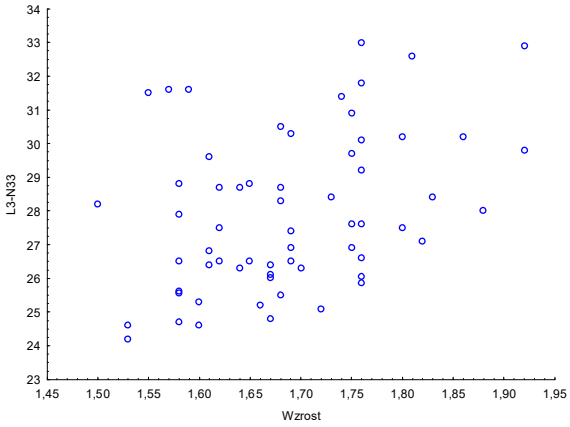
	Koronacki,Mielniczuk,str 261	
	1 Kolokwium	2 Egzamin_końcowy
1	7	20
2	11	24



Interpretacja. Powyższy wykres rozproszenia sugeruje, że dużym (małym) wartościom wyniku kolokwium odpowiadają duże (małe) wartości egzaminu końcowego. Mówimy w tym przypadku o **zależności dodatniej** miedzy zmiennymi w odróżnieniu od zależności ujemnej, gdy dużym wartościom jednej zmiennej odpowiadają małe wartości drugiej zmiennej.

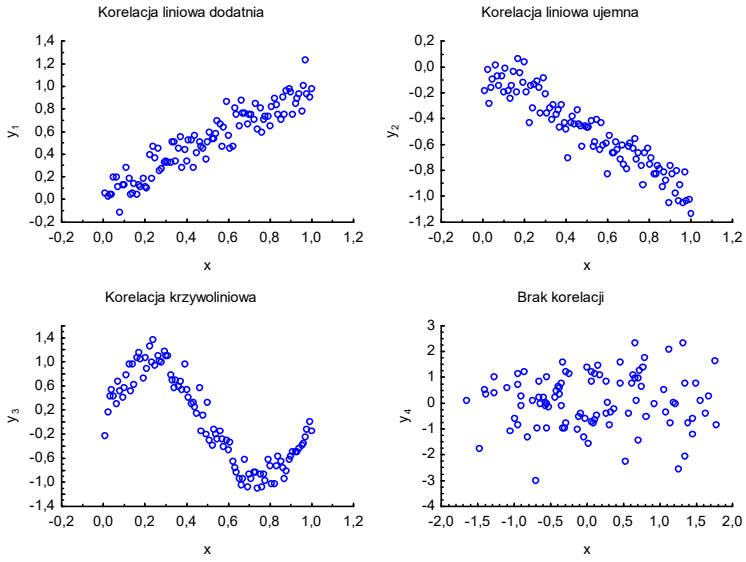
Przykład 2. Zależność latencji L3-N33 dla zdrowych osobników od ich wzrostu. Latencja L3-N33, jest to czas (w milisekundach) od momentu wzbudzenia potencjału w tzw. korzeniu L3 do osiągnięcia przez potencjał pierwszego maksimum lokalnego.

	Koronacki,Mielniczuk,str 21 i 314		
	1 Wzrost	2 L5-P40	3 L3-N33
1	1,67	50,15	26,40
2	1,59	44,50	31,60
3	1,61	48,60	29,60



Interpretacja. Chociaż zależność między zmiennymi jest podobnie jak w poprzednim przykładzie dodatnia, to siła zależności jest teraz znacznie słabsza, o czym świadczy duże rozproszenie chmury punktów.

Wykresy rozrzutu dla typowych zależności



Podstawowe statystyki próbkowe-przypomnienie.

Rozważmy n-elementową próbę prostą $(X_1,Y_1),\ldots,(X_n,Y_n)$ (czyli ciąg i.i.d.) z pewnej populacji o skończonych momentach do rzędu 2 włącznie. Statystyki

$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$, $\overline{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i$, są nieobciążonymi estymatorami wartości oczekiwanych

$\mu_X=E(X)$ i $\mu_Y=E(Y)$, natomiast statystyki

$S_X^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$, $S_Y^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2$, $S_{XY}=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right)$

są nieobciążonymi estymatorami odpowiednio $\sigma_X^2=V(X)$, $\sigma_Y^2=V(Y)$, $\sigma_{XY}=\mathrm{cov}(X,Y)$.

Współczynnik korelacji próbkowej definiujemy wzorem

$$r=\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{S_X}\right)\left(\frac{Y_i-\overline{Y}}{S_Y}\right)=\frac{S_{XY}}{\sqrt{S_X^2\,S_Y^2}}=\frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}}.$$

Współczynnik korelacji między wynikiem kolokwium a egzaminu końcowego jest równy r=0,973 a miedzy latencją L3-N33 a wzrostem jest równy r=0,51. Komentarz

Liniowa zależność między dwoma zmiennymi, prosta regresji.

Rozważmy sytuację, gdy wykres rozproszenia dla próby wartości $(x_1,y_1),\ldots,(x_n,y_n)$ wykazuje na wyraźną zależność liniową między zmiennymi x i y (np. kolokwium i egzamin końcowy). Rozważmy dowolną, ale ustaloną prostą $\hat{y}=b_0+b_1x$. Jeżeli przyjmiemy, że ta prosta ma przybliżać daną chmurę punktów, wartość $\hat{y}_i=b_0+b_1x_i$ można interpretować jako wartość y przewidywaną na podstawie rozpatrywanej prostej dla wartości zmiennej objaśniającej równej x_i. Błąd oszacowania, czyli tzw. wartość resztowa lub **rezyduum** wynosi $e_i=y_i-\hat{y}_i$. Dla prostej adekwatnie opisującej charakter zależności wartości rezyduów powinny być jak najmniejsze. Oczywiste jest, że przesunięcie prostej w kierunku ustalonego punktu próby w celu zmniejszenia odstępstwa prostej od tego punktu powoduje z reguły jej odsunięcie od innego punktu próby. Przyjmiemy podobnie, jak w definicji wariancji, za wskaźnik rozproszenia sumę kwadratów wszystkich odchyłek (rezyduów)

$$S(b_0,b_1)=\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2=\sum_{i=1}^{n}e_i^2=\sum_{i=1}^{n}\left(y_i-(b_0+b_1x_i)\right)^2 . \qquad \text{(ML1.1)}$$

Parametry b_0 i b_1 dobierzemy tak, aby suma kwadratów odchyłek była minimalna. Tak wyznaczoną prostą będziemy nazywać **prosta regresji opartą na metodzie najmniejszych kwadratów (MNK)**. Wyznaczymy teraz jawną postać tej prostej. Z WK istnienia ekstremum mamy dwa równania

$$\begin{cases} \dfrac{\partial S}{\partial b_0} = -2\sum_{i=1}^{n}\left(y_i - (b_0 + b_1 x_i)\right) = 0, & \text{(ML1.2)} \\ \dfrac{\partial S}{\partial b_1} = -2\sum_{i=1}^{n} x_i\left(y_i - (b_0 + b_1 x_i)\right) = 0. & \text{(ML1.3)} \end{cases}$$

Stąd

$$\begin{cases} \sum_{i=1}^{n} y_i - nb_0 - b_1\sum_{i=1}^{n} x_i = 0, & \text{(ML1.4)} \\ \sum_{i=1}^{n} x_i y_i - b_0\sum_{i=1}^{n} x_i - b_1\sum_{i=1}^{n} x_i^2 = 0. & \text{(ML1.5)} \end{cases}$$

Powyższy układ można zapisać w postaci

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

Z równania (ML1.4)otrzymujemy, że

$$b_0 = \tfrac{1}{n}\left(\sum_{i=1}^{n} y_i - b_1\sum_{i=1}^{n} x_i\right) = \bar{y} - b_1\bar{x}, \qquad \text{(ML1.6)}$$

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

$$\sum_{i=1}^{n} x_i y_i - (\bar{y} - b_1\bar{x})\sum_{i=1}^{n} x_i - b_1\sum_{i=1}^{n} x_i^2 = 0 \Leftrightarrow \sum_{i=1}^{n} x_i(y_i - \bar{y}) = b_1\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right).$$

Zatem, ponieważ $\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$,

$$b_1 = \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\ . \qquad \text{(ML1.7)}$$

Dwie ostatnie równości wynikają z faktu, że $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$ i $\sum_{i=1}^{n}(y_i - \bar{y}) = 0.$

Ponadto, korzystając z definicji próbkowego współczynnika korelacji możemy napisać

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{r(n-1)s_y s_x}{(n-1)s_x^2} = r\frac{s_y}{s_x}, \qquad \text{(ML1.8)}$$

gdzie oznaczenia s_y używamy, gdy rozpatrujemy odchylenie standardowe ustalonych wartości $y_1,...,y_n$ zmiennych $Y_1,...,Y_n$. Wyznaczoną MNK prostą regresji $\hat{y} = b_0 + b_1 x$ można wygodnie zapisać w postaci

$$\frac{\hat{y} - \bar{y}}{s_y} = r \frac{x - \bar{x}}{s_x},$$

z której widać, że prosta regresji MNK przechodzi przez punkt $(\bar{x}, \bar{y})$.

Pominiemy na razie kwestię warunku wystarczającego istnienia ekstremum (minimum globalne).

Rozważmy teraz pewne własności rezyduów $e_i = y_i - \hat{y}_i = y_i - (b_0 + b_1 x_i)$.

Z równania (ML1.2) wynika, że

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - \hat{y}_i) = \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i)) \underset{\text{ML1.2}}{=} 0. \qquad \text{(ML1.9)}$$

Podobnie z równania (ML1.3) wynika, że $\sum_{i=1}^{n} x_i e_i = 0$, co w połączeniu z (ML1.9) i faktem, że $\hat{y}_i = b_0 + b_1 x_i$ otrzymujemy

$$\sum_{i=1}^{n} \hat{y}_i e_i = 0. \qquad \text{(ML1.10)}$$

gdyż $\sum_{i=1}^{n} (b_0 + b_1 x_i) e_i = b_0 \sum_{i=1}^{n} e_i + b_1 \sum_{i=1}^{n} x_i e_i = 0$.

Rozważany problem konstrukcji prostej regresji opartej na MNK można przedstawić geometrycznie.

Przyjmując oznaczenia $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix}, \mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \bar{\mathbf{y}} = \begin{bmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{bmatrix} = \bar{y} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix},$

zależności $y_i = b_0 + b_1 x_i + e_i$, i=1,...,n można zapisać wektorowo

$$\mathbf{y} = b_0 \mathbf{1} + b_1 \mathbf{x} + \mathbf{e} = \hat{\mathbf{y}} + \mathbf{e}$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = b_0 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + b_1 \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

Wektor obserwacji zmiennej zależnej y tzn. $\mathbf{y}=\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ jest więc sumą liniowej kombinacji

$\hat{\mathbf{y}}=\begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix}=b_0\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}+b_1\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ wektorów $\mathbf{1}=\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ i $\mathbf{x}=\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ oraz wektora rezyduów $\mathbf{e}=\begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$.

MNK każe dobrać liniową kombinację $\hat{\mathbf{y}}=b_0\mathbf{1}+b_1\mathbf{x}$ tak, aby zminimalizować kwadrat normy

wektora rezyduów $\|\mathbf{e}\|^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}e_i^2=\sum_{i=1}^{n}(y_i-(b_0+b_1x_i))^2=S(b_0,b_1)$.

Wiadomo, z kursu algebry liniowej, że wektor $\hat{\mathbf{y}}=b_0\mathbf{1}+b_1\mathbf{x}$ realizujący ten warunek jest projekcją ortogonalną wektora obserwacji $\mathbf{y}$ na podprzestrzeń liniową X_0=span{**1**,**x**} przestrzeni R^n z naturalnym (euklidesowym) iloczynem skalarnym.

Rzut ortogonalny wektora $\mathbf{y}$ na przestrzeń X_0 wyznaczamy mnożąc skalarnie

$\mathbf{y}=b_0\mathbf{1}+b_1\mathbf{x}+\mathbf{e}=\hat{\mathbf{y}}+\mathbf{e}$ przez wektory bazowe przestrzeni X_0=span{**1**,**x**} (zakładamy że wektory **1** i **x** są liniowo niezależne).

$$\begin{cases} b_0\langle\mathbf{1},\mathbf{1}\rangle+b_1\langle\mathbf{x},\mathbf{1}\rangle=\langle\mathbf{y},\mathbf{1}\rangle \\ b_0\langle\mathbf{1},\mathbf{x}\rangle+b_1\langle\mathbf{x},\mathbf{x}\rangle=\langle\mathbf{y},\mathbf{x}\rangle \end{cases}\Leftrightarrow\begin{bmatrix} \langle\mathbf{1},\mathbf{1}\rangle & \langle\mathbf{x},\mathbf{1}\rangle \\ \langle\mathbf{1},\mathbf{x}\rangle & \langle\mathbf{x},\mathbf{x}\rangle \end{bmatrix}\begin{bmatrix} b_0 \\ b_1 \end{bmatrix}=\begin{bmatrix} \langle\mathbf{y},\mathbf{1}\rangle \\ \langle\mathbf{y},\mathbf{x}\rangle \end{bmatrix}\Leftrightarrow\begin{bmatrix} n & \sum_{i=1}^{n}x_i \\ \sum_{i=1}^{n}x_i & \sum_{i=1}^{n}x_i^2 \end{bmatrix}\begin{bmatrix} b_0 \\ b_1 \end{bmatrix}=\begin{bmatrix} \sum_{i=1}^{n}y_i \\ \sum_{i=1}^{n}x_iy_i \end{bmatrix}.$$

Otrzymaliśmy dokładnie ten sam układ równań jak z WK istnienia ekstremum $S(b_0,b_1)$.

Zwróćmy uwagę, że z ortogonalności wektora rezyduów **e** do wektorów **1** i **x** mamy inne uzasadnienie **(**ML1.9) i **(**ML1.10)

$$\langle\mathbf{e},\mathbf{1}\rangle=0\Leftrightarrow\sum_{i=1}^{n}e_i=0,\ \langle\mathbf{e},\mathbf{x}\rangle=0\Leftrightarrow\sum_{i=1}^{n}x_ie_i=0,\ \langle\mathbf{e},\hat{\mathbf{y}}\rangle=0\Leftrightarrow\sum_{i=1}^{n}\hat{y}_ie_i=0.$$

Macierz $\begin{bmatrix} \langle\mathbf{1},\mathbf{1}\rangle & \langle\mathbf{x},\mathbf{1}\rangle \\ \langle\mathbf{1},\mathbf{x}\rangle & \langle\mathbf{x},\mathbf{x}\rangle \end{bmatrix}$, to **macierz Grama**, która jest macierzą pełnego rzędu (tu 2), jeżeli wektory **1** i **x** są liniowo niezależne. Stąd $\begin{bmatrix} b_0 \\ b_1 \end{bmatrix}=\begin{bmatrix} \langle\mathbf{1},\mathbf{1}\rangle & \langle\mathbf{x},\mathbf{1}\rangle \\ \langle\mathbf{1},\mathbf{x}\rangle & \langle\mathbf{x},\mathbf{x}\rangle \end{bmatrix}^{-1}\begin{bmatrix} \langle\mathbf{y},\mathbf{1}\rangle \\ \langle\mathbf{y},\mathbf{x}\rangle \end{bmatrix}$

Oznaczając przez $\mathbf{X}=[\mathbf{1},\mathbf{x}]=\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$ **macierz planu eksperymentu** i zapisując macierz Grama w postaci

$$\begin{bmatrix} \langle \mathbf{1},\mathbf{1} \rangle & \langle \mathbf{x},\mathbf{1} \rangle \\ \langle \mathbf{1},\mathbf{x} \rangle & \langle \mathbf{x},\mathbf{x} \rangle \end{bmatrix} = \mathbf{X}^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^{n} x_i^2 \end{bmatrix}, \text{ oraz}$$

$$\begin{bmatrix} \langle \mathbf{y},\mathbf{1} \rangle \\ \langle \mathbf{y},\mathbf{x} \rangle \end{bmatrix} = [\mathbf{1},\mathbf{x}]^{\mathrm{T}}\mathbf{y} = \mathbf{X}^{\mathrm{T}}\mathbf{y} \ ,$$

mamy układ równań $[\mathbf{X}^{\mathrm{T}}\mathbf{X}]\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \mathbf{X}^{\mathrm{T}}\mathbf{y}$ i stąd przy założeniu nieosobliwości $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ otrzymujemy odpowiednio

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = [\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \ ,$$

$$\hat{\mathbf{y}} = \mathbf{X}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} = \mathbf{H}\mathbf{y}, \ \mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y} .$$

Kwadratowa (*n*,*n*) macierz **H** zwana macierzą daszkową (hat matrix) jest macierzą idempotentną ($\mathbf{H}^2=\mathbf{H}$) i symetryczną.

Wiadomo z kursu algebry, że macierz idempotentna **H** reprezentuje rzut na podprzestrzeń $\mathcal{R}(\mathbf{H})=Im(\mathbf{H})= X_0=\mathrm{span}\{\mathbf{1},\mathbf{x}\}$ (obraz endomorfizmu **H**) wzdłuż jądra $\mathcal{N}(\mathbf{H})=Ker(\mathbf{H})$, natomiast symetria macierzy **H** gwarantuje, że rozważany rzut jest rzutem ortogonalnym (zob. uzupełnienie dotyczące przestrzeni unitarnych). Macierz **I-H** jest również symetryczną i idempotentną macierzą reprezentującą rzut ortogonalny na dopełnienie ortogonalne $\mathcal{R}(\mathbf{H})^{\perp}=\mathcal{N}(\mathbf{H})$.

Widać, że $[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1} = \frac{1}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$ i łatwo otrzymujemy jawną postać macierzy daszkowej $\mathbf{H} = \mathbf{X}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}} = [H_{ij}]_{n\times n} = \left[\frac{1}{n} + \frac{(x_i-\bar{x})(x_j-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]_{n\times n}$.

Uzasadnienie, że $\hat{\mathbf{y}} = \mathbf{X}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} = \mathbf{H}\mathbf{y}$ wyznaczony MNK prowadzi do minimum globalnego jest natychmiastowe. Weźmy dowolny wektor $\tilde{\mathbf{y}}$ z podprzestrzeni X_0 .

$\|\mathbf{y}-\tilde{\mathbf{y}}\|^2 = \|\mathbf{y}-\hat{\mathbf{y}}+\hat{\mathbf{y}}-\tilde{\mathbf{y}}\|^2 = \langle \mathbf{y}-\hat{\mathbf{y}}+\hat{\mathbf{y}}-\tilde{\mathbf{y}}, \mathbf{y}-\hat{\mathbf{y}}+\hat{\mathbf{y}}-\tilde{\mathbf{y}} \rangle = \|\mathbf{y}-\hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}}-\tilde{\mathbf{y}}\|^2$, bo

$\langle \mathbf{y}-\hat{\mathbf{y}}, \hat{\mathbf{y}}-\tilde{\mathbf{y}} \rangle = 0$.

Wyrażenie $\|\mathbf{y}-\tilde{\mathbf{y}}\|^2 = \|\mathbf{y}-\hat{\mathbf{y}}\|^2 + \|\tilde{\mathbf{y}}-\hat{\mathbf{y}}\|^2$ osiąga wartość minimalną gdy $\tilde{\mathbf{y}} = \hat{\mathbf{y}}$.

Rozkład całkowitej zmienności zmiennej objaśnianej- ocena dobroci dopasowania MNK

W przypadku gdy dysponujemy tylko obserwacjami zmiennej objaśnianej $y_1,...,y_n$ ich zmienność moglibyśmy oceniać za pomocą ich wariancji s_y^2 lub pomijając czynnik $(n\text{-}1)^{-1}$ za pomocą całkowitej sumy kwadratów (total sum of squares) $SST=\sum_{i=1}^{n}(y_i-\bar{y})^2=\|\mathbf{y}-\bar{\mathbf{y}}\|^2$, gdzie $\bar{\mathbf{y}}=\bar{y}\,\mathbf{1}$. Zauważmy, że

$$SST=\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i+\hat{y}_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2=SSE+SSR\,,\quad \text{(ML1.11)}$$

bo $\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})=b_1\sum_{i=1}^{n}(y_i-\hat{y}_i)(x_i-\bar{x})=0$ (z ML1.11) lub ortogonalność $\mathbf{y}-\hat{\mathbf{y}}\perp \mathrm{X}_0$.

$SSE=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\|\mathbf{y}-\hat{\mathbf{y}}\|^2=RSS$ to suma kwadratów błędów (*error sum of squares residual sum of squares-deviance*) a

$SSR=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2=\|\hat{\mathbf{y}}-\bar{\mathbf{y}}\|^2$ to regresyjna suma kwadratów (*regresion sum of squares*).

W przypadku, gdy chmura punktów na wykresie rozproszenia jest silnie skupiona wokół prostej MNK składnik *SSE* jest mały w porównaniu ze składnikiem *SST*. Zatem stosunek

$$R^2=\frac{SSR}{SST}=1-\frac{SSE}{SST}$$

zwany **współczynnikiem determinacji wielokrotnej** określa stopień, w jakim zależność liniowa między zmienną objaśnianą a objaśniającą tłumaczy zmienność wykresu rozproszenia. Dekompozycja (ML1.11)

$$SST=SSE+SSR\Leftrightarrow \|\mathbf{y}-\bar{\mathbf{y}}\|^2=\|\mathbf{y}-\hat{\mathbf{y}}\|^2+\|\hat{\mathbf{y}}-\bar{\mathbf{y}}\|^2,$$

jest konsekwencją tw. Pitagorasa w przestrzeni R^n. Z postaci prostej regresji $\frac{\hat{y}-\bar{y}}{s_y}=r\frac{x-\bar{x}}{s_x}$ wynika,

zależność między wektorami $\hat{\mathbf{y}}-\bar{\mathbf{y}}=r\frac{s_y}{s_x}(\mathbf{x}-\bar{\mathbf{x}})$ a stąd mamy

Stwierdzenie ML1. Prawdziwa jest równość

$$\frac{SSR}{SST}=\frac{\|\hat{\mathbf{y}}-\bar{\mathbf{y}}\|^2}{\|\mathbf{y}-\bar{\mathbf{y}}\|^2}=r^2\frac{s_y^2}{s_x^2}\frac{\|\mathbf{x}-\bar{\mathbf{x}}\|^2}{\|\mathbf{y}-\bar{\mathbf{y}}\|^2}=r^2\frac{s_y^2}{s_x^2}\frac{(n-1)s_x^2}{(n-1)s_y^2}=r^2.$$

Uwaga. W przykładzie 1 *SST*=1320,63 ; *SSE*=69,25 ; *SSR*=1251,39 więc współczynnik r^2=0,95. Natomiast w przykładzie 2 *SST*=325,48 ; *SSE*=278,62 ; *SSR*=46,86 więc współczynnik r^2=0,14.

Wektor $\overline{\mathbf{y}} = \begin{bmatrix} \overline{y} \\ \vdots \\ \overline{y} \end{bmatrix} = \overline{y} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ jest rzutem ortogonalnym wektora obserwacji $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ na podprzestrzeń span$\{\mathbf{1}\}$ przestrzeni R^n. Stąd wektor $\mathbf{y} - \overline{\mathbf{y}}$ jest elementem n-1 wymiarowej przestrzeni span$\{\mathbf{1}\}^{\perp}$.

Podobnie wektor $\hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix}$ jest elementem 2 wymiarowej podprzestrzeni X_0=span$\{\mathbf{1},\mathbf{x}\}$ przestrzeni R^n. Stąd wektor $\mathbf{y} - \hat{\mathbf{y}}$ jest elementem n-2 wymiarowej przestrzeni span$\{\mathbf{1},\mathbf{x}\}^{\perp}$. Natomiast wektor $\hat{\mathbf{y}} - \overline{\mathbf{y}}$ jest elementem 1 wymiarowej przestrzeni będącej dopełnieniem ortogonalnym przestrzeni span$\{\mathbf{1}\}$ do przestrzeni span$\{\mathbf{1},\mathbf{x}\}$, gdyż jest wektorem 2 wymiarowej przestrzeni span$\{\mathbf{1},\mathbf{x}\}$ spełniającym warunek $\langle \hat{\mathbf{y}} - \overline{\mathbf{y}}, \mathbf{1} \rangle = 0$. Dekompozycję ML1.11 uzupełnić możemy równaniem dotyczącym wymiarów (stopni swobody) rozważanych przestrzeni

$$\begin{aligned} \|\mathbf{y} - \overline{\mathbf{y}}\|^2 &= \|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \overline{\mathbf{y}}\|^2 \\ SST &= SSE + SSR \quad . \\ (n-1) &= (n-2) + 1 \end{aligned} \qquad \text{ML1.11'}$$

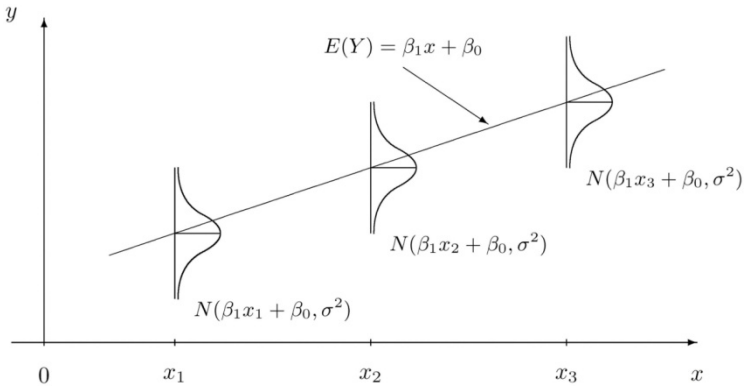
Model zależności liniowej

Zauważmy, że przedstawione podejście opiera się jedynie na analizie danych w postaci wykresu rozproszenia i nie wymaga żadnych założeń dotyczących zależności między zmienną objaśnianą a objaśniającą. Możemy jednak abstrahować od konkretnych wartości $(x_1,y_1),\ldots, (x_n,y_n)$ i potraktować b_0 i b_1 dane wzorami ML1.6 i ML1.7 jako zmienne losowe zdefiniowane na podstawie próby losowej $(x_1,Y_1),\ldots, (x_n,Y_n)$, gdzie Y_i jest objaśnianą zmienną losową odpowiadającą (nielosowej) wartości zmiennej x_i zmiennej objaśniającej x. Badanie probabilistycznych własności współczynników b_0 i b_1 będzie wymagało sformułowania modelu zależności pomiędzy wartościami x_i a zmiennymi Y_i. Ogólne założenie dotyczące powiązania tych wartości wygląda następująco:

Dla pewnych stałych β_0 i β_1 zachodzi

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad i = 1,\ldots,n \qquad \text{ML1.12}$$

gdzie ε_i są niezależnymi zmiennymi losowymi o tym samym rozkładzie z wartością oczekiwaną równą 0 i (nieznaną) wariancją σ^2. Przyjmujemy ponadto, że wartości x_i nie są wszystkie równe jednej liczbie (tzn. wektory **1** i **x** są liniowo niezależne).



Prosta $y = \beta_0 + \beta_1 x$ nosi nazwę prostej regresji, a współczynniki β_0 i β_1 są odpowiednio jej **wyrazem wolnym** i **współczynnikiem kierunkowym**. Zmienne losowe ε_i noszą nazwę **błędów losowych w modelu regresji**, a ich wariancja σ^2 wariancji błędów w modelu. Model ML1.12 zależy więc od trzech nieznanych parametrów β_0, β_1 i σ^2.

W eksperymencie obserwujemy wartości zmiennej objaśniającej x równe x_i, $i=1,\ldots,n$ i odpowiadające jej zmienne objaśniane Y_i : $(x_1,Y_1),\ldots,(x_n,Y_n)$. Przy przyjęciu, że spełniony jest warunek ML1.12 naszym celem jest wnioskowanie o nieznanych parametrach β_0, β_1 i σ^2. Oczywiście naturalnymi estymatorami współczynników β_0 i β_1 są zmienne losowe

$$b_0 = \tfrac{1}{n}\Big(\sum_{i=1}^{n} Y_i - b_1 \sum_{i=1}^{n} x_i\Big) = \overline{Y} - b_1 \overline{x}, \quad b_1 = \frac{\sum_{i=1}^{n} (x_i - \overline{x}) Y_i}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

określone w równaniach ML1.6 i ML1.7 i $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Wyznaczona przez nie prosta MNK $y = b_0 + b_1 x$ jest oszacowaniem nieznanej prostej regresji $y = \beta_0 + \beta_1 x$. Adekwatność tego oszacowania zależy od własności współczynników b_0 i b_1.

Elementarnym rachunkiem można wykazać

Stwierdzenie ML2. Zmienne losowe b_0 i b_1 są nieobciążonymi estymatorami współczynników β_0 i β_1, tzn. $E(b_0) = \beta_0$ i $E(b_1) = \beta_1$

a ich wariancje są wynoszą

$$V(b_0) = \sigma_{b_0}^2 = \sigma^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \right) \qquad \text{ML1.16}$$

$$V(b_1) = \sigma_{b_1}^2 = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \qquad \text{ML1.17}$$

Pominiemy na razie te rachunki i łatwo udowodnimy powyższe stwierdzenie później w ogólniejszej postaci.

Estymator wariancji błędów σ^2 oprzemy na następującej intuicji. Rezydua $Y_i - \hat{Y}_i$ będące odchyłkami obserwacji Y_i od wartości prostej regresji MNK są empirycznymi odpowiednikami odchyłek obserwacji Y_i od nieznanej prostej regresji $y = \beta_0 + \beta_1 x$. Zatem wariancja rezyduów w próbie powinna być naturalnym oszacowaniem wariancji błędów σ^2 . Pamiętając, że średnia rezyduów jest równa 0 i zastępując w definicji wariancji próbkowej czynnik $(n\text{-}1)^{-1}$ czynnikiem $(n\text{-}2)^{-1}$ otrzymamy następujący estymator wariancji σ^2

$$S^2 = \frac{1}{n-2} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \frac{SSE}{n-2}$$

zwany błędem średniokwadratowym.

Wykażemy później, że $E(S^2) = \sigma^2$ czyli, że błąd średniokwadratowy S^2 jest nieobciążonym estymatorem wariancji σ^2 .

Z równości ML1.16 i ML1.17 i własności błędu średniokwadratowego S^2 możemy wyznaczyć błędy standardowe estymatorów b_0 i b_1

$$SE_{b_0} = S \sqrt{\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}} \quad , \quad SE_{b_1} = \frac{S}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}} .$$

Przyjmijmy oznaczenia $\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$. Jeżeli przyjmiemy dodatkowe założenie:

Rozkład każdego z błędów ε_i jest normalny

czyli ε_i ,i=1,...,n tworzą prostą próbę losową z rozkładu normalnego $N(0,\sigma^2)$, to model regresji liniowej ML1.12 można zapisać w postaci

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \quad , \quad \boldsymbol{\varepsilon}\sim N(\mathbf{0},\sigma^2\mathbf{I})$$

możemy wyznaczyć rozkłady estymatorów współczynników prostej regresji jak również rozkłady ich wersji studentyzowanych.

Uwaga. Wszystkie rozważania dotyczące modelu regresji w zapisie macierzowym są prawdziwe także dla modelu regresji wielokrotnej.

Rzeczywiście $\mathbf{b}=(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ i $\mathbf{Y}\sim N(\mathbf{X}\boldsymbol{\beta},\sigma^2\mathbf{I})$. Stąd

$$E(\mathbf{b})=(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E(\mathbf{Y})=(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}=\boldsymbol{\beta}$$

$$V(\mathbf{b})=(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T V(\mathbf{Y})\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}=(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}=\sigma^2(\mathbf{X}^T\mathbf{X})^{-1},$$

więc $\hat{\boldsymbol{\beta}}=\mathbf{b}\sim N(\boldsymbol{\beta},\sigma^2(\mathbf{X}^T\mathbf{X})^{-1})$.

W szczególności dla regresji jednokrotnej, biorąc pod uwagę postać

$[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}=\frac{1}{n\sum_{i=1}^n(x_i-\bar{x})^2}\begin{bmatrix}\sum_{i=1}^n x_i^2 & -n\bar{x}\\ -n\bar{x} & n\end{bmatrix}$ mamy

Twierdzenie ML1. Rozkład estymatora b_1 jest rozkładem normalnym $N(\beta_1,\frac{\sigma^2}{\sum_{i=1}^n(x_i-\bar{x})^2})$. Ponadto dla studentyzowanego estymatora b_1 zachodzi

$$\frac{b_1-\beta_1}{SE_{b_1}}=\frac{(b_1-\beta_1)\sqrt{\sum_{i=1}^n(x_i-\bar{x})^2}}{S}\sim \mathrm{t}_{n-2} \qquad \text{ML1.18}$$

Rozkład estymatora b_0 jest rozkładem normalnym $N(\beta_0,\sigma^2(\frac{1}{n}+\frac{\bar{x}^2}{\sum_{i=1}^n(x_i-\bar{x})^2})$. Ponadto dla studentyzowanego estymatora b_0 zachodzi

$$\frac{b_0-\beta_0}{SE_{b_0}}=\frac{(b_0-\beta_0)}{S\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{\sum_{i=1}^n(x_i-\bar{x})^2}}}\sim \mathrm{t}_{n-2} \qquad \text{ML1.19}$$

Równości ML1.18 i 19 umożliwiają konstrukcję przedziałów ufności dla współczynników β_0 i β_1.

Stwierdzenie ML1.3. Przedziały ufności na poziomie 1-α dla współczynników β_0 i β_1 są odpowiednio postaci

$$b_0\pm t_{1-\alpha/2,n-2}\times SE_{b_0}=b_0\pm t_{1-\alpha/2,n-2}S\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{\sum_{i=1}^n(x_i-\bar{x})^2}},$$

$$b_1\pm t_{1-\alpha/2,n-2}\times SE_{b_1}=b_1\pm t_{1-\alpha/2,n-2}\frac{S}{\sqrt{\sum_{i=1}^n(x_i-\bar{x})^2}},$$

gdzie $t_{1-\alpha/2,n-2}$ oznacza kwantyl rzędu $1-\alpha/2$ z rozkładu t-Studenta z n-2 stopniami swobody.

Równości ML1.18 i ML1.19 można również wykorzystać do skonstruowania statystyk testowych dla testowania na poziomie istotności α hipotezy

H_0: $\beta_1=\beta_{1,0}$ przeciwko alternatywie H_1: $\beta_1\neq\beta_{1,0}$,

gdzie $\beta_{1,0}$ jest pewną ustaloną liczbą (zwykle $\beta_{1,0}$ =0). Statystyka testowa mająca postać

$$t=\frac{b_1-\beta_{1,0}}{SE_{b_1}}=\frac{(b_1-\beta_{1,0})\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}{S} \qquad \text{ML1.20}$$

ma przy prawdziwości hipotezy H_0 rozkład t_{n-2} . Zatem obszar krytyczny testu H_0 przeciwko alternatywie H_1 na poziomie istotności α ma postać

$$\{t:|t|\geq t_{1-\alpha/2,n-2}\}.$$

Analogicznie, przy zachowaniu tej samej hipotezy zerowej i zmianie hipotezy alternatywnej na prawostronną H_1: $\beta_1>\beta_{1,0}$ obszar krytyczny ma postać

$$\{t:t\geq t_{1-\alpha,n-2}\}$$

a dla hipotezy alternatywnej lewostronnej H_1: $\beta_1<\beta_{1,0}$ obszar krytyczny ma postać

$$\{t:t\leq -t_{1-\alpha,n-2}=t_{\alpha,n-2}\}.$$

Podobnie dla testowania na poziomie istotności α hipotezy

H_0: $\beta_0=\beta_{0,0}$ przeciwko alternatywie H_1: $\beta_0\neq\beta_{0,0}$,

gdzie $\beta_{0,0}$ jest pewną ustaloną liczbą (zwykle $\beta_{0,0}$ =0) statystyka testowa mająca postać

$$t=\frac{b_0-\beta_0}{SE_{b_0}}=\frac{(b_0-\beta_{0,0})}{S\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}}$$

ma przy prawdziwości hipotezy H_0 rozkład t_{n-2} . Konstrukcja obszarów krytycznych dla hipotez alternatywnych dwustronnych i jednostronnych jest identyczna jak w poprzednim przypadku.

Uwaga. Pakiety statystyczne podają z reguły p-wartość dla odpowiednich testów dwustronnych.

Inne podejście do testowania hipotezy H_0: $\beta_1=0$ przeciwko alternatywie H_1: $\beta_1\neq0$ można zaproponować na podstawie rozkładu zmienności zmiennej objaśnianej ML1.11'. Okazuje się, że przy spełnieniu hipotezy H_0: $\beta_1=0$ zmienne losowe $\frac{SSR}{\sigma^2}$, $\frac{SSE}{\sigma^2}$ są niezależne i mają rozkłady odpowiednio χ_1^2 i χ_{n-2}^2, stąd statystyka $F=\frac{SSR/1}{SSE/(n-2)}$ ma rozkład Snedecora $F_{1,n-2}$. Test o zbiorze krytycznym $\{F:F\geq F_{1-\alpha,1,n-2}\}$ jest testem H_0: $\beta_1=0$ przeciwko H_1: $\beta_1\neq0$ na poziomie α. Jest to zgodne z intuicją gdyż dla $\beta_1\neq0$ regresyjna suma kwadratów $SSR=\|\hat{\mathbf{y}}-\bar{\mathbf{y}}\|^2$ powinna przyjmować duże wartości w

stosunku do $SSE = \|\mathbf{y} - \hat{\mathbf{y}}\|^2$. Zatem duże wartości statystyki F powinny wskazywać na niespełnienie hipotezy H_0.

Okazuje się , że w wyniku tego rozumowania nie otrzymujemy żadnego nowego testu hipotezy H_0: $\beta_1=0$. W sytuacji jednej zmiennej objaśniającej prawdziwy jest związek $F=t^2$, gdzie t jest dane przez ML1.20.

Problem prognozy.

Przypuśćmy, że dopasowanie modelu do danych jest zadowalające: wartość współczynnika determinacji jest duża, a rezydua nie wskazują na wyraźne odstępstwa od założeń modelowych (problem ten zostanie omówiony później). Wtedy postulowany model może być użyty do prognozowania wartości zmiennej objaśnianej, gdy zmienna objaśniająca przyjmie nową wartość x_0 zwykle różną od pozostałych wartości zmiennej objaśniającej. Nowa wartość x_0 nie powinna znacząco odbiegać od "centrum" zbioru wartości zmiennej objaśniającej $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. W przeciwnym przypadku dokonamy nieuzasadnionej **ekstrapolacji** na zakres wartości x, o którym nie mamy żadnych (albo bardzo mało) informacji. Zakładając, że powyższy warunek jest spełniony chcemy odpowiedzieć na dwa różne pytania:

jaka jest **wartość średnia** zmiennej objaśnianej w sytuacji, gdy zmienna objaśniająca $x = x_0$,

jaka jest **przyszła wartość** zmiennej objaśnianej przy tym samym warunku $x = x_0$.

Zauważmy, że drugi problem jest inny niż dotychczas rozważane, gdzie estymowaliśmy (szacowaliśmy) pewien nieznany ale stały parametr rozkładu. W drugim problemie staramy się szacować pewną zmienną losową.

Z analizy modelu wynika, że możemy przyjąć iż obserwacja $Y(x_0)$ dla $x = x_0$ spełnia równanie

$$Y(x_0) = \beta_0 + \beta_1 x_0 + \varepsilon_0 \quad , \qquad \text{ML1.21}$$

gdzie ε_0 jest zmienną losową o rozkładzie $N(0,\sigma^2)$ niezależną od zmiennych losowych $\varepsilon_1,...,\varepsilon_n$.

Przewidywanie (prognoza) wartości średniej (oczekiwanej) zmiennej losowej $Y(x_0)$.

Z równania ML1.21 po obliczeniu wartości oczekiwanej mamy

$$\mu_{Y(x_0)} = E(Y(x_0)) = \beta_0 + \beta_1 x_0 .$$

Oczywistym oszacowaniem tej wartości będzie wartość

$$\hat{Y}(x_0) = b_0 + b_1 x_0 = \begin{bmatrix} 1 & x_0 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 1 & x_0 \end{bmatrix} \mathbf{b} \ . \qquad \text{ML1.22}$$

prostej MNK dla argumentu $x = x_0$.

Z nieobciążoności estymatorów b_0 i b_1 wynika, że $\hat{Y}(x_0)$ jest nieobciążonym estymatorem wartości oczekiwanej $\mu_{Y(x_0)} = E(\hat{Y}(x_0)) = E(b_0 + b_1 x_0) = \beta_0 + \beta_1 x_0$.

Biorąc pod uwagę $\mathbf{b} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1})$ i $[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1} = \frac{1}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$ wyliczamy

$$\sigma^2_{\hat{Y}(x_0)} = V(\hat{Y}(x_0)) = \begin{bmatrix} 1 & x_0 \end{bmatrix} V(\mathbf{b}) \begin{bmatrix} 1 \\ x_0 \end{bmatrix} = \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right). \qquad \text{ML1.23}$$

Definiując jak poprzednio błąd standardowy

$$SE_{\hat{Y}(x_0)} = S\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

możemy udowodnić następujące stwierdzenie

Stwierdzenie ML1.4 . $\hat{Y}(x_0) \sim N(\mu_{\hat{Y}(x_0)}, \sigma^2_{Y(x_0)})$, $\frac{\hat{Y}(x_0) - \mu_{\hat{Y}(x_0)}}{SE_{\hat{Y}(x_0)}} \sim t_{n-2}$.

Na tej podstawie można skonstruować przedział ufności na poziomie 1-α dla wartości średniej $\mu_{\hat{Y}(x_0)} = \beta_0 + \beta_1 x_0$ **:**

$$\hat{Y}(x_0) \pm t_{1-\alpha/2, n-2} SE_{\hat{Y}(x_0)} . \qquad \text{ML1.24}$$

Zauważmy, że długość przedziału ufności nie jest stała dla wszystkich x_0: im punkt x_0 jest bardziej oddalony od średniej, tym przedział jest dłuższy a prognoza wartości średniej bardziej niepewna.

Przewidywanie (prognoza) przyszłej wartości zmiennej losowej $Y(x_0)$.

Podobnie jak w przypadku prognozy wartości średniej, jako estymator przyszłej wartości $Y(x_0)$ służy nam $\hat{Y}(x_0)$ dane wzorem ML1.22. Oceńmy, jak duża jest zmienność różnicy $\hat{Y}(x_0) - Y(x_0)$.

Ponieważ zmienne losowe $\hat{Y}(x_0)$ i $Y(x_0)$ są niezależne, to

$$\sigma^2_{\hat{Y}(x_0)-Y(x_0)} = V(\hat{Y}(x_0)) + V(Y(x_0)) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right).$$

Stąd $SE_{\hat{Y}(x_0)-Y(x_0)} = S\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right).$

Można teraz analogicznie do stwierdzenia ML1.4 sformułować

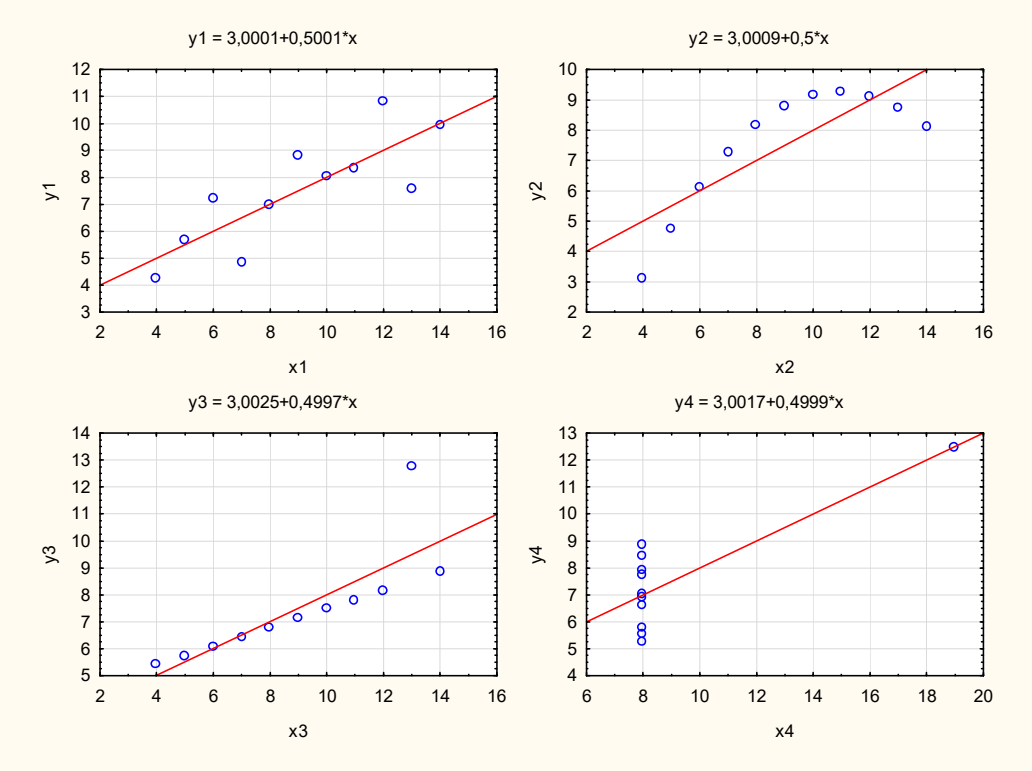
Stwierdzenie ML1.5 . $\hat{Y}(x_0)-Y(x_0) \sim N(\mu_{\hat{Y}(x_0)-Y(x_0)}, \sigma^2_{\hat{Y}(x_0)-Y(x_0)})$, $\dfrac{\hat{Y}(x_0)-Y(x_0)}{SE_{\hat{Y}(x_0)-Y(x_0)}} \sim t_{n-2}$.

Na tej podstawie przedział ufności dla zmiennej $Y(x_0)$ na poziomie 1-α jest postaci

$$\hat{Y}(x_0) \pm t_{1-\alpha/2,n-2} SE_{\hat{Y}(x_0)-Y(x_0)}. \qquad \text{ML1.25}$$

Omówione powyżej wnioskowanie dotyczące estymacji parametrów, testowania ich istotności oraz prognozy wartości średniej i przyszłej zmiennej losowej $Y(x_0)$ istotnie zależy od poprawności postulowanego modelu. Dlatego jest niezwykle ważne, aby ocenić, czy dane nie wskazują na istotne odstępstwa od przyjętych założeń.

Dane Anscomba i potrzeba głębszej diagnostyki ML



W każdym przypadku wyestymowany model (z dokładnością do zaokrągleń) jest postaci $y_i = 0{,}5x_i + 3$ z tym samym wskaźnikiem dopasowania $R^2 = 0{,}666$ i poprawionym $R^2 = 0{,}629$ oraz taką samą istotnością parametrów mierzoną p-wartością.