

Statystyka i rachunek prawdopodobieństwa

Wykład 14: Regresja liniowa

Mariusz Tarnopolski

Instytut Astronomii UMK

Statystyka ©2026



Kowariancja

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

zmiennych losowych X i Y może przyjmować różne wartości, często nieporównywalne. Np. $\text{Cov}(X_1, Y_1) = 1$ może być *duże*, zaś $\text{Cov}(X_2, Y_2) = 1000$ może być *małe*.

Kowariancja

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

zmiennych losowych X i Y może przyjmować różne wartości, często nieporównywalne. Np. $\text{Cov}(X_1, Y_1) = 1$ może być *duże*, zaś $\text{Cov}(X_2, Y_2) = 1000$ może być *małe*. W sytuacjach (zob. dalej) gdy chcemy mieć możliwość porównywania siły zależności między zmiennymi losowymi, użytecznym jest zdefiniowane **współczynnika korelacji**:

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{\mathbb{E}[(X - m_1)(Y - m_2)]}{\sigma_X \sigma_Y} = \frac{\mu_{11}}{\sqrt{\mu_{20}\mu_{02}}}$$

Jest on:

- bezwymiarowy
- miarą współzależności liniowej zmiennych X i Y
- ograniczony do przedziału $-1 \leq \rho \leq 1$

Twierdzenie 1. $|\rho| \leq 1$.

Dowód. Weźmy wyrażenie

$$\mathbb{E} \left\{ \left[\frac{X - \mathbb{E}(X)}{\sqrt{V(X)}} \pm \frac{Y - \mathbb{E}(Y)}{\sqrt{V(Y)}} \right]^2 \right\} =$$

$$\mathbb{E} \left\{ \left[\frac{X - \mathbb{E}(X)}{\sqrt{V(X)}} \right]^2 \right\} + \mathbb{E} \left\{ \left[\frac{Y - \mathbb{E}(Y)}{\sqrt{V(Y)}} \right]^2 \right\} \pm 2\mathbb{E} \left\{ \left[\frac{X - \mathbb{E}(X)}{\sqrt{V(X)}} \frac{Y - \mathbb{E}(Y)}{\sqrt{V(Y)}} \right] \right\}$$

$$= 2 \pm 2\rho$$

Skoro wyrażenie po lewej jest zawsze nieujemne (bo jest to całka z funkcji nieujemnej lub suma dodatnich składników) to również

$2 \pm 2\rho \geq 0$, skąd (rozpatrując oba znaki \pm) ostatecznie $-1 \leq \rho \leq 1$. ■

Twierdzenie 2. Jeśli X i Y są niezależne, to $\rho = 0$.

Dowód. Skoro (slajd 34, [wykład 7](#)) dla zmiennych niezależnych zachodzi $\mu_{11} = Cov(X, Y) = 0$, to—o ile wariancje $V(X) > 0$ i $V(Y) > 0$ —to również $\rho = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}} = 0$. ■

Jeśli np. X ma $V(X) = 0$ (tj. X jest stałe), to $Cov(X, Y) = 0$, ale ρ jest nieokreślone (bo symbol $0/0$).

Caveat emptor!

Twierdzenie odwrotne do powyższego nie jest, w ogólnym przypadku, prawdziwe—zmienne nieskorelowane nie muszą być niezależne.

Przykład. Niech zmienna X ma symetryczny (względem zera) rozkład prawdopodobieństwa, oraz $Y = X^2$. Wtedy

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{\text{Cov}(X, X^2)}{\sqrt{V(X)V(X^2)}}$$

Liczmy kowariancję:

$$\text{Cov}(X, X^2) = \mathbb{E}(X^3) - \mathbb{E}(X)\mathbb{E}(X^2)$$

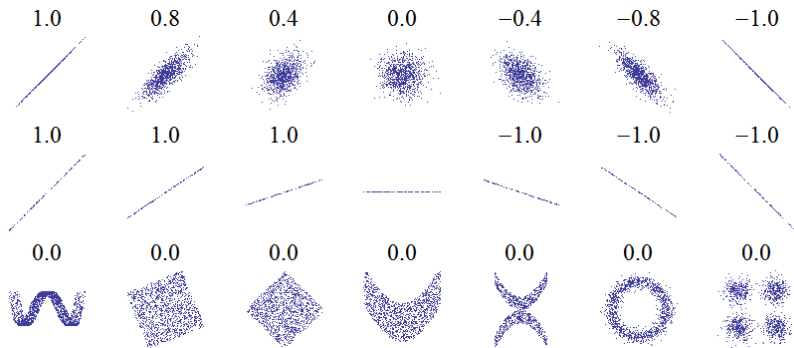
ale dla rozkładów symetrycznych ($\forall k \in \mathbb{N}$)

$$\mathbb{E}(X) = \mathbb{E}(X^3) = \dots = \mathbb{E}(X^{2k+1}) = \dots = 0$$

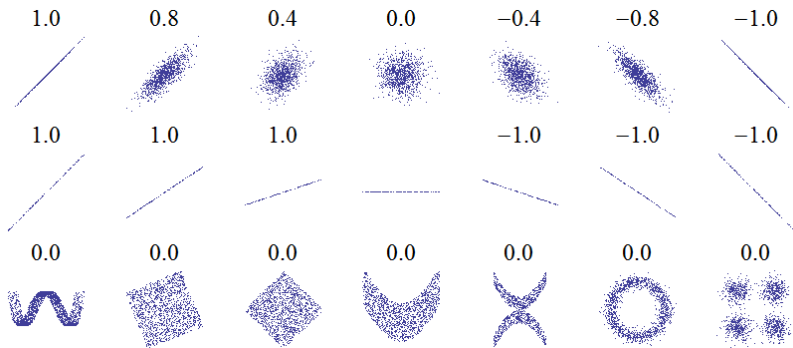
zatem [zakładając $V(X) > 0$, $V(X^2) > 0$] zachodzi

$$\rho(X, X^2) = 0$$

mimo, że zmienne X i X^2 są zależne.



¹Zob. slajd 20



Współczynnik determinacji ρ^2 —określa jaka część wariancji jednej zmiennej jest określona przez wariancję drugiej zmiennej.¹

Np. $\rho = 0.5$ przekłada się na 25%.

Zauważmy, że $> 50\%$ uzyskuje się dopiero dla $\rho > 1/\sqrt{2} \approx 0.707$.

¹Zob. slajd 20

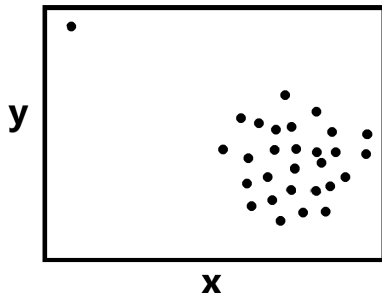
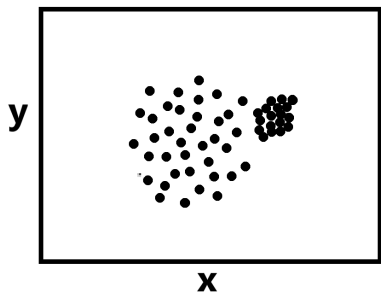
Słownictwo dot. siły związku korelacyjnego

$ \rho $	siła korelacji
0 – 0.2	brak
0.2 – 0.4	słaba
0.4 – 0.7	średnia
0.7 – 0.9	silna
0.9 – 1	bardzo silna

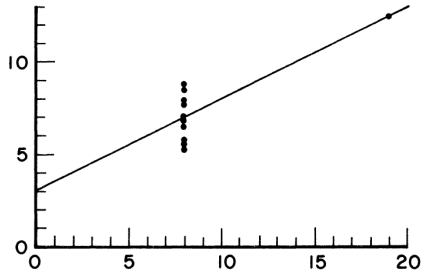
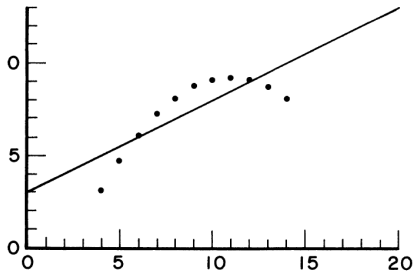
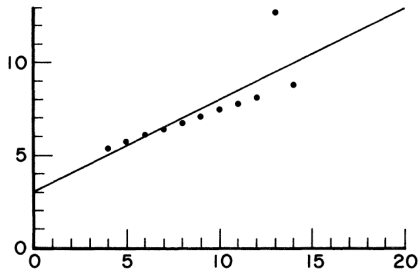
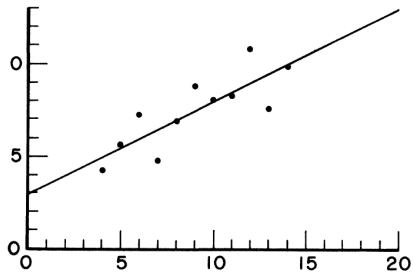
Słownictwo dot. siły związku korelacyjnego

$ \rho $	siła korelacji
0 – 0.2	brak
0.2 – 0.4	słaba
0.4 – 0.7	średnia
0.7 – 0.9	silna
0.9 – 1	bardzo silna

Korelacje należy interpretować ostrożnie:



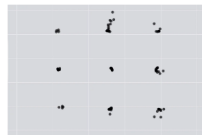
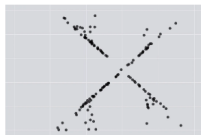
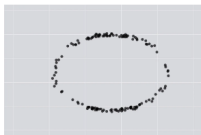
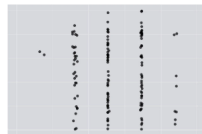
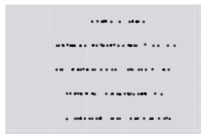
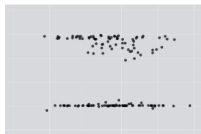
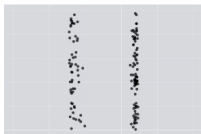
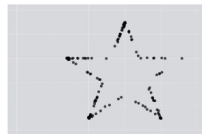
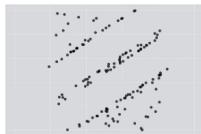
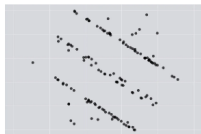
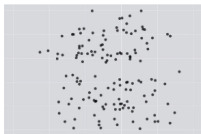
Kwartet Anscombe'a ($\rho = 0.82$):



DataSaurus:



X Mean: 54.26
Y Mean: 47.83
X SD : 16.76
Y SD : 26.93
Corr. : -0.06



W przypadku istotnej korelacji liniowej pojawia się zatem kluczowe pytanie: czy jeśli zmienne losowe X i Y są zależne, to czy możemy przewidywać jedną na podstawie wiedzy o drugiej?

W przypadku istotnej korelacji liniowej pojawia się zatem kluczowe pytanie: czy jeśli zmienne losowe X i Y są zależne, to czy możemy przewidywać jedną na podstawie wiedzy o drugiej?

Tak, ponieważ poza momentami zdefiniowanymi na [wykładzie 7](#) istnieje jeszcze jedna ich grupa—**momenty warunkowe**. Są to momenty rozkładów warunkowych (zob. [wykład 6](#)).

Warunkową wartością oczekiwaną $\mathbb{E}(X|Y = y)$ zmiennej losowej X pod warunkiem, że zmienna losowa Y przyjmuje wartość y nazywamy wartość oczekiwaną rozkładu warunkowego zmiennej X :

Warunkową wartością oczekiwaną $\mathbb{E}(X|Y = y)$ zmiennej losowej X pod warunkiem, że zmienna losowa Y przyjmuje wartość y nazywamy wartością oczekiwaną rozkładu warunkowego zmiennej X :

- dla zmiennej losowej dyskretnej:

$$\mathbb{E}(X|Y = y_k) = \sum_i x_i P(X = x_i|Y = y_k) = \frac{1}{p_{\bullet k}} \sum_i x_i p_{ik}$$

- dla zmiennej losowej ciągłej:

$$\mathbb{E}(X|Y = y) = \int_{-\infty}^{\infty} x f(x|y) dx = \frac{1}{f_2(y)} \int_{-\infty}^{\infty} x f(x, y) dx$$

Warunkową wartością oczekiwaną $\mathbb{E}(X|Y = y)$ zmiennej losowej X pod warunkiem, że zmienna losowa Y przyjmuje wartość y nazywamy wartość oczekiwaną rozkładu warunkowego zmiennej X :

- dla zmiennej losowej dyskretnej:

$$\mathbb{E}(X|Y = y_k) = \sum_i x_i P(X = x_i|Y = y_k) = \frac{1}{p_{\bullet k}} \sum_i x_i p_{ik}$$

- dla zmiennej losowej ciągłej:

$$\mathbb{E}(X|Y = y) = \int_{-\infty}^{\infty} x f(x|y) dx = \frac{1}{f_2(y)} \int_{-\infty}^{\infty} x f(x, y) dx$$

Warunkowa wartość oczekiwana zmiennej X pod warunkiem Y jest zatem zależna (funkcyjnie) od y , zatem możemy ją oznaczyć przez $m_1(y) = \mathbb{E}(X|Y = y)$.

Warunkową wartością oczekiwaną $\mathbb{E}(Y|X = x)$ zmiennej losowej Y pod warunkiem, że zmienna losowa X przyjmuje wartość x nazywamy wartość oczekiwaną rozkładu warunkowego zmiennej Y :

- dla zmiennej losowej dyskretnej:

$$\mathbb{E}(Y|X = x_i) = \sum_k y_k P(Y = y_k | X = x_i) = \frac{1}{p_{i\bullet}} \sum_k y_k p_{ik}$$

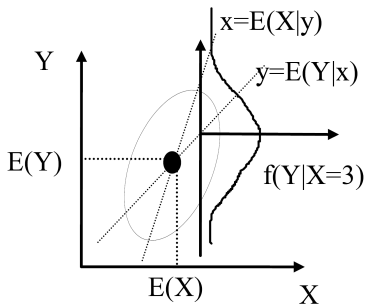
- dla zmiennej losowej ciągłej:

$$\mathbb{E}(Y|X = x) = \int_{-\infty}^{\infty} y f(y|x) dy = \frac{1}{f_1(x)} \int_{-\infty}^{\infty} y f(x, y) dy$$

Warunkowa wartość oczekiwana zmiennej Y pod warunkiem X jest zatem zależna (funkcyjnie) od x , zatem możemy ją oznaczyć przez $m_2(x) = \mathbb{E}(Y|X = x)$.

Krzywa regresji I rodzaju zmiennej X względem zmiennej Y to zbiór punktów płaszczyzny (x, y) o współrzędnych $[m_1(y), y]$.

Krzywa regresji I rodzaju zmiennej Y względem zmiennej X to zbiór punktów płaszczyzny (x, y) o współrzędnych $[x, m_2(x)]$.



Jeśli zmienna losowa dwuwymiarowa (X, Y) jest:

- 1 dyskretna, to krzywa regresji I rodzaju jest zbiorem skończonej (lub przeliczalnej) liczby punktów;
- 2 ciągła, to krzywa regresji I rodzaju jest pewną krzywą.

Jeśli zmienne X i Y są niezależne, to

$$m_1(y) = \mathbb{E}(X|Y = y) = \mathbb{E}(X)$$

$$m_2(x) = \mathbb{E}(Y|X = x) = \mathbb{E}(Y)$$

- Krzywa regresji zmiennej X względem Y leży na prostej równoległej do osi Oy .
- Krzywa regresji zmiennej Y względem X leży na prostej równoległej do osi Ox .
- Te dwie proste przecinają się w punkcie o współrzędnych $[m_1(y), m_2(x)]$ pod kątem 90° .

Twierdzenie 3. Wartość oczekiwana kwadratu odchylenia wartości zmiennej X od krzywej regresji I rodzaju zmiennej X jest minimalna:

$$\mathbb{E} \left\{ [X - m_1(Y)]^2 \right\} = \min.$$

czyli średnie odchylenie kwadratowe zmiennej X od funkcji $g(X)$ jest najmniejsze gdy $g(X)$ jest równe $m_1(Y)$, tj. jest linią regresji.

Podobnie, dla linii regresji I rodzaju zmiennej Y względem X zachodzi

$$\mathbb{E} \left\{ [Y - m_2(X)]^2 \right\} = \min.$$

Dowód. Dla zmiennej losowej (X, Y) typu ciągłego:

$$\mathbb{E} \left\{ [Y - g(X)]^2 \right\} = \int_{-\infty}^{\infty} f_1(x) \left(\int_{-\infty}^{\infty} [y - g(x)]^2 f(y|x) dy \right) dx$$

prawa strona przybiera wartość minimalną gdy $g(X) = \mathbb{E}(Y|X = x)$, bo wtedy czynnik w dużym nawiasie okrągłym jest wariancją w rozkładzie warunkowym (ponieważ wariancja minimalizuje moment drugiego rzędu—zob. slajd 21, [wykład 7](#)) ■

Krzywe regresji I rodzaju w ogólności nie są prostymi, jednak często chcemy wyznaczyć taką prostą, że spośród wszystkich prostych na płaszczyźnie (x, y) odchylenia zmiennej losowej od tej konkretnej, poszukiwanej prostej są najmniejsze. Ogólniej, chodzi o sposób wyznaczania krzywej parametrycznej o dowolnej, zadanej z góry postaci—w szczególności *dopasowanie* jej do pewnego zbioru punktów $\{(x_i, y_i)\}_{i=1}^N$.

Krzywe regresji I rodzaju w ogólności nie są prostymi, jednak często chcemy wyznaczyć taką prostą, że spośród wszystkich prostych na płaszczyźnie (x, y) odchylenia zmiennej losowej od tej konkretnej, poszukiwanej prostej są najmniejsze. Ogólniej, chodzi o sposób wyznaczania krzywej parametrycznej o dowolnej, zadanej z góry postaci—w szczególności *dopasowanie* jej do pewnego zbioru punktów $\{(x_i, y_i)\}_{i=1}^N$.

Krzywa regresji II rodzaju $g(x)$ zmiennej losowej Y względem X to pewna krzywa (funkcja), której parametry są określone metodą najmniejszych kwadratów—tj. przez minimalizację wartości przeciętnej odchylenia kwadratowego zmiennej losowej Y od krzywej $g(x)$:

$$\mathbb{E} \left\{ [Y - g(x)]^2 \right\} = \min.$$

(Podobnie określa się krzywą regresji II rodzaju zmiennej X względem Y .)

Krzywe regresji I rodzaju w ogólności nie są prostymi, jednak często chcemy wyznaczyć taką prostą, że spośród wszystkich prostych na płaszczyźnie (x, y) odchylenia zmiennej losowej od tej konkretnej, poszukiwanej prostej są najmniejsze. Ogólniej, chodzi o sposób wyznaczania krzywej parametrycznej o dowolnej, zadanej z góry postaci—w szczególności *dopasowanie* jej do pewnego zbioru punktów $\{(x_i, y_i)\}_{i=1}^N$.

Krzywa regresji II rodzaju $g(x)$ zmiennej losowej Y względem X to pewna krzywa (funkcja), której parametry są określone metodą najmniejszych kwadratów—tj. przez minimalizację wartości przeciętnej odchylenia kwadratowego zmiennej losowej Y od krzywej $g(x)$:

$$\mathbb{E} \left\{ [Y - g(x)]^2 \right\} = \min.$$

(Podobnie określa się krzywą regresji II rodzaju zmiennej X względem Y .)

W szczególności, jeśli $g(x) = ax + b$, to mówimy o prostej regresji II rodzaju (lub w skrócie: regresji liniowej) zmiennej Y względem X .

Najoptimalniejsze (w sensie najmniejszych kwadratów odchyłeń) wartości parametrów a i b znajdujemy z warunku minimalizacji:

$$\mathbb{E} \left\{ [Y - (aX + b)]^2 \right\} = \min.$$

Różniczkując ostatnie wyrażenie po a oraz b oraz przyrównując do zera:

$$\frac{\partial}{\partial a} \mathbb{E} \left\{ [Y - (aX + b)]^2 \right\} = -2\mathbb{E} [(Y - aX - b)X] = 0$$

$$\frac{\partial}{\partial b} \mathbb{E} \left\{ [Y - (aX + b)]^2 \right\} = -2\mathbb{E} [Y - aX - b] = 0$$

co po przekształceniach

$$\mathbb{E}(XY) - a\mathbb{E}(X^2) - b\mathbb{E}(X) = 0$$

$$\mathbb{E}(Y) - a\mathbb{E}(X) - b = 0$$

Powyższe jest nazywane **układem równań normalnych**. Zapisując go poprzez momenty:

$$m_{11} - am_{20} - bm_{10} = 0$$

$$m_{01} - am_{10} - b = 0$$

Jego rozwiązanie to

$$a = \frac{m_{11} - m_{01}m_{10}}{m_{20} - m_{10}^2} \equiv \frac{\mu_{11}}{\mu_{20}} \equiv \frac{\text{Cov}(X, Y)}{V(X)}$$

$$b = m_{01} - am_{10}$$

Zatem równanie prostej regresji zmiennej Y względem X jest ostatecznie postaci

$$y = \frac{\mu_{11}}{\mu_{20}} (x - m_{10}) + m_{01}$$

Podobnie otrzymuje się równanie prostej regresji zmiennej X względem Y :

$$x = \frac{\mu_{11}}{\mu_{02}} (y - m_{01}) + m_{10}$$

Obie proste regresji przechodzą przez punkt (m_{10}, m_{01}) (zwany **środkiem ciężkości populacji**).

Rozkładów zmiennych losowych X i Y zwykle nie znamy—momenty oraz parametry estymujemy z próby.

Współczynnik kierunkowy

$$a = \frac{Cov(X, Y)}{V(X)} = \frac{Cov(X, Y)}{\sigma_X^2}$$

zaś współczynnik korelacji

$$\rho = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{Cov(X, Y)}{\sigma_X^2} \frac{\sigma_X}{\sigma_Y}$$

zatem

$$\rho = a \frac{\sigma_X}{\sigma_Y}$$

Dla regresji liniowej $Y = aX + b + \varepsilon$ zachodzi

$$V(Y) = V(aX + b) + V(\varepsilon)$$

gdzie

- $V(aX + b) = a^2V(X) = a^2\sigma_X^2$ to wariancja *wyjaśniona* przez model
- $V(\varepsilon)$ to wariancja pochodząca z fluktuacji statystycznych (szumu, niepewności pomiarowych itp.—gdyby tego nie było mielibyśmy model całkowicie deterministyczny, a nie stochastyczny, czyli wszystkie punkty leżałyby idealnie na linii prostej)

Zatem

$$a^2 V(X) = \left(\rho \frac{\sigma_Y}{\sigma_X} \right)^2 \sigma_X^2 = \rho^2 \sigma_Y^2 = \rho^2 V(Y)$$

czyli skoro $\rho^2 = \frac{a^2 V(X)}{V(Y)}$, to ρ^2 to część wariancji zmiennej Y , która pochodzi od wariancji członu $aX + b$ —części deterministycznej modelu:

$$V(Y) = a^2 V(X) + V(\varepsilon) = \rho^2 V(Y) + V(\varepsilon) \quad / : V(Y) \Rightarrow$$

$$1 = \rho^2 + \frac{V(\varepsilon)}{V(Y)}$$

Podsumowując: wariancja Y rozkłada się na część pochodzącą od $aX + b$ (deterministyczną) i część losową/stochastyczną pochodzącą od ε ; po podstawieniu relacji między a i ρ okazuje się, że udział tej pierwszej w wariancji Y to dokładnie ρ^2 .

Ponieważ jednak zazwyczaj „momenty oraz parametry estymujemy z próby” (slajd 19), to estymator współczynnika korelacji $\hat{\rho}$ z próby n -elementowej wynosi

$$\hat{\rho} = \frac{\hat{c}(x, y)}{S_x S_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

gdzie nieobciążony estymator kowariancji to

$$\hat{c}(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Dopasowanie linii regresji do próby odbywa się koncepcyjnie tak jak powyżej; momenty i parametry są estymowane z tej próby i przybierają formy:

$$\hat{a} = \frac{\mu_{11}}{\mu_{20}} = \frac{\hat{c}(x, y)}{S_X^2} = \hat{\rho}(x, y) \frac{S_Y}{S_X} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\hat{b} = m_{01} - \hat{a}m_{10} = \bar{y} - \hat{a}\bar{x}$$

Dopasowanie linii regresji do próby odbywa się koncepcyjnie tak jak powyżej; momenty i parametry są estymowane z tej próby i przybierają formy:

$$\hat{a} = \frac{\mu_{11}}{\mu_{20}} = \frac{\hat{c}(x, y)}{S_X^2} = \hat{\rho}(x, y) \frac{S_Y}{S_X} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\hat{b} = m_{01} - \hat{a}m_{10} = \bar{y} - \hat{a}\bar{x}$$

W celu znalezienia linii dopasowania metodą najmniejszych kwadratów minimalizujemy wyrażenie:

$$\sum_{i=1}^n (y_i - ax_i - b)^2 = \min.$$

co w efekcie daje takie same wyrażenia na \hat{a} oraz \hat{b} jak wyżej. Są to estymatory zgodne i nieobciążone parametrów a oraz b .

Przy założeniu, że populacja ma dwuwymiarowy rozkład normalny okazuje się (pozostawiamy bez wyprowadzenia), że rozkład parametru $a \sim \mathcal{N}(a, S_a)$, gdzie

$$S_a = \sqrt{\frac{S_Y^2 (1 - \rho^2)}{S_X^2 (n - 2)}}$$

jest niepewnością (błędem), tj. odchyleniem standardowym, współczynnika kierunkowego a . Podobnie, dla wyrazu wolnego $b \sim \mathcal{N}(b, S_b)$, gdzie

$$S_b = S_a \sqrt{S_X^2 + \bar{X}^2}$$

Niech y_i oznaczają wartości prostej regresji $y_i = ax_i + b$, zaś y'_i wartości znalezionej metodą najmniejszych kwadratów linii $y'_i = \hat{a}x_i + \hat{b}$. Wtedy (bez dowodu) dla danego x_i statystyka

$$t = \frac{(\hat{a}x_i + \hat{b}) - (ax_i + b)}{S_{\hat{a}x + \hat{b}}} = \frac{y'_i - y_i}{S'_{y_i}}$$

ma rozkład t -Studenta z $(n - 2)$ stopniami swobody, gdzie

$$S'_{y_i} = S_a \left[S_X^2 + (x_i - \bar{x})^2 \right]$$

Niech y_i oznaczają wartości prostej regresji $y_i = ax_i + b$, zaś y'_i wartości znalezionej metodą najmniejszych kwadratów linii $y'_i = \hat{a}x_i + \hat{b}$. Wtedy (bez dowodu) dla danego x_i statystyka

$$t = \frac{(\hat{a}x_i + \hat{b}) - (ax_i + b)}{S_{\hat{a}x + \hat{b}}} = \frac{y'_i - y_i}{S'_{y_i}}$$

ma rozkład t -Studenta z $(n - 2)$ stopniami swobody, gdzie

$$S'_{y_i} = S_a \left[S_X^2 + (x_i - \bar{x})^2 \right]$$

Statystyka ta służy do wyznaczania obszaru ufności dla prostej regresji y dla każdego x_i z prawdopodobieństwem ufności $1 - \alpha$ jako

$$y'_i - S'_{y_i} t \left(1 - \frac{\alpha}{2}, n - 2 \right) < y(x_i) < y'_i + S'_{y_i} t \left(1 - \frac{\alpha}{2}, n - 2 \right)$$

Obszar ufności jest wtedy ograniczony krzywymi:

$$Y = 2.45 - 0.032 * X$$

Współczynnik korelacji: $r = -0.39$

