



**GDAŃSK UNIVERSITY
OF TECHNOLOGY**

FACULTY OF MANAGEMENT AND ECONOMICS

REGRESJA LINIOWA

1. LINIA REGRESJI

Zmienna zależna

Zmienna niezależna

$$\hat{y} = a + bx$$

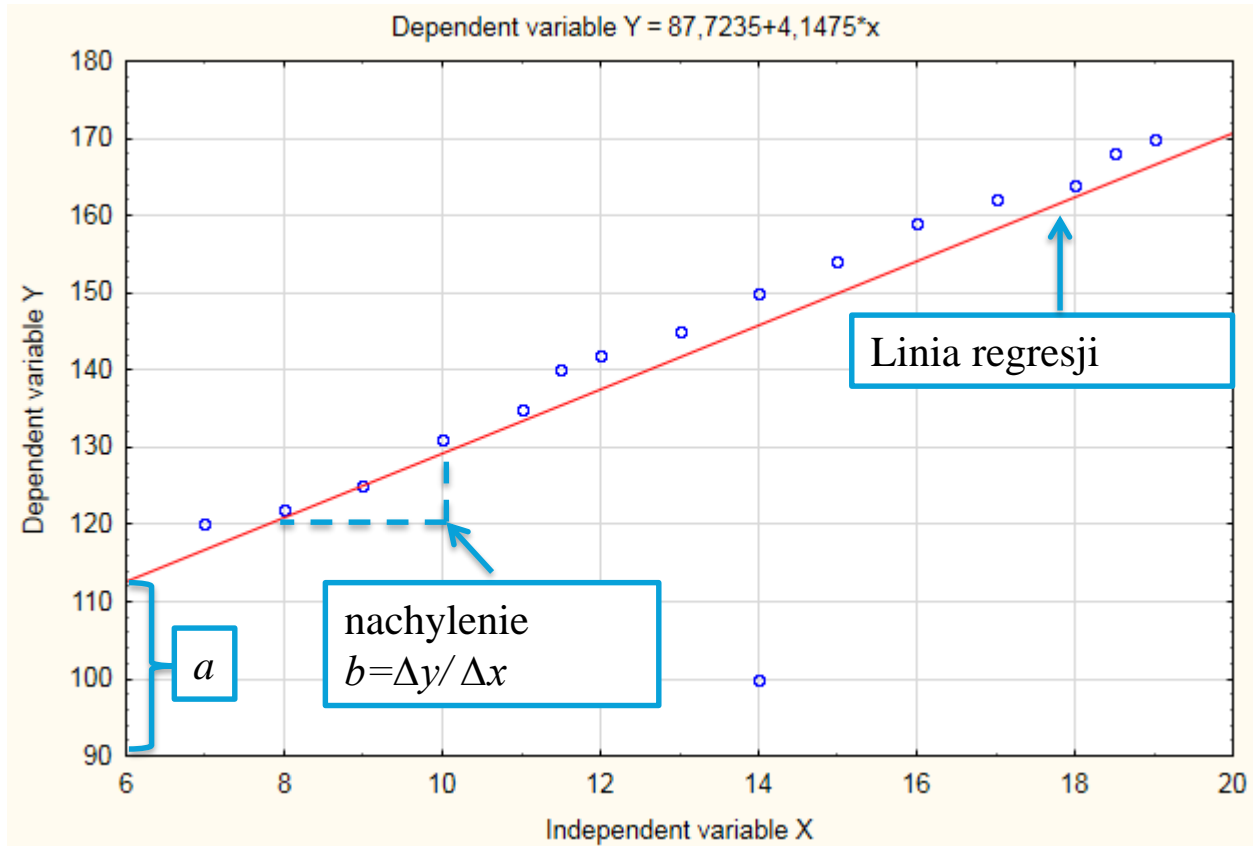
Stała

Współczynnik regresji

Estymacja

Bada zależność liniową pomiędzy zmiennymi

1. INTERPRETACJA GEOMETRYCZNA



$$\hat{y} = a + bx$$

„The line of best fit”

1. WSPÓŁCZYNNIKI- WZORY

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}$$

$$a = \bar{y} - b\bar{x}$$

ZAD 1. LINIA REGRESJI

Tabela przedstawia wskaźnik urodzeń (%) oraz wzrostu PNB w 12-stu państwach w 2011 r. Wyznacz i zinterpretuj linię regresji (zmienna zależna : Birth rate).

Country	Birth rate	GNP growth [%]
Brazil	30	5,1
Colombia	29	3,2
Costa Rica	30	3
India	35	1,4
Mexico	36	3,8
Peru	36	1
Philippines	34	2,8
Senegal	48	-0,3
South Korea	24	6,9
Sri Lanka	27	2,5
Taiwan	21	6,2
Thailand	30	4,6

Country	y_i	x_i	x_i^2	$x_i y_i$
Brazil	30	5,1	26,01	153
Colombia	29	3,2	10,24	92,8
Costa Rica	30	3	9	90
India	35	1,4	1,96	49
Mexico	36	3,8	14,44	136,8
Peru	36	1	1	36
Philippines	34	2,8	7,84	95,2
Senegal	48	-0,3	0,09	-14,4
South Korea	24	6,9	47,61	165,6
Sri Lanka	27	2,5	6,25	67,5
Taiwan	21	6,2	38,44	130,2
Thailand	30	4,6	21,16	138
Sum	380	40,2	184,04	1139,7

$$\bar{y} = \frac{380}{12} = 31.67$$

$$\bar{x} = \frac{40.2}{12} = 3.35$$

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{12 * 1139.7 - 40.2 * 380}{12 * 184.04 - (40.2)^2} = -2.7$$

$$a = \bar{y} - b\bar{x} = 31.67 - 3.35 * (-2.7) = 40.71$$

2. WSPÓŁCZYNNIK DETERMINACJI

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2$$

Total sum of squares (TSS)

$$ESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n y_i - b \sum_{i=1}^n x_i y_i$$

Error sum of squares (ESS)

$$RSS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = TSS - ESS$$

Regression sum of squares (RSS)

$$R^2 = \frac{RSS}{TSS}$$

$$R^2 \in \langle 0, 1 \rangle$$

Country	y_i	x_i	x_i^2	$x_i y_i$	y_i^2
Brazil	30	5,1	26,01	153	900
Colombia	29	3,2	10,24	92,8	841
Costa Rica	30	3	9	90	900
India	35	1,4	1,96	49	1225
Mexico	36	3,8	14,44	136,8	1296
Peru	36	1	1	36	1296
Philippines	34	2,8	7,84	95,2	1156
Senegal	48	-0,3	0,09	-14,4	2304
South Korea	24	6,9	47,61	165,6	576
Sri Lanka	27	2,5	6,25	67,5	729
Taiwan	21	6,2	38,44	130,2	441
Thailand	30	4,6	21,16	138	900
Sum	380	40,2	184,04	1139,7	12564

$$\bar{y} = \frac{380}{12} = 31.67 \quad \bar{x} = \frac{40.2}{12} = 3.35$$

$$TSS = \sum_{i=1}^n y_i^2 - n\bar{y}^2 =$$

$$= 12564 - 12 * (31.67)^2 = 530.667$$

$$ESS = \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n y_i - b \sum_{i=1}^n x_i y_i =$$

$$= 12564 - 40.71 * 380 - (-2.7) * 1139.7 =$$

$$= 170.75$$

$$RSS = TSS - ESS =$$

$$= 530.67 - 170.75 = 359.91$$

$$R^2 = \frac{RSS}{TSS} = \frac{359.91}{530.67} \approx 0.67$$

3. ANALIZA BŁĘDÓW

Błąd standardowy estymacji

$$\hat{y} = \underset{(S(a))}{a} + \underset{(S(b))}{b} x + /- S_y$$

Błąd standardowy szacunku parametru a

Błąd standardowy szacunku parametru b

3. WZORY

$$S_Y = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-k}} = \sqrt{\frac{\sum_{i=1}^n y_i^2 - a \sum_{i=1}^n y_i - b \sum_{i=1}^n x_i y_i}{n-k}}$$

$$S(b) = \frac{S_y}{\sqrt{\sum_{i=1}^n x_i^2 - n\bar{x}^2}}$$

$$S(a) = \sqrt{\frac{S_y^2 \sum_{i=1}^n x_i^2}{n \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)}}$$

Country	y_i	x_i	x_i^2	$x_i y_i$	y_i^2
Brazil	30	5,1	26,01	153	900
Colombia	29	3,2	10,24	92,8	841
Costa Rica	30	3	9	90	900
India	35	1,4	1,96	49	1225
Mexico	36	3,8	14,44	136,8	1296
Peru	36	1	1	36	1296
Philippines	34	2,8	7,84	95,2	1156
Senegal	48	-0,3	0,09	-14,4	2304
South Korea	24	6,9	47,61	165,6	576
Sri Lanka	27	2,5	6,25	67,5	729
Taiwan	21	6,2	38,44	130,2	441
Thailand	30	4,6	21,16	138	900
Sum	380	40,2	184,04	1139,7	12564

$$S_Y = \sqrt{\frac{\sum_{i=1}^n y_i^2 - a \sum_{i=1}^n y_i - b \sum_{i=1}^n x_i y_i}{n - k}} =$$

$$= \sqrt{\frac{12564 - 40.71 * 380 - (-2.7) * 1139.7}{12 - 2}} = 4.13$$

$$S(b) = \frac{S_y}{\sqrt{\sum_{i=1}^n x_i^2 - n\bar{x}^2}} =$$

$$= \frac{4.13}{\sqrt{184.04 - 12 * 3.35^2}} = 0.59$$

$$S(a) = \sqrt{\frac{S_y^2 \sum_{i=1}^n x_i^2}{n \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)}} =$$

$$= \sqrt{\frac{4.13^2 * 184.04}{12 * (184.04 - 12 * 3.35^2)}} = 2.3$$

Dziękuję za uwagę



**GDAŃSK UNIVERSITY
OF TECHNOLOGY**

FACULTY OF MANAGEMENT AND ECONOMICS