

Statystyka opisowa

Robert Pietrzykowski

email: robert_pietrzykowski@sggw.pl

www.ekonometria.info

Na dziś...

- Sprawy bieżące

	Data	Zjazd	Ćwiczenia	Data	Wykład
2015	26-09-15	1	Szereg szczegółowy. Miary położenia i zmienności.	25-09-15	Wstęp. Miary położenia i zmienności. Rozkłady normalny.
	10-10-15	2	Szereg rozdzielczy. Miary położenia i zmienności. Asymetria i kurtoza.	09-10-15	Wsp. Giniego, asymetria i kurtoza
	24-10-15	3	Wsp. Giniego. Indeksy proste.	23-10-15	Indeksy proste i złożone.
	14-11-15	4	SPRAWDZIAN PIERWSZY	13-11-15	Szeregi czasowe
	28-11-15	5	Indeksy agregatowe	27-11-15	Korelacja i regresja.
	12-12-15	6	Szeregi czasowe.	11-12-15	Statystyka matematyczna. Estymatory. Hipotezy statystyczne
2016	09-01-16	7	Korelacja i regresja. REPETYTORIUM	08-01-16	Przedziały ufności. REPETYTORIUM
	23-01-16	8	SPRAWDZIAN DRUGI	22-01-16	SPRAWDZIAN ZERO.
				29-01-16	SESJA SPRAWDZIAN I
	13-02-16	Sesja pop	POPRAWA ĆWICZEŃ	12-02-16	SESJA POP SPRAWDZIAN II

ZMIANA: Termin I egzaminu 29-01-2016 zamiast 05-02-2016

Badanie zależności (zmienne losowe)

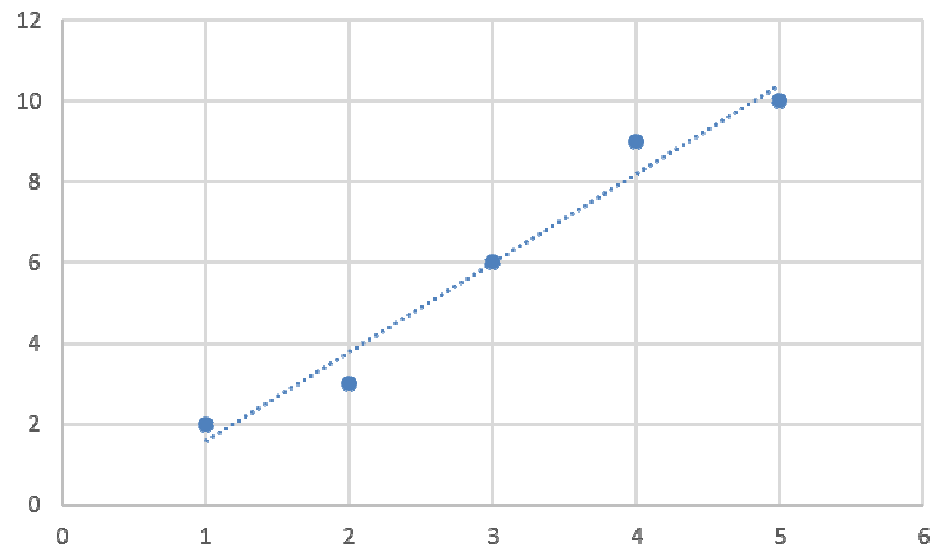
	X	Y
	1	2
	2	3
	3	6
	4	9
	5	10
Sumy	15	30
Średnie	3	6

X – procent bezrobotnych w gminie

Y – procent studentów w gminie

$$R = 0.9839$$

$$Y(x) = 2.2x - 0.6$$



Badanie zależności (zmienne losowe)

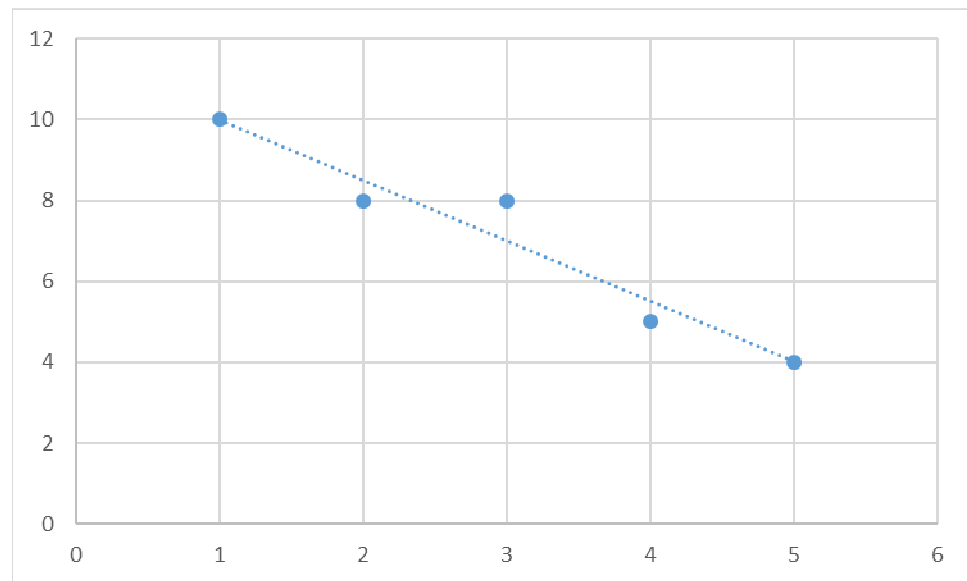
X	Y
1	10
2	8
3	8
4	5
5	4
Sumy	15 35
Średnie	3 7

Sumy
Średnie

X – procent bezrobotnych w gminie
Y – procent studentów w gminie

$$R = -0.9683$$

$$Y(x) = -1.5x + 11.5$$



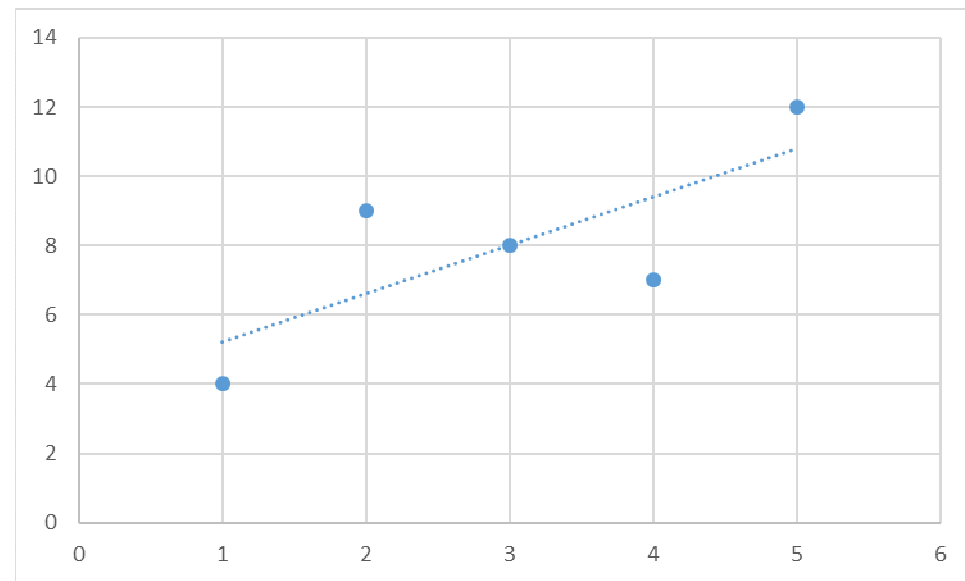
Badanie zależności (zmienne losowe)

	X	Y
	1	4
	2	9
	3	8
	4	7
	5	12
Sumy	15	40
Średnie	3	8

X – procent bezrobotnych w gminie
Y – procent studentów w gminie

$$R = 0.7592$$

$$Y(x) = 1.4x + 3.8$$



Badanie zależności (zmienne losowe)

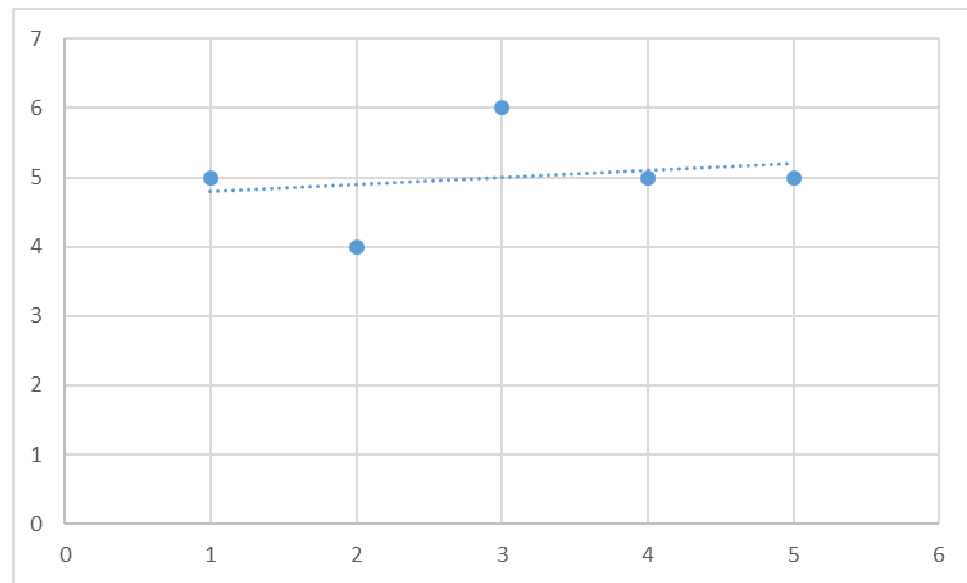
X	Y
1	5
2	4
3	6
4	5
5	5
Sumy	15 25
Średnie	3 5

Sumy
Średnie

X – procent bezrobotnych w gminie
Y – procent studentów w gminie

$$R = 0.2236$$

$$Y(x) = 0.1x + 4.7$$



- 0.9683

0.9839

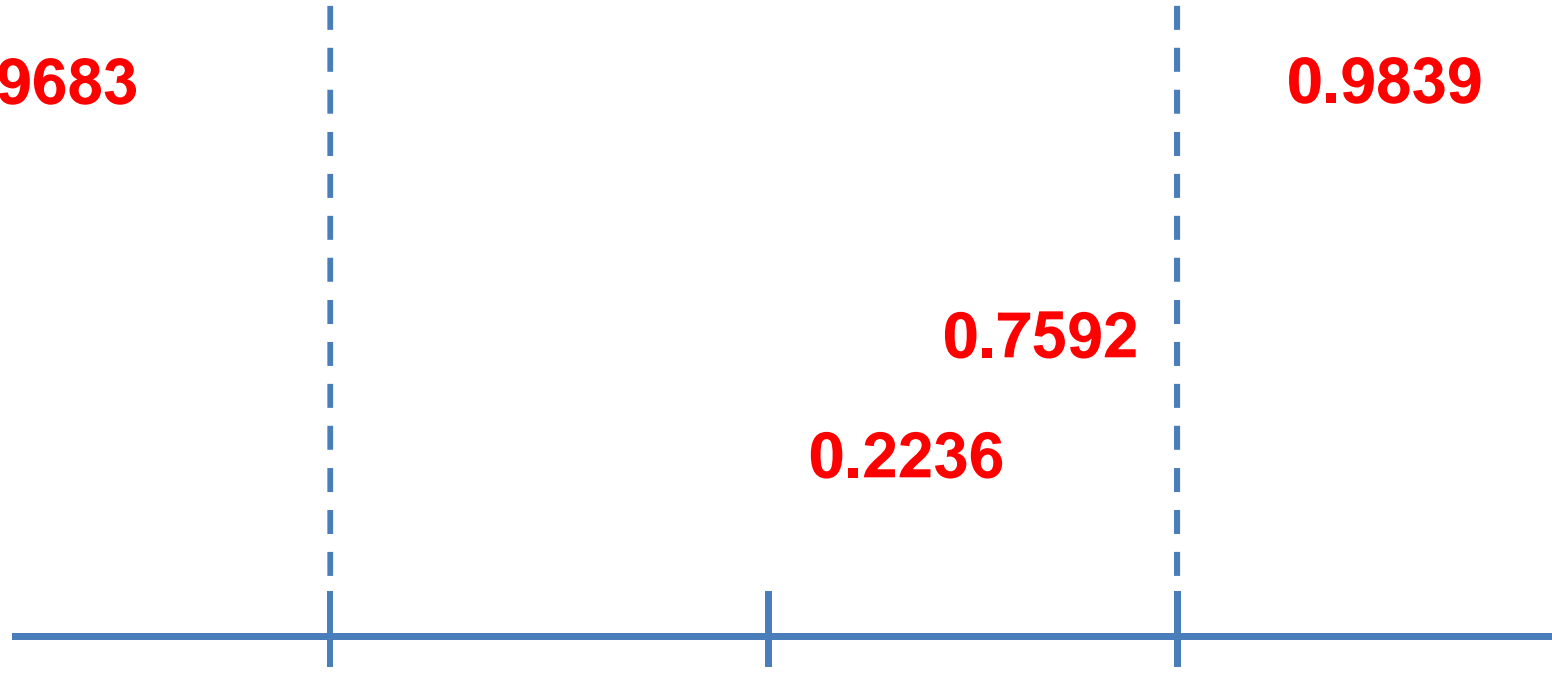
0.7592

0.2236

- 0.8783

0.0000

0.8783



$n \backslash \alpha$	0.100	0.050	0.025	0.010
3	0.9877	0.9969	0.9992	0.9999
4	0.9000	0.9500	0.9750	0.9900
5	0.8054	0.8783	0.9237	0.9587
6	0.7293	0.8114	0.8680	0.9172
7	0.6694	0.7545	0.8166	0.8745
8	0.6215	0.7067	0.7713	0.8343
9	0.5822	0.6664	0.7318	0.7977
10	0.5494	0.6319	0.6973	0.7646
11	0.5214	0.6021	0.6669	0.7348
12	0.4973	0.5760	0.6400	0.7079
13	0.4762	0.5529	0.6159	0.6835
14	0.4575	0.5324	0.5943	0.6614
15	0.4409	0.5140	0.5748	0.6411

Hipotezą statystyczną nazywamy dowolne przypuszczenie dotyczące rozkładu prawdopodobieństwa obserwowanej zmiennej losowej lub charakterystyki tegoż rozkładu, o prawdziwości którego wnioskujemy na podstawie zaobserwowanych wartości tej zmiennej losowej.

Testem hipotezy statystycznej nazywamy postępowanie mające na celu odrzucenie lub nie odrzucenie hipotezy statystycznej.

Statystyką testową nazywamy funkcję próby na podstawie której wnioskuje się o odrzuceniu lub nie hipotezy statystycznej.

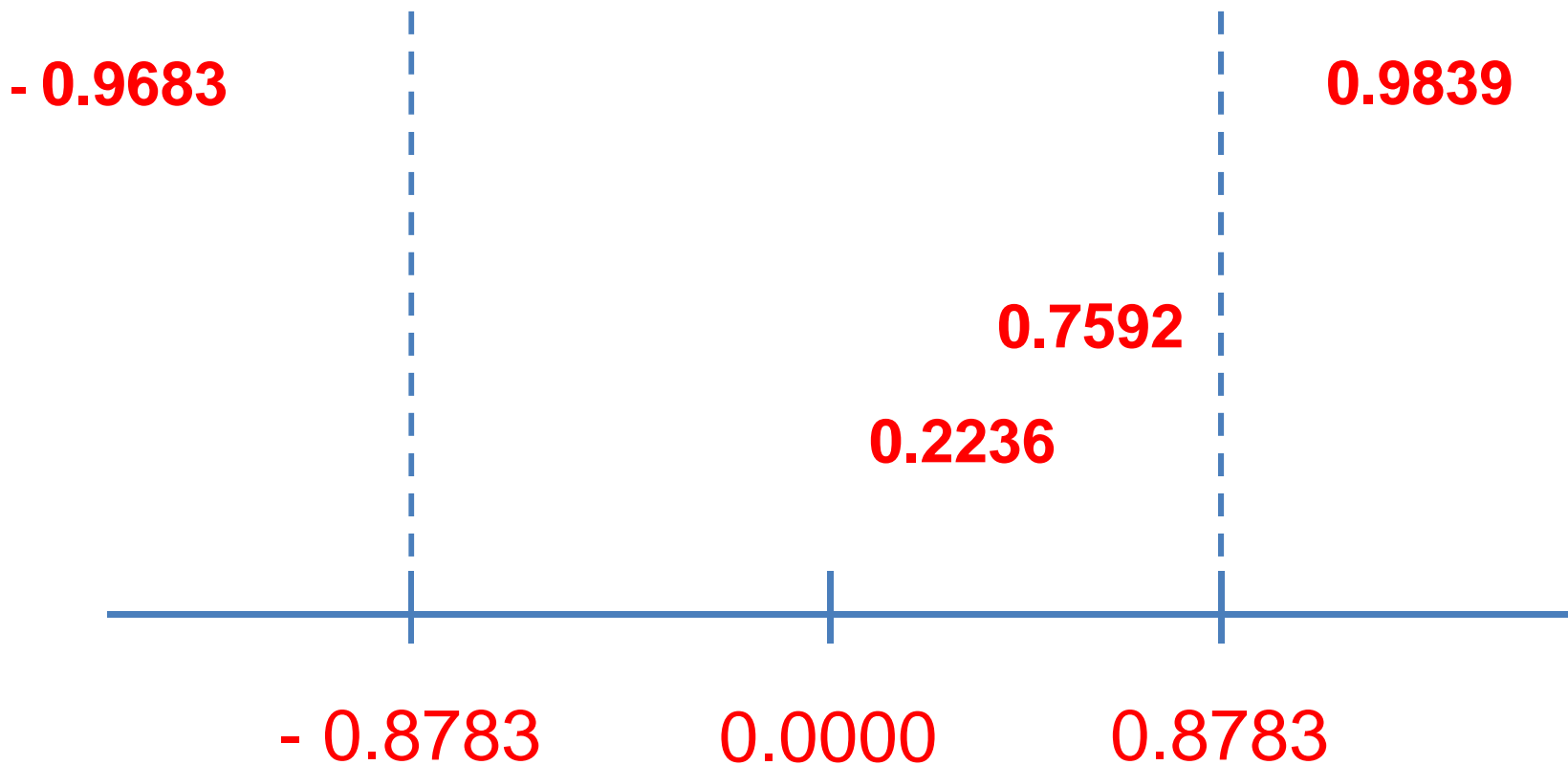
Rzeczywistość: hipoteza H_0	Wniosek o hipotezie H_0	
	nie odrzucać	odrzuścić
prawdziwa	prawidłowy	nieprawidłowy
nieprawdziwa	nieprawidłowy	prawidłowy

Błędem I rodzaju nazywamy błąd wnioskowania polegający na odrzuceniu hipotezy, gdy w rzeczywistości jest ona prawdziwa.

Błędem II rodzaju nazywamy błąd wnioskowania polegający na nieodrżuceniu hipotezy, gdy w rzeczywistości jest ona fałszywa.

Poziomem istotności nazywamy dowolną liczbę z przedziału $(0, 1)$ określającą prawdopodobieństwo popełnienia błędu I rodzaju.

Oznaczenie: α



$$H_0: \rho = 0 \quad H_1: \rho \neq 0$$

$|R| > r(n, \alpha)$ hipotezę odrzucamy

$|R| \leq r(n, \alpha)$ hipotezy nie odrzucamy

Badanie zależności (nie spełnienie założeń związanych z rozkładem lub zmienna X deterministyczna)

	X	Y
	1	4
	2	9
	3	8
	4	7
	5	12
Sumy	15	40
Średnie	3	8

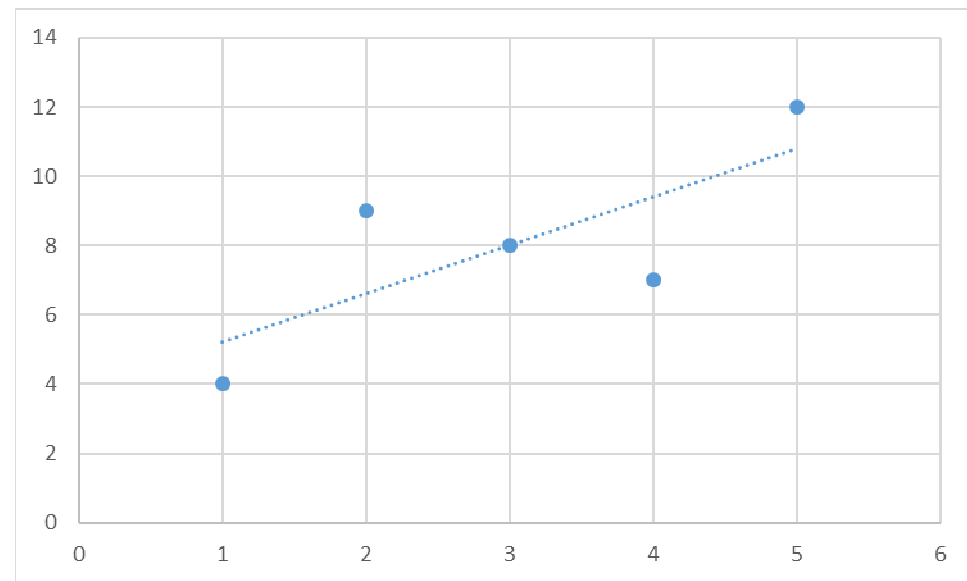
Sumy
Średnie

Źródło	Sumy kw.	Stop.sw	Śr kw.	Femp	F(α,1,3)
Regresja	19,6	1	19,6	4,08	10,13
Błąd	14,4	3	4,8		
Ogółem	34	4	8,5		

X – procent bezrobotnych w gminie
Y – procent studentów w gminie

$$R = 0.7592$$

$$Y(x) = 1.4x + 3.8$$



1. Model: $y(x) = \beta_0 + \beta_1 x + \varepsilon_i$
2. Hipoteza $H_0: \beta_1 = 0$, $H_1: \beta_1 \neq 0$
3. Test statystyczny: F
zmienna X deterministyczna
4. Jeżeli hipoteza jest prawdziwa to statystyka testowa ma rozkład Fischera $F(k, n-k-1)$ stopniami swobody.
5. $F_{emp} > F(k, n-k-1)$ hipotezę odrzucamy
 $F_{emp} \leq F(k, n-k-1)$ hipotezy nie odrzucamy

Wartości krytyczne rozkładu F-Snedecora

	0,05	0,01
	1	1
1	161,45	4052,18
2	18,51	98,50
3	10,13	34,12
4	7,71	21,20
5	6,61	16,26
6	5,99	13,75
7	5,59	12,25
8	5,32	11,26
9	5,12	10,56
10	4,96	10,04

$$F(0,05;1;3) = 10,13$$

$$F(0,01;1;3) = 34,12$$

Badanie zależności (nie spełnienie założeń związanych z rozkładem, zmienna X deterministyczna)

	X	Y
	1	4
	2	9
	3	8
	4	7
	5	12
Sumy	15	40
Średnie	3	8

Źródło	Sumy kw.	Stop.sw	Śr kw.	Femp	F($\alpha,1,3$)
Regresja	19,6	1	19,6	4,08	10,13
Błąd	14,4	3	4,8		
Ogółem	34	4	8,5		

4,08 < 10,13 hipotezy $H_0: \beta_1 = 0$ nie odrzucamy

X – procent bezrobotnych w gminie
Y – procent studentów w gminie

~~**R = 0.7592**~~

$Y(x) = 1.4x + 3.8$

