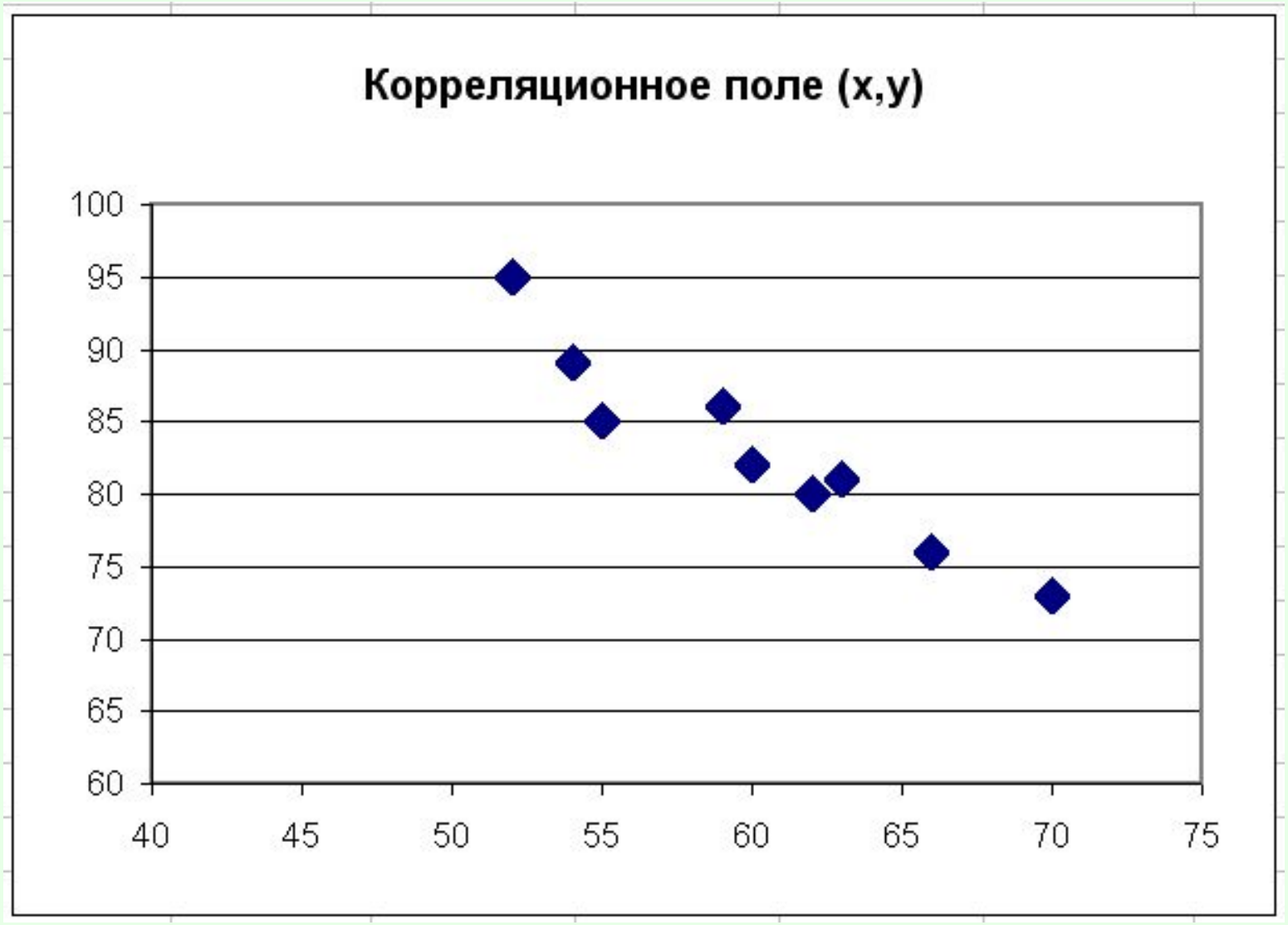
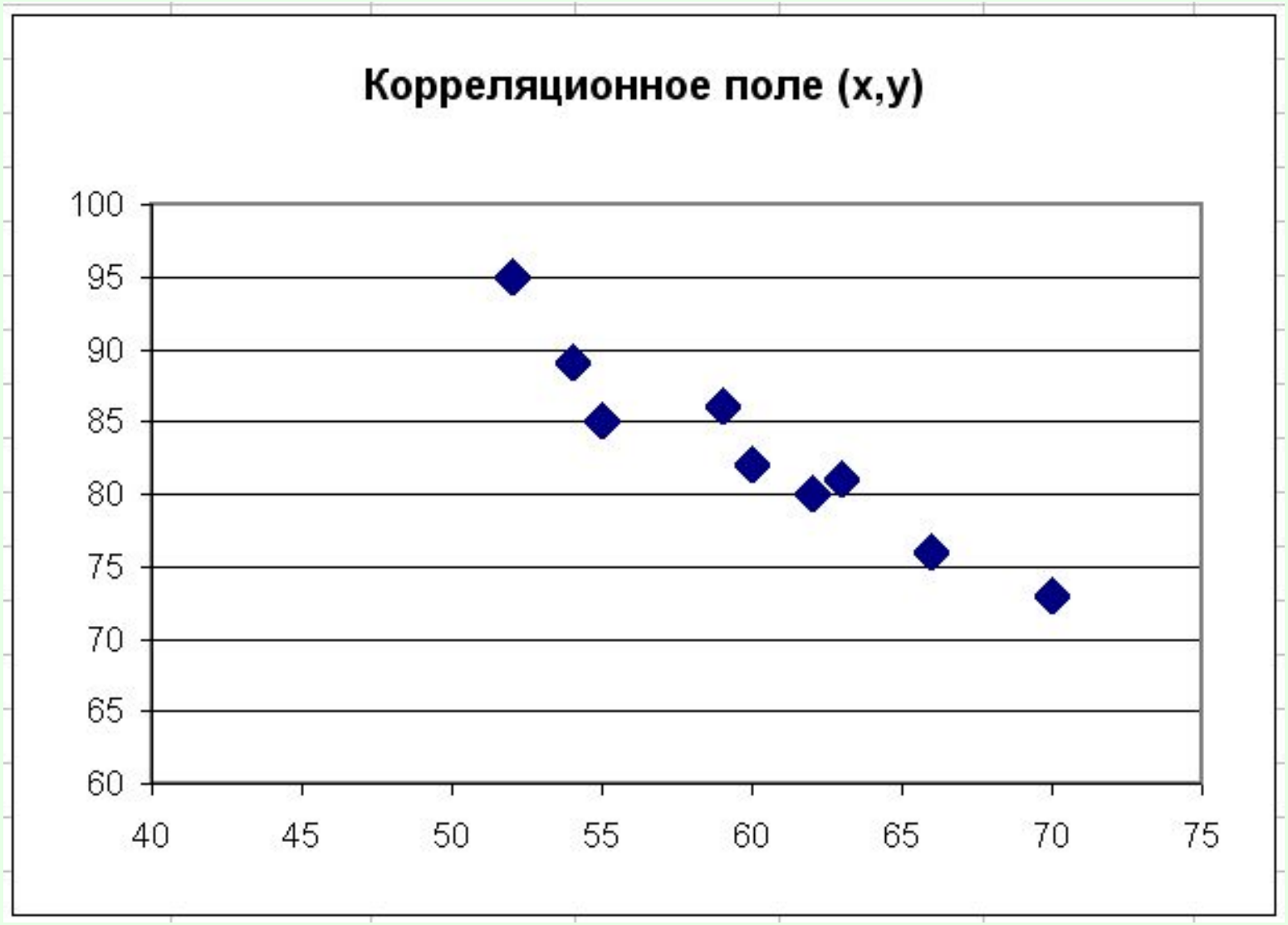
Оценка влияния количественных показателей друг на друга (параметризация)

Рассмотрим зависимость между **ценой** (x) и **спросом** (y). Данные наблюдений изобразим точками на графике



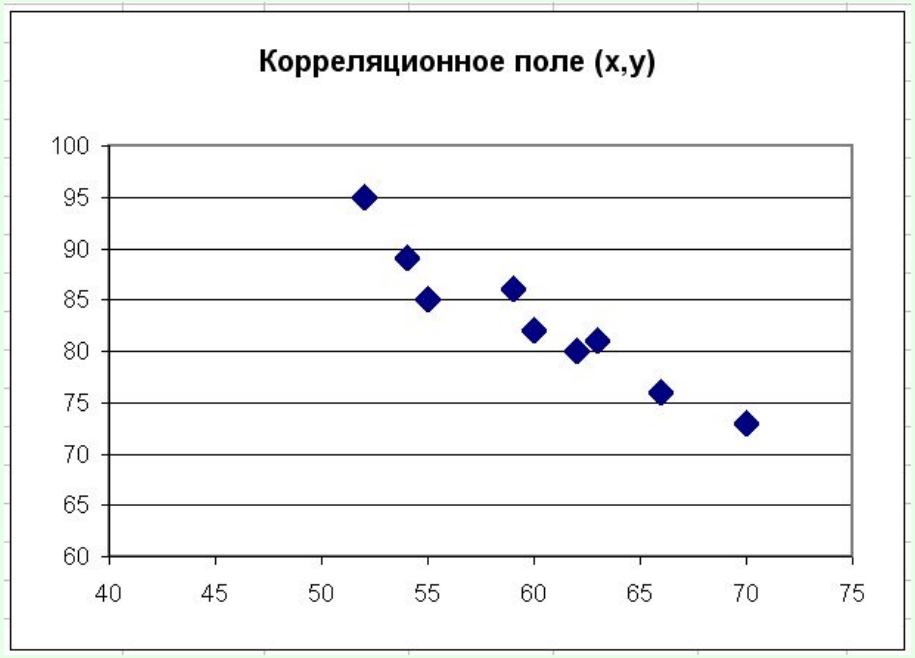
На графике видно, что между величинами есть сильная линейная связь



Как линейную связь оценить количественно?

С помощью линейного коэффициента корреляции

$$r_{xy} = \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_i (y_i - \bar{y})^2 \sum_i (x_i - \bar{x})^2}} = \frac{\mathrm{cov}(x, y)}{\sigma_x \cdot \sigma_y} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{y^2} - \bar{y}^2}\sqrt{\overline{x^2} - \bar{x}^2}}$$



Свойства линейного коэффициента корреляции:

1) $\left| r_{xy} \right| \le 1$

2) $r_{xy} > 0$: между $\boldsymbol{x}$ и $\boldsymbol{y}$ связь **прямая**,

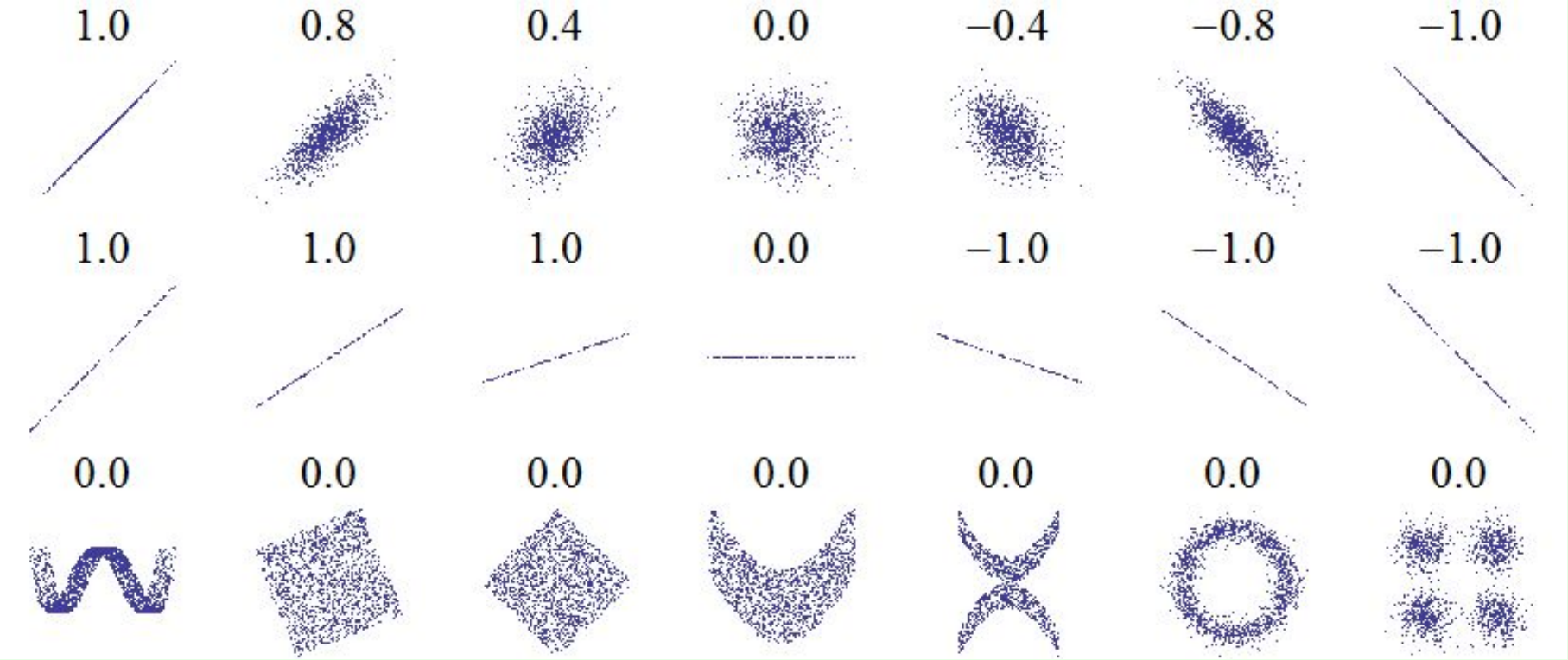
$r_{xy} < 0$: **обратная**

3) $\left| r_{xy} \right| > 0{,}7$: связь достаточно сильная

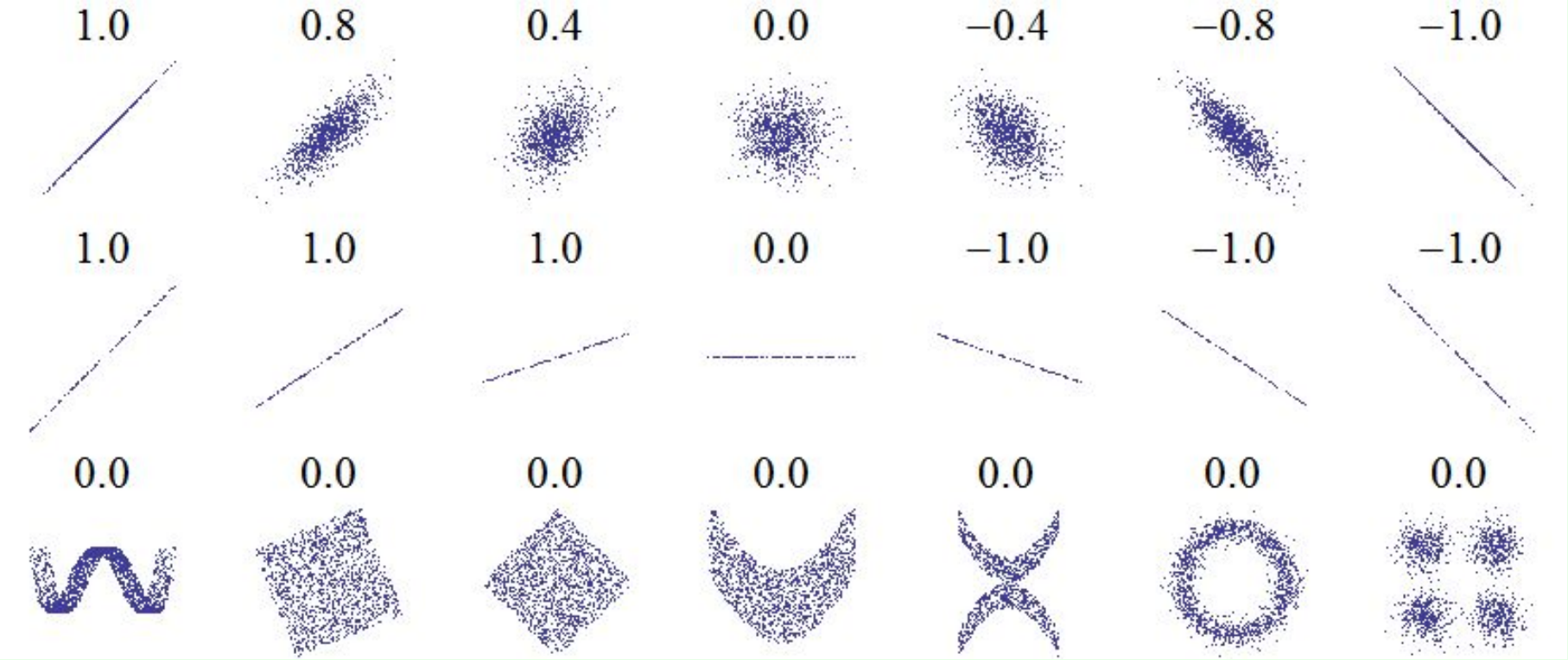
$\left| r_{xy} \right| < 0{,}3$: связь слабая

$\left| r_{xy} \right| \approx 0$: линейная связь отсутствует

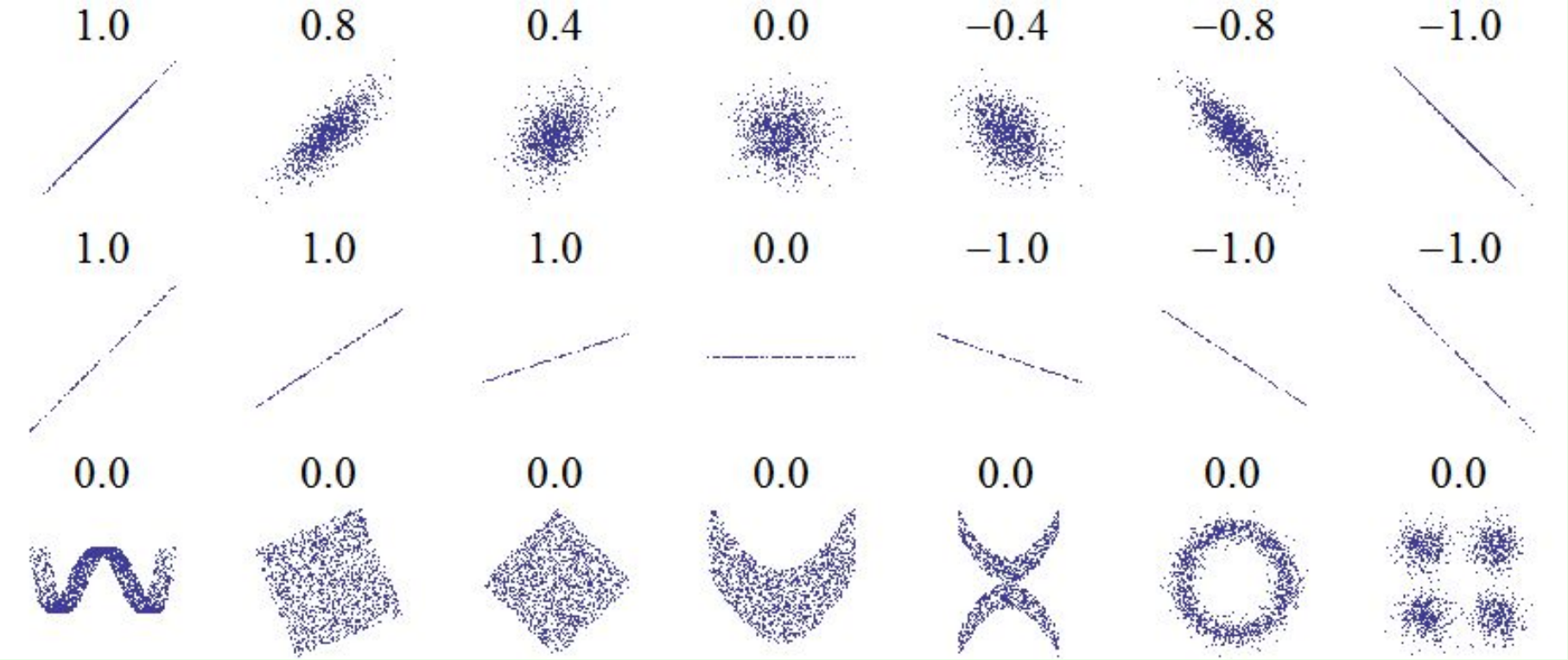
Различные графики распределений пар (x,y) с коэффициентами корреляции



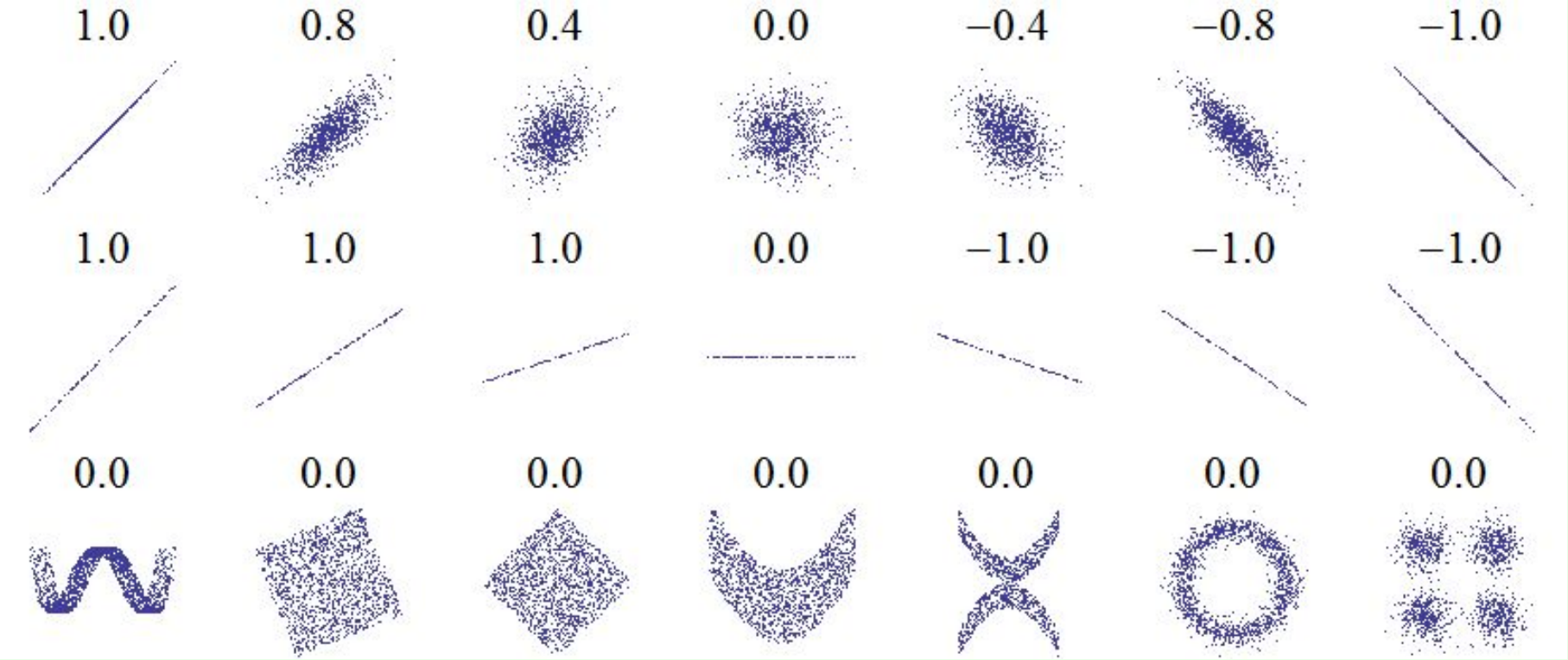
Коэффициент корреляции отражает линейную зависимость (**верхняя строка**)



но не описывает кривую зависимости (**средняя строка**)



и совсем не подходит для описания сложных, нелинейных зависимостей (**нижняя строка**)



Корреляция между переменными может оказаться ложной.

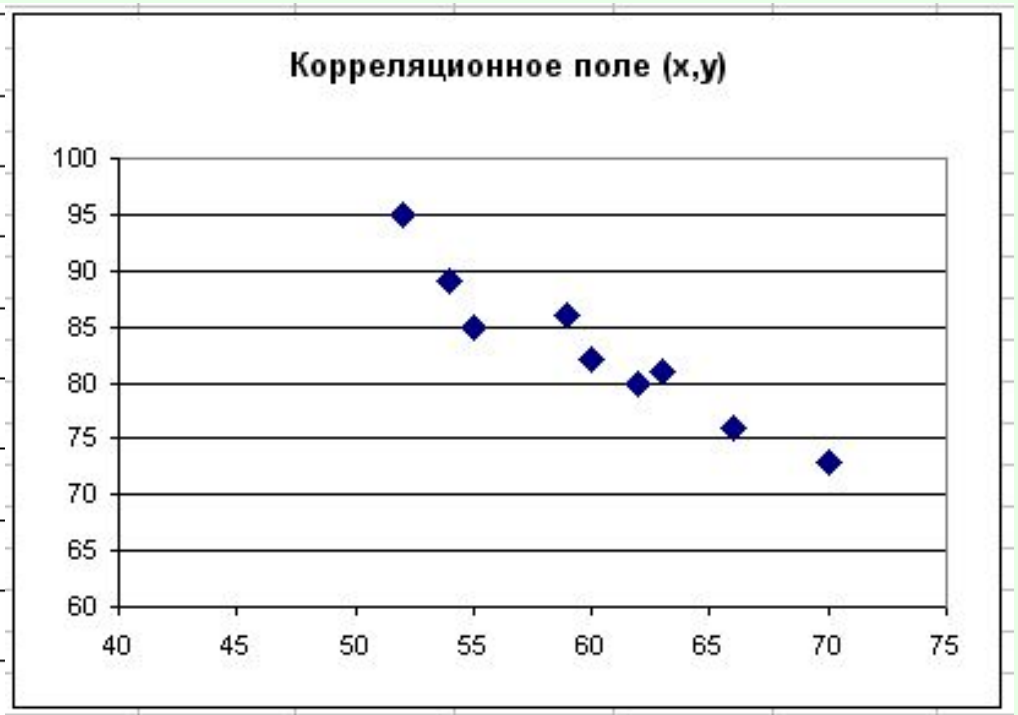
Например, из-за наличия у них **трендов**

Тренд – длительная тенденция изменения экономических показателей

Для выявления *ложной корреляции* можно измерить корреляцию не самих показателей **x** и **y**, а их первых разностей **Δx** и **Δy**.

Пример 1. Оценить влияние фактора x на показатель y.

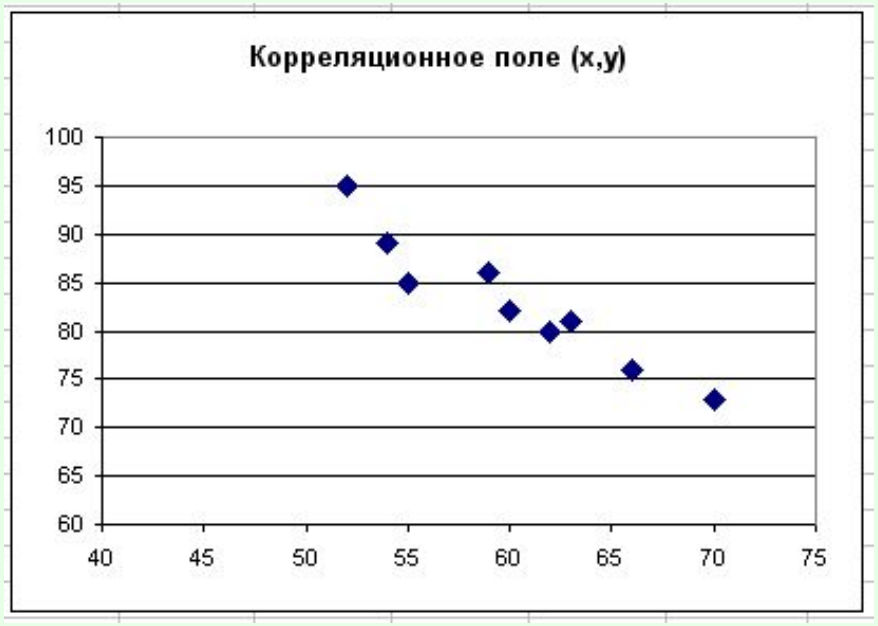
i	y	x			
1	95	52			
2	89	54			
3	85	55			
4	86	59			
5	82	60			
6	80	62			
7	81	63			
8	76	66			
9	73	70			
сумма	**747**	**541**	**44604**	**62357**	**32795**
среднее	**83**	**60,11**	**4956,00**	**6928,56**	**3643,89**



$$r_{xy} = \frac{\overline{y \cdot x} - \overline{y} \cdot \overline{x}}{\sqrt{\overline{y^2} - \overline{y}^2} \cdot \sqrt{\overline{x^2} - \overline{x}^2}} = \frac{4956 - 83 \cdot 60{,}11}{\sqrt{6928{,}56 - 83^2} \cdot \sqrt{3643{,}89 - 60{,}11^2}} = -0{,}9558$$

Пример 1. Оценить влияние фактора x на показатель y.

$$\left| r_{xy} \right| = 0{,}9558 > 0{,}7$$



Большое значение коэффициента корреляции говорит о сильной обратной линейной связи между рассматриваемыми количественными показателями. **Проверим, не является ли она ложной?**

Проверим, является ли корреляция ложной?

Найдем корреляцию между первыми разностями

i	y	x	$\Delta y=y_i-y_{i-1}$	$\Delta x=x_i-x_{i-1}$	$\Delta y\cdot\Delta x$	Δy^2	Δx^2
1	95	52					
2	89	54	-6	2	-12	36	4
3	85	55	-4	1	-4	16	1
4	86	59	1	4	4	1	16
5	82	60	-4	1	-4	16	1
6	80	62	-2	2	-4	4	4
7	81	63	1	1	1	1	1
8	76	66	-5	3	-15	25	9
9	73	70	-3	4	-12	9	16
сумма	**747**	**541**	**-22**	**18**	**-46**	**108**	**52**
среднее	**83**	**60,11**	**-2,75**	**2,25**	**-5,75**	**13,50**	**6,50**

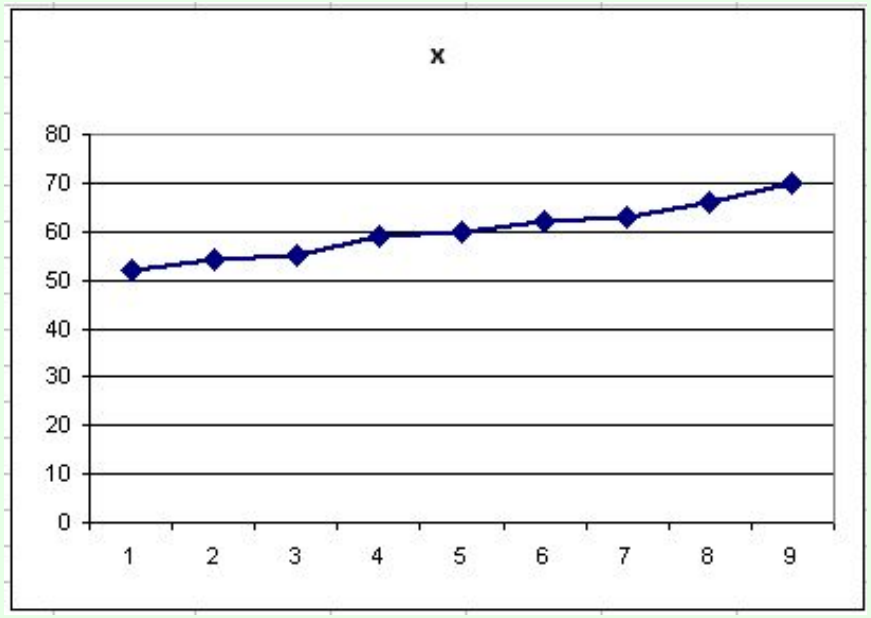
$$r_{xy}=\frac{\overline{\Delta y\cdot\Delta x}-\overline{\Delta y}\cdot\overline{\Delta x}}{\sqrt{\overline{\Delta y^2}-\overline{\Delta y}^2}\cdot\sqrt{\overline{\Delta x^2}-\overline{\Delta x}^2}}=\frac{-5{,}75-(-2{,}75)\cdot 2{,}25}{\sqrt{13{,}5-(-2{,}75)^2}\cdot\sqrt{6{,}5-2{,}25^2}}=0{,}1498<0{,}3$$

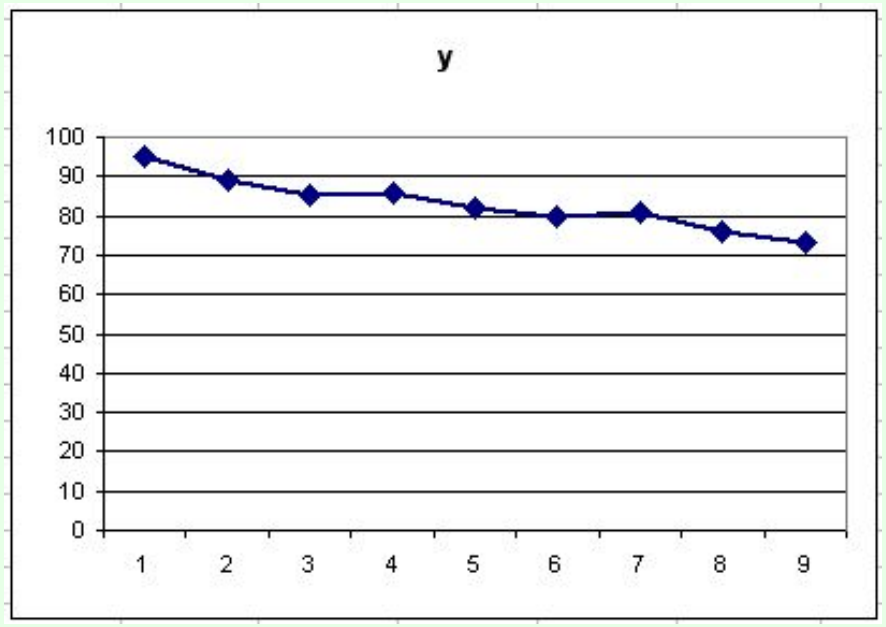
Корреляция между первыми разностями

$$r_{xy} = \frac{\overline{\Delta y \cdot \Delta x} - \overline{\Delta y} \cdot \overline{\Delta x}}{\sqrt{\overline{\Delta y^2} - \overline{\Delta y}^2} \cdot \sqrt{\overline{\Delta x^2} - \overline{\Delta x}^2}} = \frac{-5{,}75 - (-2{,}75) \cdot 2{,}25}{\sqrt{13{,}5 - (-2{,}75)^2} \cdot \sqrt{6{,}5 - 2{,}25^2}} = 0{,}1498 < 0{,}3$$

Получили маленькое значение, следовательно, сильная линейная связь между показателями *x* и *y* – является ложной.

Она является следствием наличия у них трендов



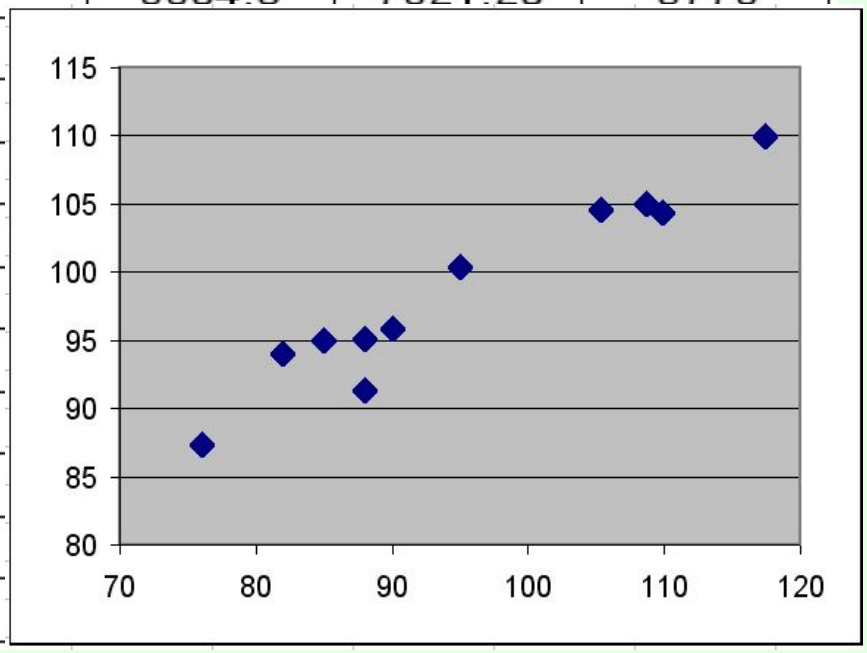


Пример 2. В таблице приведена динамика экономических показателей России: валовой внутренний продукт РФ (в процентах к предыдущему году) – показатель y (ВВП) и капитальные вложения в основные фонды РФ (в процентах к предыдущему году) – фактор x (КВОФ). Требуется оценить влияние фактора x на результативный признак y.

Год	ВВП (y,%)	КВОФ (x,%)
1991	95	85
1992	85,5	60
1993	91,3	88
1994	87,3	76
1995	95,8	90
1996	94	82
1997	100,4	95
1998	95,1	88
1999	104,6	105,3
2000	109,9	117,4
2001	105	108,7
2002	104,3	109,9

Рассчитаем коэффициент корреляции между переменными

Год	ВВП (y,%)	КВОФ (x,%)	$y \cdot x$	y^2	x^2
1991	95	85	8075	9025	7225
1992	85,5	60	5130	7310,25	3600
1993	91,3	88	8034,4	8335,69	7744
1994	87,3	76	6634,8	7621,29	5776
1995	95,8	90			
1996	94	82			
1997	100,4	95			
1998	95,1	88			
1999	104,6	105,3			
2000	109,9	117,4			
2001	105	108,7			
2002	104,3	109,9			
сумма	**1168,2**	**1105,3**			
среднее	**97,35**	**92,11**			



$$r_{xy} = \frac{\overline{y \cdot x} - \overline{y} \cdot \overline{x}}{\sqrt{\overline{y^2} - \overline{y}^2} \cdot \sqrt{\overline{x^2} - \overline{x}^2}} = \frac{9075{,}31 - 97{,}35 \cdot 92{,}11}{\sqrt{9529{,}39 - 97{,}35^2} \cdot \sqrt{8725{,}21 - 92{,}11^2}} = 0{,}9658$$

$$\left| r_{xy} \right| = 0{,}9658 > 0{,}7$$

Следовательно имеет место **сильная прямая линейная связь** между темпом роста валового внутреннего продукта РФ (ВВП) и темпом роста капитальных вложений в основные фонды РФ (КВОФ).

Проверим, не является ли она ложной?

Проверим, не является ли она ложной?

Год	ВВП (y,%)	КВОФ (x,%)	$\Delta y = y_i - y_{i-1}$	$\Delta x = x_i - x_{i-1}$	$\Delta y \cdot \Delta x$	Δy^2	Δx^2
1991	95	85					
1992	85,5	60	-9,5	-25	237,5	90,25	625
1993	91,3	88	5,8	28	162,4	33,64	784
1994	87,3	76	-4	-12	48	16	144
1995	95,8	90	8,5	14	119	72,25	196
1996	94	82	-1,8	-8	14,4	3,24	64
1997	100,4	95	6,4	13	83,2	40,96	169
1998	95,1	88	-5,3	-7	37,1	28,09	49
1999	104,6	105,3	9,5	17,3	164,35	90,25	299,29
2000	109,9	117,4	5,3	12,1	64,13	28,09	146,41
2001	105	108,7	-4,9	-8,7	42,63	24,01	75,69
2002	104,3	109,9	-0,7	1,2	-0,84	0,49	1,44
сумма	**1168,2**	**1105,3**	**9,3**	**24,9**	**971,87**	**427,27**	**2553,83**
среднее	**97,35**	**92,11**	**0,85**	**2,26**	**88,35**	**38,84**	**232,17**

$$r_{xy} = \frac{\overline{\Delta y \cdot \Delta x} - \overline{\Delta y} \cdot \overline{\Delta x}}{\sqrt{\overline{\Delta y^2} - \overline{\Delta y}^2} \cdot \sqrt{\overline{\Delta x^2} - \overline{\Delta x}^2}} = \frac{88{,}35 - 0{,}85 \cdot 2{,}26}{\sqrt{38{,}84 - 0{,}85^2} \cdot \sqrt{232{,}17 - 2{,}26^2}} = 0{,}9290 > 0{,}7$$

$$r_{xy} = \frac{\overline{\Delta y \cdot \Delta x} - \overline{\Delta y} \cdot \overline{\Delta x}}{\sqrt{\overline{\Delta y^2} - \overline{\Delta y}^2} \cdot \sqrt{\overline{\Delta x^2} - \overline{\Delta x}^2}} = \frac{88{,}35 - 0{,}85 \cdot 2{,}26}{\sqrt{38{,}84 - 0{,}85^2} \cdot \sqrt{232{,}17 - 2{,}26^2}} = 0{,}9290 > 0{,}7$$

Сильная линейная связь первых разностей переменных

Значит сильная прямая связь между темпом роста валового внутреннего продукта РФ и темпом роста капитальных вложений в основные фонды РФ, по-видимому, **не является ложной.**