ОСНОВНІ ПОНЯТТЯ РЕГРЕСІЙНОГО АНАЛІЗУ

№1 По статистическим данным о валовом выпуске продукции за 12 лет *Y*, и об основных производственных фондах *X*, приведенные в таблице. Требуется составить корреляционную таблицу наблюдаемых данных и найти выборочные уравнения линейной регрессии *Y* на *X* та *X* на *Y* и построить прямые регрессии. (Считать зависимость линейной)

год	X	Y
1	460	618
2	510	615
3	268	645
4	325	641
5	300	554
6	500	493
7	505	510
8	250	398
9	285	455
10	315	600
11	480	625
12	500	614

Построим статистическое распределение выборки для величин X и Y отдельно:

№	n_i	x_i
1.	1	250
2.	1	268
3.	1	285
4.	1	300
5.	1	315
6.	1	325
7.	1	460
8.	1	480
9.	2	500
10.	1	505
11.	1	510
12.		

n_i	y_i
1	398
1	455
1	493
1	510
1	554
1	600
1	614
1	615
1	618
1	625
1	641
1	645

Разобьем интервал для X и Y на 5 равных подинтервалов. Сюда ж отнесем статистическое распределение для усредненных вариант x_i^* та y_i^*.

Размах вариаций: $R_x = 510 - 250 = 260$, $R_y = 645 - 398 = 247$.

Длины подинтервалов: $h_x = \dfrac{R_x}{5} = 52$, $h_y = \dfrac{R_y}{5} = 49{,}4$.

Соответствующие усредненные варианты: $x_i^* = \dfrac{x_i + x_{i+1}}{2}$, $y_i^* = \dfrac{y_i + y_{i+1}}{2}$.

Будем иметь:

– для X

x_i	x_{i+1}	x_i^*	n_i

250,0	302,0	276,0	4
302,0	354,0	328,0	2
354,0	406,0	380,0	0
406,0	458,0	432,0	0
458,0	510,0	484,0	6

– для Y

y_i	y_{i+1}	y_i^*	n_i
398,0	447,4	422,7	1
447,4	496,8	472,1	2
496,8	546,2	521,5	1
546,2	595,6	570,9	1
595,6	645,0	620,3	7

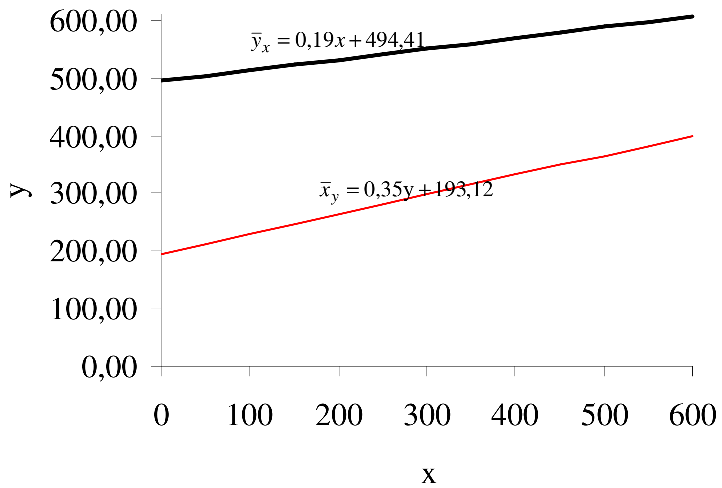
Построим корреляционную таблицу:

Y \ X	276,0	328,0	380,0	432,0	484,0	n_y
422,7	1					1
472,1	1				1	2
521,5					1	1
570,9	1					1
620,3	1	2			4	7
n_x	4	2	0	0	6	12

Расчет коэффициента корреляции проводим по серединам интервалов по формуле:

$$r^* = \bar{r}_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = 0{,}25.$$

Уравнения прямых линий регрессии Y на X и X на Y запишем, исходя из формул $\bar{y}_x - \bar{y} = \bar{r}_{xy} \dfrac{\sigma_y}{\sigma_x}(x - \bar{x})$ и $\bar{x}_y - \bar{x} = \bar{r}_{xy} \dfrac{\sigma_x}{\sigma_y}(y - \bar{y})$. Учитывая предыдущие вычисления, получим такие выборочные уравнения регрессии: $\bar{y}_x = 0{,}19x + 494{,}41$ и $\bar{x}_y = 0{,}35\text{y} + 193{,}12$. Построим их:



№2 Ранее был рассмотрен пример, касающийся установления связи между удельным электрическим сопротивлением $\rho_{\text{уд.}}$ константой изоляции k.

№	X	Y
	$\rho_{\text{уд.}}$	k

1	$5{,}0\cdot 10^{13}$	90,0
2	$1{,}5\cdot 10^{13}$	36,0
3	$9{,}8\cdot 10^{13}$	63,8
4	$8{,}0\cdot 10^{11}$	7,1
5	$7{,}5\cdot 10^{11}$	1,2
6	$3{,}9\cdot 10^{11}$	1,8
7	$7{,}8\cdot 10^{11}$	5,9
8	$7{,}0\cdot 10^{11}$	1,5

Найдено, что коэффициент корреляции равен $\bar{r}_{xy} = 0{,}718$, было показано, что он значим, т.е. связь между $\rho_{\text{уд.}}$ и k близка к линейной. Для определения уравнения этой связи воспользуемся методом наименьших квадратов:

$$a = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = 9{,}516; \qquad b = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = 7{,}88\cdot 10^{-13}.$$

Значит уравнение связи: $y = 9{,}516 + 7{,}88\cdot 10^{-13} x$.

Для проверки адекватности уравнения необходимо вычислять остаточную дисперсию:

$$S_R^2 = \frac{\sum_{i=1}^{n} (y_i - \tilde{y}_i)^2}{n-2} = 445{,}5; \qquad S_R = 21{,}1.$$

Поскольку нет информации об ошибке в определении y для проверки адекватности уравнения можно воспользоваться лишь сравнением остаточной дисперсии общей дисперсией y в эксперименте: $S_R^2 = 445{,}5$; $S_R = 21{,}1$.

Сравнение показывает, что надежность уравнения сравнительно невысокая, т.к. для адекватности необходимо, чтобы S_R^2 было меньше S_y^2, чем в 10 раз.