PAIR LINEAR REGRESSION EQUATION: $\tilde{y}_x = b_0 + b_1 x$

Example. Investigate the dependence between the variables x and y, if data are given as:

x_i	1	1,5	2	2,5	3
y_i	2,15	2,3	2,6	2,8	2,5

STEP 1. Let's fill in this table:

№	x_i	y_i	x_i^2	$x_i y_i$	y_i^2	$\tilde{y}_i$	$y_i - \tilde{y}_i$	$(y_i - \tilde{y}_i)^2$	$\frac{y_i - \tilde{y}_i}{y_i}$	$\left\lvert \frac{y_i - \tilde{y}_i}{y_i} \right\rvert$
1	**1**	**2,15**	**1**	**2,15**	**4,6225**	**2,23**	**-0,08**	**0,0064**	**-0,0372**	**0,0372**
2	**1,5**	**2,3**	**2,25**	**3,45**	**5,29**	**2,35**	**-0,05**	**0,0025**	**-0,0217**	**0,0217**
3	**2**	**2,6**	**4**	**5,2**	**6,76**	**2,47**	**0,13**	**0,0169**	**0,05**	**0,05**
4	**2,5**	**2,8**	**6,25**	**7**	**7,84**	**2,59**	**0,21**	**0,0441**	**0,075**	**0,075**
5	**3**	**2,5**	**9**	**7,5**	**9**	**2,71**	**-0,21**	**0,0441**	**-0,084**	**0,084**
Σ	**10**	**12,35**	**22,5**	**25,3**	**33,5125**			**0,114**		**0,2679**

STEP 2. Let's find numerical characteristics:

$$\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{10}{5} = 2, \qquad \bar{y} = \frac{1}{5}\sum_{i=1}^{5} y_i = \frac{12{,}35}{5} = 2{,}47,$$

$$\overline{xy} = \frac{1}{5}\sum_{i=1}^{5} x_i y_i = \frac{25{,}3}{5} = 5{,}06, \qquad \overline{x^2} = \frac{1}{5}\sum_{i=1}^{5} x_i^2 = \frac{22{,}5}{5} = 4{,}5,$$

$$\overline{y^2} = \frac{33{,}5125}{5} = 6{,}7025, \ \sigma_x = \sqrt{\sigma_x^2} = \sqrt{0{,}5} = 0{,}71$$

$$\sigma_x^2 = \overline{x^2} - (\bar{x})^2 = 4{,}5 - 2^2 = 0{,}5, \qquad \sigma_x = \sqrt{\sigma_x^2} = \sqrt{0{,}5} = 0{,}71$$

$$\sigma_y^2 = \overline{y^2} - (\bar{y})^2 = 6{,}7025 - 2{,}47^2 = 0{,}6016, \qquad \sigma_y = \sqrt{\sigma_y^2} = \sqrt{0{,}6016} = 0{,}7756,$$

$$\mu_{xy} = \overline{xy} - \bar{x} \cdot \bar{y} = 5{,}06 - 2 \cdot 2{,}47 = 0{,}12$$

STEP 3. Let's calculate coefficients:

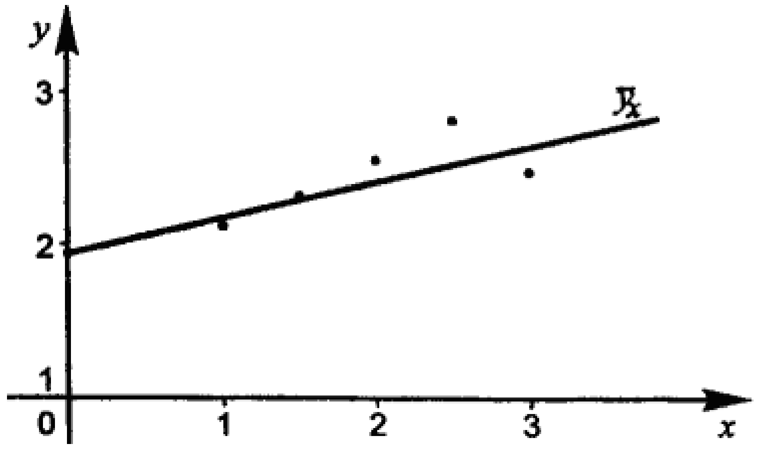
$$b_1 = \frac{\mu_{xy}}{\sigma_x^2} = \frac{0{,}12}{0{,}5} = 0{,}24 \qquad b_0 = \bar{y} - b_1 \cdot \bar{x} = 2{,}47 - 0{,}24 \cdot 2 = 1{,}99$$

The theoretical regression equation is $\tilde{y}_x = 1{,}99 + 0{,}24x$ (**the regression equation *of* y *upon* x**)

Explanation: the coefficient $b_1 = 0{,}24$ shows the increasing X by 1 unit gives the increasing Y by 0,24 units.

STEP 4. Let's plot the empirical data and the regression equation:

x_i	$\tilde{y}_i$
1	**2,23**
3	**2,71**



STEP 5. The correlation coefficient:

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \cdot \sigma_y} = \frac{0{,}12}{0{,}71 \cdot 0{,}2272} = 0{,}7471$$

Since this coefficient ***r*** is greater than 0.7 ($|r_{xy}| > 0{,}7$), then this linear correlation is strong (сильная).

STEP 6. The *mean error*:

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \tilde{y}_i}{y_i}\right| \cdot 100\% \qquad \text{or} \qquad \overline{A} = \frac{1}{5} \cdot 0,2679 \cdot 100\% \approx 0,0536 \cdot 100\% = 5,36\%$$

The *mean error* $\overline{A}$ is less than 10% (the allowable limit) then we can use this equation for a forecasting.

STEP 7. The determination coefficient:

$$R^2 = r_{xy}^2 = 0,7471^2 = 0,5582$$

It means that 55,82% of the total variation in ***y*** can be explained by the linear relationship between ***x*** and ***y*** (as described by the regression equation). The other 44,18% of the total variation in ***y*** remains unexplained.

STEP 8. F-test (Fisher test) for verification of a statistical significance of an equation:

$$F_{tabl} = \frac{r_{xy}^2}{1 - r_{xy}^2}(n-2)$$

$$F_{emp} = \frac{0,7471^2}{1 - 0,7471^2}(5-2) \approx 3,79$$

The tabular value: $F_{tabl}(k_1; k_2) = F_{tabl}(1; n-2) = F_{tabl}(1; 5-2) = F_{tabl}(1; 3) = 10,1$

Let's compare with the tabular value: $F_{tabl} > F_{emp}$ or $10{,}1 > 3{,}79$

Conclusion: the hypothesis H_0 is accepted, i.e. the statistical nonsignificance of this regression equation is accepted with 95%..

STEP 9. The variance and the root-mean square deviation of residuals (остаточная дисперсия):

$$\sigma^2_{residuals} = \frac{\sum_i (y_i - \tilde{y}_i)^2}{n-2} = \frac{0{,}114}{5-2} = 0{,}038$$

$$\sigma_{residuals} = \sqrt{\sigma^2_{residuals}} = \sqrt{0{,}038} = 0{,}195$$

STEP 10. The forecasting value: let's take x=4

$$\tilde{y}_{x=4} = 1{,}99 + 0{,}24 \cdot 4 = 2{,}95$$

Fisher table

$\alpha = 0{,}05$

k_2	k_1											
	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,5	19,0	19,2	19,2	19,3	19,3	19,3	19,4	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,13	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,16	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,!0	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60