PhD Misiura Ie.Iu. (доцент Місюра Є.Ю.)

Theme: Relationship of random variables in economics. Correlation dependence. Elements of regression analysis. Forecasting the characteristics of the foreign trade market

Part 1. Correlation theory

Let’s start the study of dependences between variables **with two variables**, which are denoted as X and Y. There are two basic types of dependences: functional and stochastic (statistical).

A functional dependence between two variables ***x*** and ***y*** is the defined value of y corresponds to each value of x, and the equation $y = f(x)$ takes place.

At a functional dependence graphs of equations

y=f(x) and **x=φ(y)**

are identical, i.e. it’s not important that x or y will be dependent or independent variable. Thus, a functional dependence does not respond to a direction of cause-and-effect relations (причинно-слідчі зв’язки).

Example 1. Let’s take the equation

$$y = 2x - 4$$

Hence $x = \frac{y}{2} + 2$

Graphs of these lines coincide. Both straight lines pass through points

x=0, y=-4 and x=2, y=0

For random variables we can or can’t establish a functional dependence. There exists a dependence between such variables, at which with a change of one variable a distribution of the other variable is changed. Such a dependence is called ***stochastic (statistical)***.

There are **two components at a statistical dependence**:

1) stochastic, i.e. it is directly related with a dependence between a functional factor and a factor-argument;

2) random, i.e. it is related with an influence of random factors on a dependence between a functional factor and a factor-argument.

A lack of the second component leads to a functional dependence.

At statistical relationships **cause-and-effect relations (причинно-слідчі зв'язки) are important**, i.e. it is important that a variable will be a functional factor or an independent factor.

The particular case of a statistical dependence is ***a correlation dependence*** between two random variables, at which with a change of one out of two variable a mean value of the other variable is changed.

Basic problems of correlation theory. A correlation table, a correlation field, empirical regression lines

Correlation theory **solves two basic problems**:

1. A defining a presence of correlation (if values of $\overline{y}_x$ are same for all values of ***x***, then a correlation dependence is absent), an establishment of a form of a dependence (a type of a regression function) and defining parameters a regression equation.

Correlation theory **solves two basic problems**:

2. A defining ***a strength (міснота)*** of a correlation. A strength of a correlation is defined as a measure of values diffusion (розсіювання) relative to a conditional mean and characterizes a degree of an influence of a changing value **X** on a changing value **Y**.

For defining a presence of a correlation we fill a correlation table using empirical data. ***A correlation table*** is formed for grouped observations (згрупованих спостережень) and contains intervals (classses) of a changing variables **X** and **Y** and frequencies of an joint appearance (сумісна поява) of the given pair values of ***x*** and **y** (see a table).

A correlation table

Intervals of Y \ Intervals of X	y_k \ x_i	$x_0^* - x_1^*$	$x_1^* - x_2^*$	$\ldots$	$x_{i-1}^* - x_i^*$	$\ldots$	$x_{l-1}^* - x_l^*$	m_y
		x_1	x_2	$\ldots$	x_i	$\ldots$	x_l	
$y_0^* - y_1^*$	y_1	m_{11}	m_{12}	$\ldots$	m_{1i}	$\ldots$	m_{1l}	m_{y_1}
$y_1^* - y_2^*$	y_2	m_{21}	m_{22}	$\ldots$	m_{2i}	$\ldots$	m_{2l}	m_{y_2}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$y_{k-1}^* - y_k^*$	y_k	m_{k1}	m_{k2}	$\ldots$	m_{ki}	$\ldots$	m_{kl}	m_{y_k}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$y_{s-1}^* - y_s^*$	y_s	m_{s1}	m_{s2}	$\ldots$	m_{si}	$\ldots$	m_{sl}	m_{y_s}
m_{x_i}		m_{x_1}	m_{x_2}	$\ldots$	m_{x_i}	$\ldots$	m_{x_l}	n

Let’s denote

n as the total number of observations;

m$_{ki}$ as the frequency of an joint appearance of two random variables *x* and *y* , besides

$$\sum_k \sum_i m_{ki} = n$$

Let's fill horizontal cells of a correlation table as intervals of the factor ***X***, vertical cells as intervals of a functional factor ***Y***. The frequency m_{ki} is located on the intersection of a column x_i and a row y_k. Let's denote x_i and y_k as midpoints of intervals.

The last column and row are defined as the total sums of rows and columns

$$m_{x_i} = \sum_{k=1}^{s} m_{ki} \qquad m_{y_k} = \sum_{i=1}^{l} m_{ki}$$

besides

$$\sum_{i=1}^{l} m_{x_i} = \sum_{k=1}^{s} m_{y_k} = n$$

Let’s use a correlation table and get the correspondence between each value of one out of variables ***X*** or ***Y*** and a distribution row of the other variable.

For example, for $X = x_1$ we have a distribution row:

y	y_1	y_2	$\cdots$	y_k	$\cdots$	y_s
m	m_{11}	m_{21}	$\cdots$	m_{k1}	$\cdots$	m_{s1}

Let's use a correlation table and get the correspondence between each value of one out of variables ***X*** or ***Y*** and a distribution row of the other variable.

For example, for $X = x_2$ we have a distribution row:

y	y_1	y_2	$\dots$	y_k	$\dots$	y_s
m	m_{12}	m_{22}	$\dots$	m_{k2}	$\dots$	m_{s2}

For example, for $X = x_1$ we have a distribution row:

y	y_1	y_2	...	y_k	...	y_s
m	m_{11}	m_{21}	...	m_{k1}	...	m_{s1}

Let's calculate the conditional mean:

$$\overline{y}_{x=x_1} = \frac{y_1 m_{11} + y_2 m_{21} + ... + y_k m_{k1} + ... + y_s m_{s1}}{m_{11} + m_{21} + ... + m_{k1} + ... + m_{s1}} = \frac{\sum_{k=1}^{s} y_k m_{k1}}{m_{x_1}}$$

For example, for $X = x_2$ we have a distribution row:

y	y_1	y_2	...	y_k	...	y_s
m	m_{12}	m_{22}	...	m_{k2}	...	m_{s2}

Let’s calculate the conditional mean:

$$\overline{y}_{x=x_2} = \frac{y_1 m_{12} + y_2 m_{22} + \ldots + y_k m_{k2} + \ldots + y_s m_{s2}}{m_{11} + m_{22} + \ldots + m_{k2} + \ldots + m_{s2}} = \frac{\sum_{k=1}^{s} y_k m_{k2}}{m_{x_2}}$$

Let's generalize:

$$\overline{y}_{x=x_1} = \frac{y_1 m_{11} + y_2 m_{21} + \ldots + y_k m_{k1} + \ldots + y_s m_{s1}}{m_{11} + m_{21} + \ldots + m_{k1} + \ldots + m_{s1}} = \frac{\sum_{k=1}^{s} y_k m_{k1}}{m_{x_1}}$$

$$\overline{y}_{x=x_2} = \frac{y_1 m_{12} + y_2 m_{22} + \ldots + y_k m_{k2} + \ldots + y_s m_{s2}}{m_{11} + m_{22} + \ldots + m_{k2} + \ldots + m_{s2}} = \frac{\sum_{k=1}^{s} y_k m_{k2}}{m_{x_2}}$$

and get:

$$\overline{y}_{x=x_i} = \frac{\sum_{k=1}^{s} y_k m_{ki}}{m_{x_i}}$$

Let's generalize and get: $\overline{y}_{x=x_i} = \dfrac{\sum_{k=1}^{s} y_k m_{ki}}{m_{x_i}}$

Thus, we have the dependence:

x	x_1	x_2	...	x_i	...	x_l
$\overline{y}_x$	$\overline{y}_{x_1}$	$\overline{y}_{x_2}$	...	$\overline{y}_{x_i}$	...	$\overline{y}_{x_l}$

Thus, we have the dependence:

x	x_1	x_2	...	x_i	...	x_l
$\overline{y}_x$	$\overline{y}_{x_1}$	$\overline{y}_{x_2}$	...	$\overline{y}_{x_i}$	...	$\overline{y}_{x_l}$

Like the previous variable $Y = y_k$ the conditional mean $\overline{x}_{y=y_k}$ is calculated as:

$$\overline{x}_{y=y_k} = \frac{\sum_{i=1}^{l} x_i m_{ki}}{m_{y_k}}$$

Like the previous variable $Y = y_k$ the conditional mean $\bar{x}_{y=y_k}$ is calculated as:

$$\bar{x}_{y=y_k} = \frac{\sum_{i=1}^{l} x_i m_{ki}}{m_{y_k}}$$

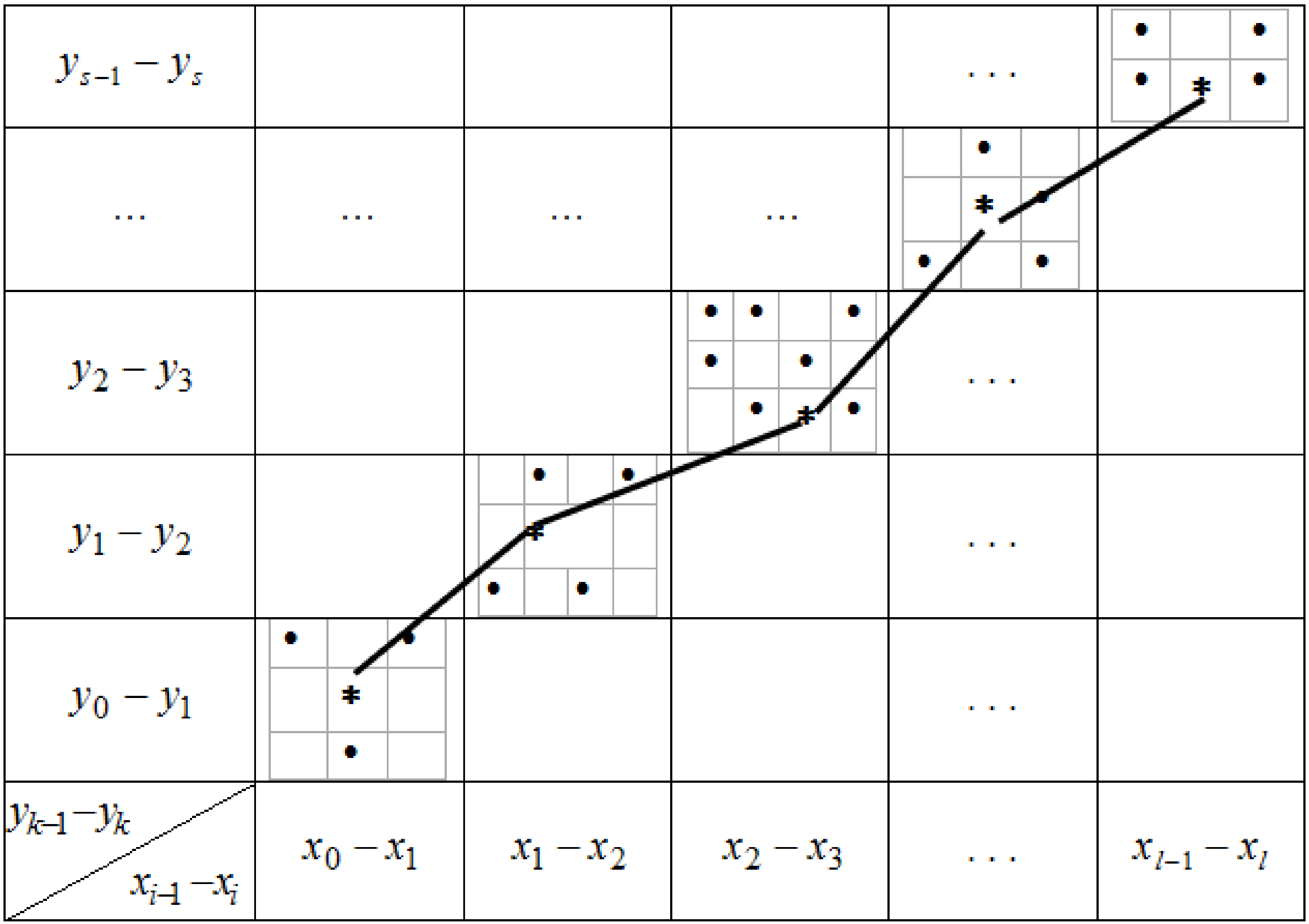
Thus, we have the dependence:

y	y_1	y_2	...	y_k	...	y_s
$\bar{x}_y$	$\bar{x}_{y_1}$	$\bar{x}_{y_2}$	...	$\bar{x}_{y_k}$	...	$\bar{x}_{y_s}$

Correlation tables give the possibility to define a presence of a correlation dependence: if $\overline{y}_x$ $\left(\overline{x}_y\right)$ are changed from a column to a column (from a row to a row), then there exists a correlation between variables ***X*** and ***Y***.

A correlation field (кор.поле) is a graphic presentation of a correlation table. A correlation field is called a point graph, where each point is a result of a separate observations of values of two variables.

Example


$y_{s-1} - y_s$
$\dots$
$y_2 - y_3$
$y_1 - y_2$
$y_0 - y_1$
$y_{k-1} - y_k$
$x_{i-1} - x_i$
$x_0 - x_1$
$x_1 - x_2$
$x_2 - x_3$
$\dots$
$x_{l-1} - x_l$

A correlation field can be plotted with the help of a correlation table.

Let’s write down intervals of variables ***X*** and ***Y***, в we mark points into each cell which is locates as the intersection of a column and a row, the number of points equals values m_{ki}. If we have clouds (згустки, центри групування), which decentre, then we can define a presence of a correlation between variables ***X*** and ***Y***.

Let's mark points $\left(x_i, \overline{y}_{x_i}\right)$ or $\left(\overline{x}_{y_k}, y_k\right)$ and connect them with the help of polygonal line.

The first polygonal line $\left(\overline{y}_x\right)$, which connects points with coordinates $\left(x_i, \overline{y}_{x_i}\right)$, is called ***an empirical regression line Y upon X***.

The second polygonal line $\left(\overline{x}_y\right)$, which connects points with coordinates $\left(\overline{x}_{y_k}, y_k\right)$, is called ***an empirical regression line X upon Y***.

Let’s use a notion **«a conditional mean»** $\overline{y}_{x_i}$

Let’s take a random variable **Y**, related with a variable **X**, and each value of **X** corresponds to some value of **Y**.

Example

Y \ X	2	4	6	8	m_y	$\overline{x_{y_i}}$
1	1	2				
3	1	6	3			
5		5	5			
7			1	1		
m_x						
$\overline{y_{x_j}}$						

Example.

Let's calculate frequencies m_x and m_y

Y \ X	2	4	6	8	m_y	$\overline{x_{y_i}}$
1	1	2			3	
3	1	6	3		10	
5		5	5		10	
7			1	1	2	
m_x	2	13	9	1	25	
$\overline{y_{x_j}}$						

For example, at $x = 2$ the variable **y** takes values: $y_1 = 1$, $y_2 = 3$.

Then arithmetic mean **Y**, which corresponds to $x = 2$:

$$\overline{y}_{x=2} = \frac{1 \cdot 1 + 3 \cdot 1}{2} = \frac{4}{2} = 2$$

Example

Y \ X	2	4	6	8	m_y	$\overline{x}_{y_i}$
1	1	2			3	
3	1	6	3		10	
5		5	5		10	
7			1	1	2	
m_x	2	13	9	1	25	
$\overline{y}_{x_j}$						

For example, at $x=4$ the variable **y** takes values: $y_1=1$, $y_2=3$, $y_3=5$.

Then arithmetic mean **Y**, which corresponds to $x=4$:

$$\overline{y}_{x=4}=\frac{1\cdot 2+3\cdot 6+5\cdot 5}{13}=\frac{45}{13}=3{,}5$$

For example, at $x = 6$ the variable **y** takes values: $y_2 = 3$, $y_3 = 5$, $y_4 = 7$.

Then arithmetic mean **Y**, which corresponds to $x = 6$:

$$\overline{y}_{x=6} = \frac{3 \cdot 3 + 5 \cdot 5 + 7 \cdot 1}{9} = \frac{41}{9} = 4{,}6$$

For example, at $x = 8$ the variable **y** takes values: $y_4 = 7$.

Then arithmetic mean **Y**, which corresponds to $x = 8$:

$$\overline{y}_{x=8} = \frac{7 \cdot 1}{1} = 7$$

Y \ X	2	4	6	8	m_y	$\overline{x_{y_i}}$
1	1	2				
3	1	6	3			
5		5	5			
7			1	1		
m_x	**2**	**13**	**9**	**1**	**25**	
$\overline{y_{x_j}}$	**2**	**3,5**	**4,6**	**7**		

A correlation dependence Y on X is called a functional dependence of a conditional mean $\overline{y}_x$ on x :

$$\overline{y}_x = f(x)$$

A conditional mean $\bar{x}_{y_j}$ is called an arithmetic mean of values **X**, which correspond to

$$y = y_j$$

For example, at $y = 1$ the variable **y** takes values: $x_1 = 2,\ x_2 = 4$.

Then arithmetic mean **Y**, which corresponds to $y = 1$:

$$\bar{x}_{y=1} = \frac{2 \cdot 1 + 4 \cdot 2}{3} = \frac{10}{3} = 3{,}3$$

Task. Find:

$$\bar{x}_{y=3} = ?$$

$$\bar{x}_{y=5} = ?$$

$$\bar{x}_{y=7} = ?$$

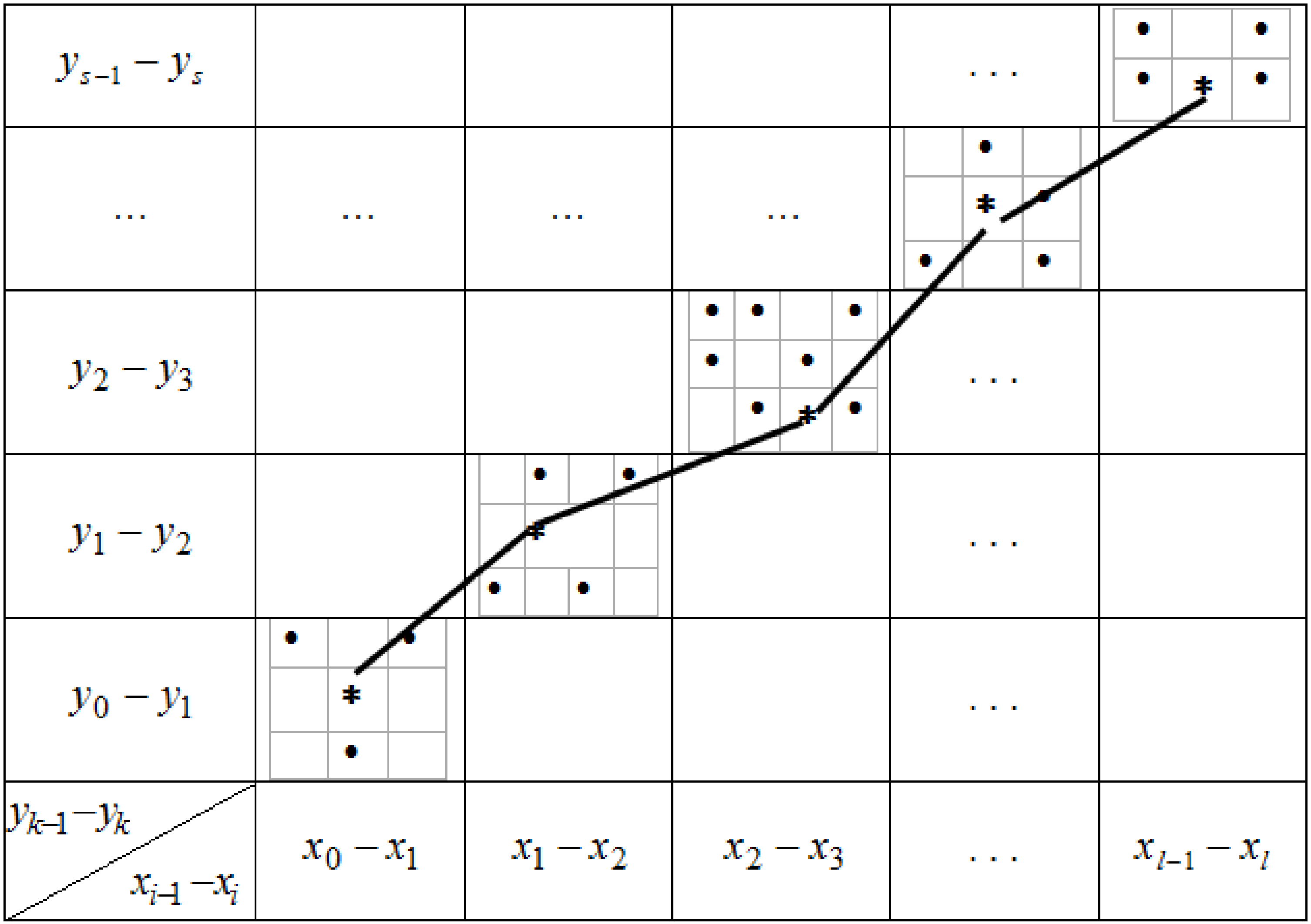
According to results of the same trails dependences can be obtained: $\overline{y}_x$ and $\overline{x}_y$ i.e. it is necessary to indicate **a factor-argument (cause, причина)** and **a functional factor (слідство, effect).**

The equation $\overline{y}_x = f(x)$ is called **the *empirical* regression equation *of Y upon X***, and the function $f(x)$ **the regression of *Y upon X***, and its graph is ***the empirical regression line of Y upon X***.

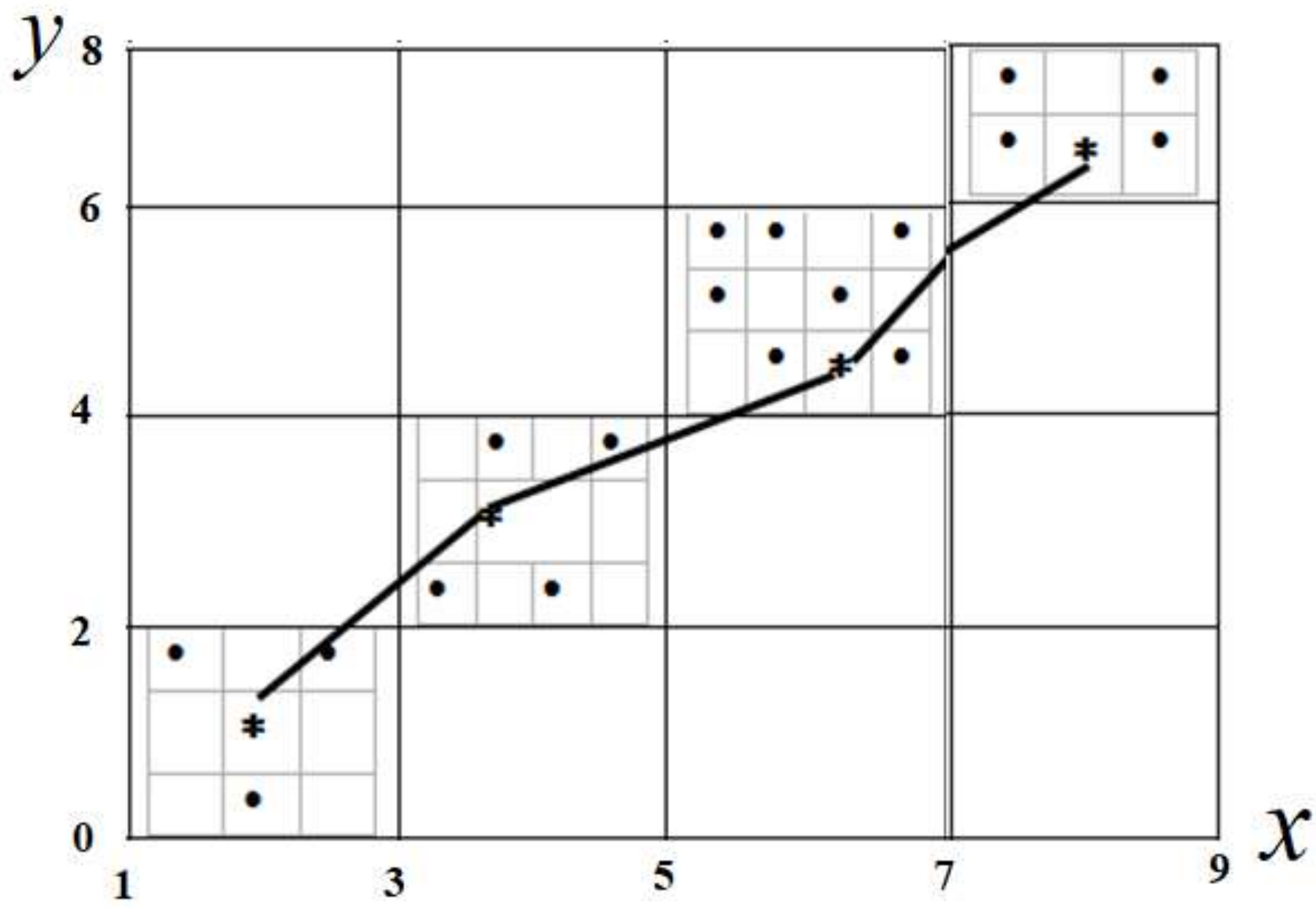
The equation $\overline{y}_x = f(x)$ is called **the empirical regression equation *of Y upon X***, and the function $f(x)$ **the regression of *Y upon X***, and its graph is ***the regression line of Y upon X***.

We can use the other equation: $\overline{x}_y = \varphi(y)$. It is **the empirical regression equation *of X upon Y.***

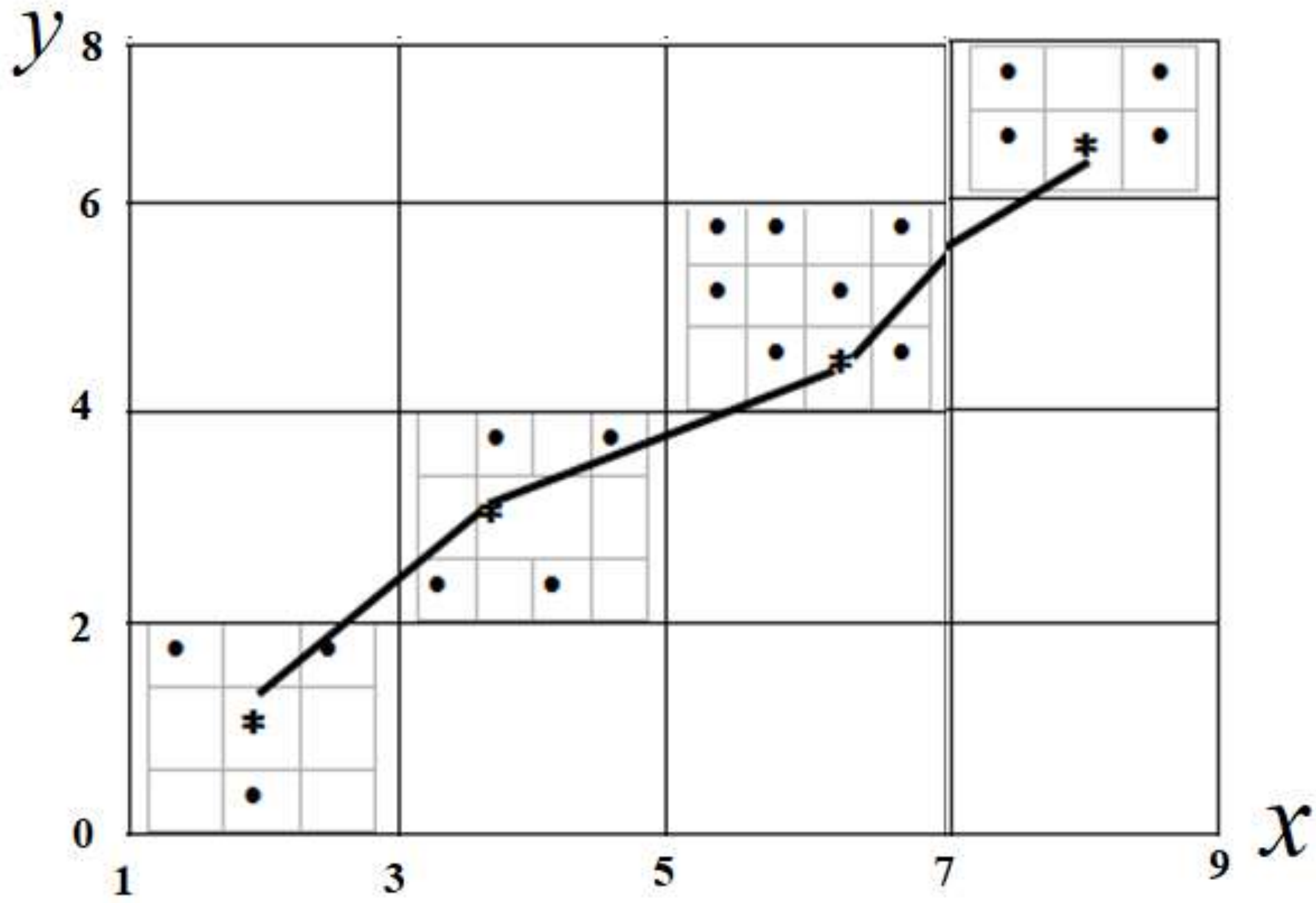
Example



Example. Let's define h_x=4-2=2

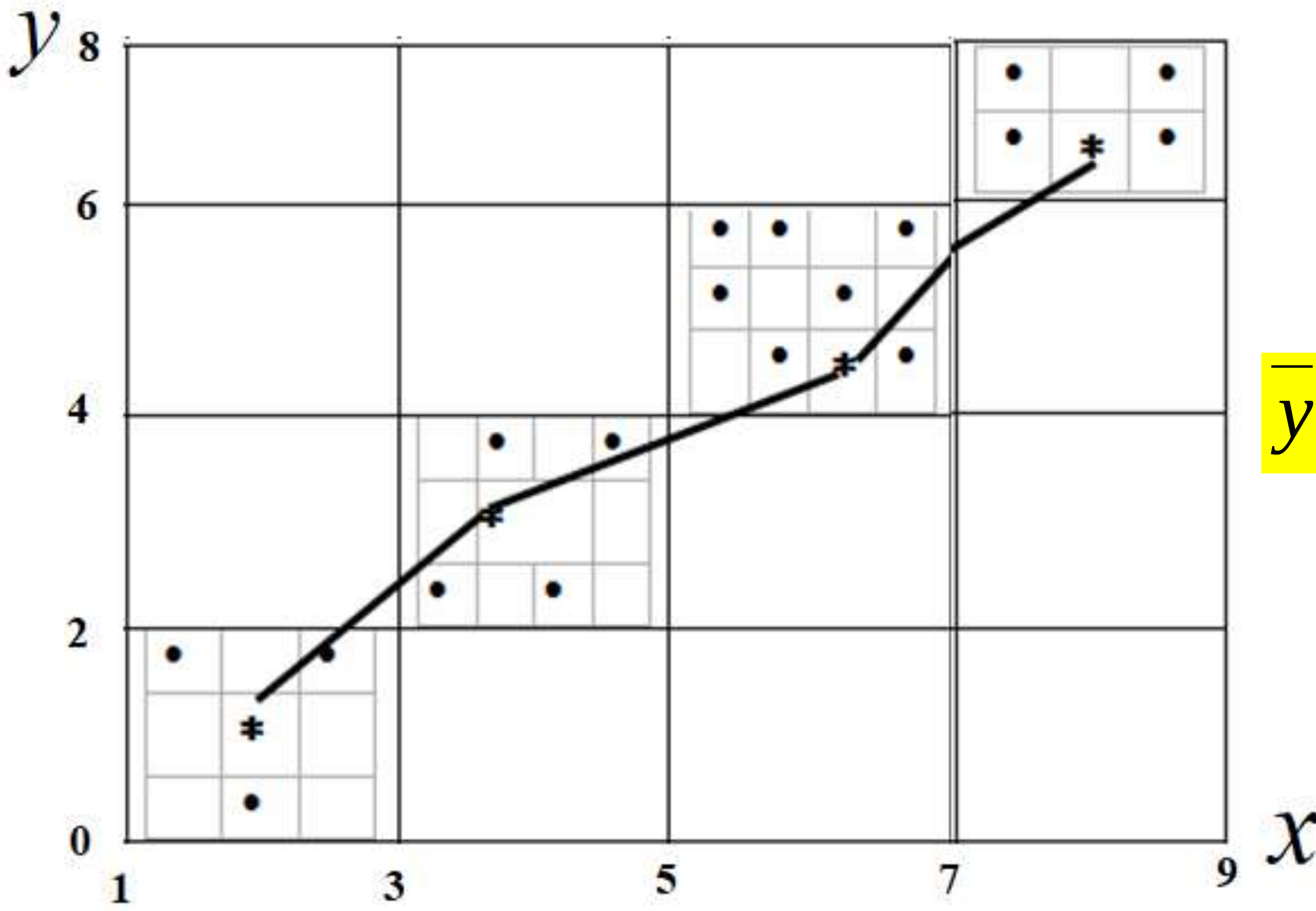


Example. Let's define h_y=3-1=2



Example.

It's the empirical regression equation *of Y upon X*



$$\overline{y}_x = f(x)$$

Notions of a regression and a correlation are related. **A correlation analysis** estimates a strength (тіснота) stochastic correlation (зв'язки), **a regression analysis** investigates its form. Both types of analysis are used for explanations of cause-and-effect relations between phenomenons and for defining a presence или a lack of a correlation.

Regression analysis **is a statistical tool for the investigation of relationships between variables.** Usually, the investigator seeks to ascertain the causal effect of one variable upon another—the effect of a price increase upon demand, for example, or the effect of changes in the money supply upon the inflation rate. To explore such issues, the investigator assembles data on the underlying variables of interest and employs regression to estimate the quantitative effect of the causal variables upon the variable that they influence. The investigator also typically assesses the "statistical significance" of the estimated relationships, that is, the degree of confidence that the true relationship is close to the estimated relationship.

There are such types ***a correlation and a regression***:

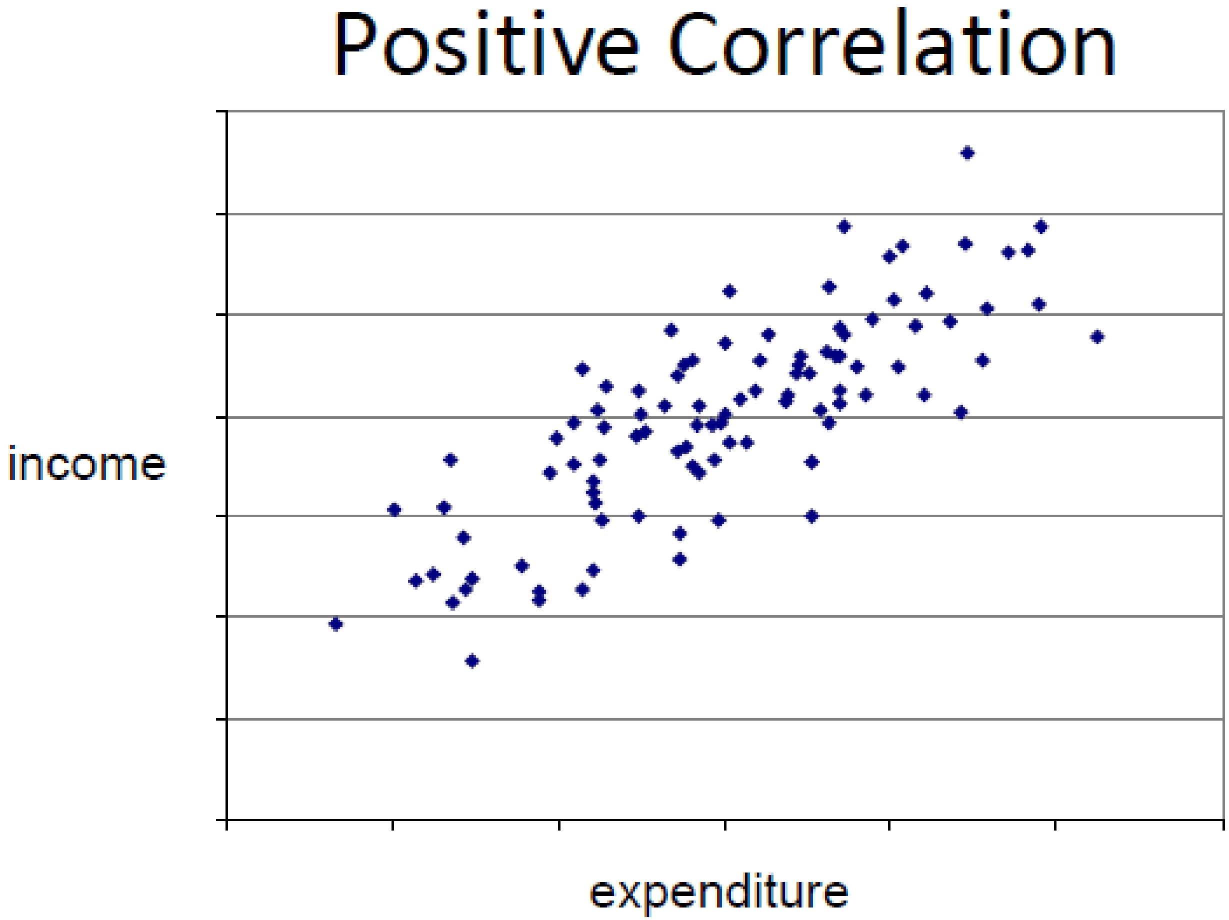
1)relative to a character of a correlation and regression we have *a positive or negative correlation and regression*. A positive value means that an increase (decrease) of an argument ***x*** is related to an increase (decrease) of a functional factor ***y***. A positive value means that an increase (decrease) of an argument ***x*** is related to a decrease (increase) of a functional factor ***y***.

Example

Correlation

- Correlation examines the relationships between pairs of variables, for example
  - between the price of doughnuts and the demand for them
  - between economic growth and life expectancy
  - between hair colour and hourly wage
  - between rankings
- Such analyses can be useful for formulating policies

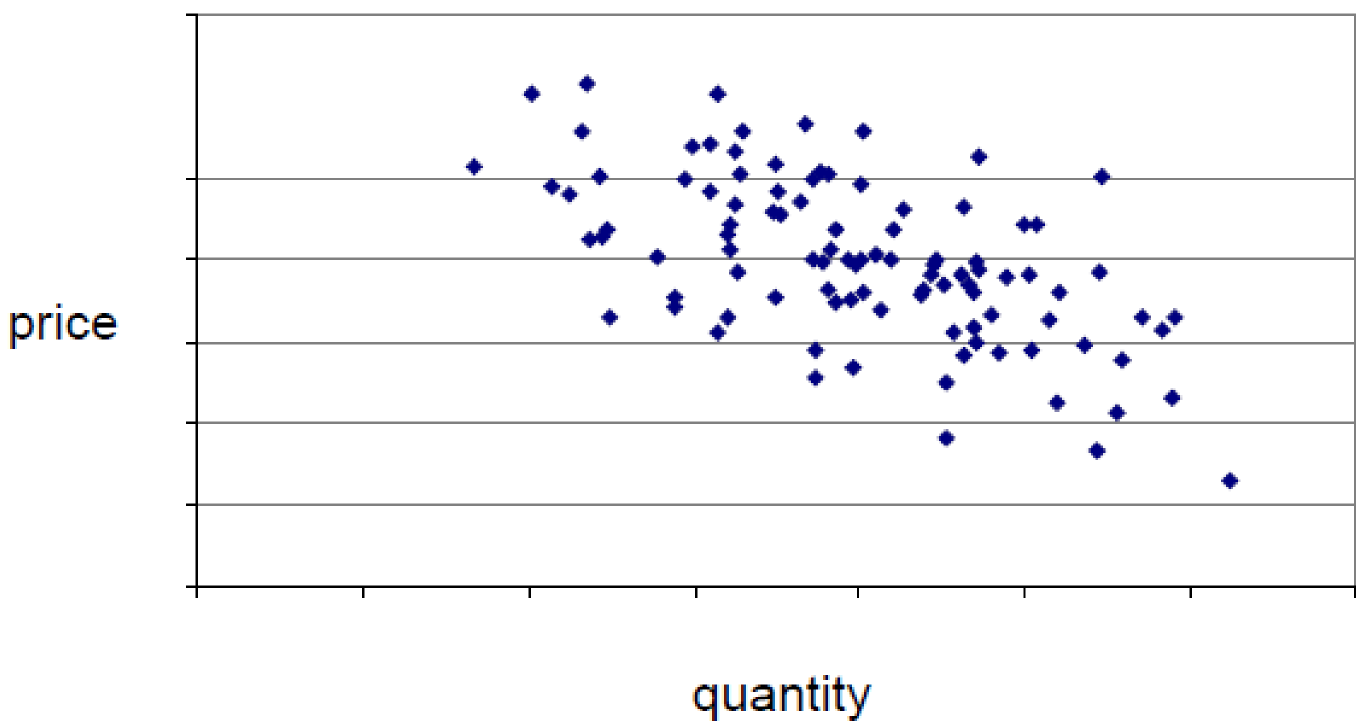
Example



E.g. income and food expenditure

Example

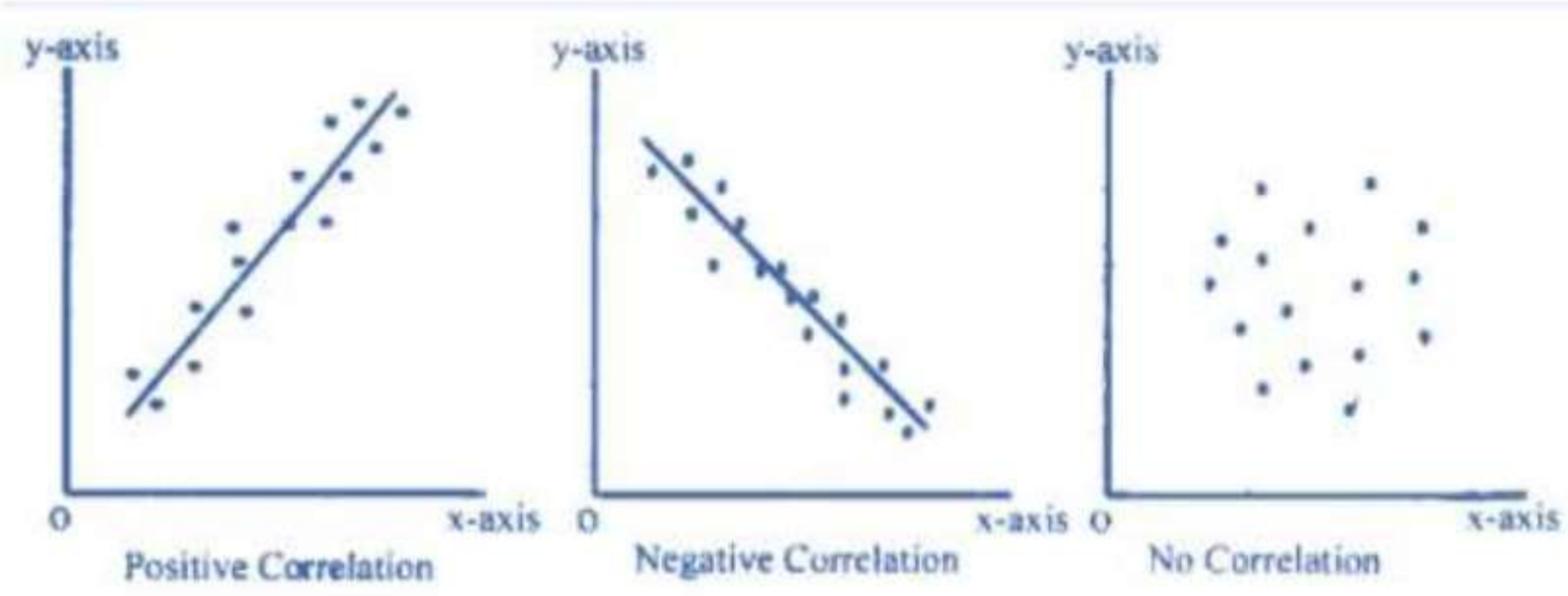
Negative Correlation



E.g. demand and price

Example

Correlations. A correlation is when two variables change together and so associate in some way. The correlation can be: ***Positive.*** The variables increase and decrease together. ***Negative.*** When one variable increases the other decreases and vice-versa. ***No correlation.*** The variables vary randomly with respect to each other.



There are such types ***a correlation and a regression***:

2) ***relative to a number of variables***: a pair or multiple correlation and regression.

Example

A pair regression is $\overline{y} = f(x)$

Example

A pair regression is $\overline{y} = f(x)$

A multiple regression is $\overline{y} = f(x_1, x_2, x_3, x_4)$

There are such types ***a correlation and a regression***:

3) ***relative to a correlation form***: a linear or nonlinear correlation and regression.

Example

A linear dependence is: $\tilde{y}_x = b_0 + b_1 x$

A nonlinear dependence is:

$$\tilde{y}_x = b_0 \cdot x^{b_1}$$ (a power function)

$$\tilde{y}_x = b_0 \cdot b_1^x$$ (an exponential function)

$$\tilde{y}_x = \frac{1}{b_0 + b_1 \cdot x}$$ (a hyperbolic function)

LEAST SQUARES METHOD (LSM)

Let's take empirical data:

x	x_1	x_2	$\cdots$	x_i	$\cdots$	x_n
y	y_1	y_2	$\cdots$	y_i	$\cdots$	y_n

Let's take empirical data:

x	x_1	x_2	$\cdots$	x_i	$\cdots$	x_n
y	y_1	y_2	$\cdots$	y_i	$\cdots$	y_n

Let's construct the theoretical regression equation in the form:

$$\tilde{y}_x = b_0 + b_1 x$$

Let's construct the theoretical regression equation in the form:

$$\tilde{y}_x = b_0 + b_1 x$$

We define parameters ***b₀*** and ***b₁*** using **least-squares method (метод найменших квадратів)** in order to the sum of squares deviations of $\hat{y}_{i\ theor.}$ from $y_{i\ emp.}$ will be least:

$$\sum_{i=1}^{n} \left(\hat{y}_{i\ theor.} - y_{i\ emp.} \right)^2 = \sum_{i=1}^{n} \varepsilon_i^2 \rightarrow \min$$

We define parameters ***a*** and ***b*** using **least-squares method (метод найменших квадратів)** in order to the sum of squares deviations of $\hat{y}_{i\ theor.}$ from $y_{i\ emp.}$ will be least:

$$\sum_{i=1}^{n}\left(\hat{y}_{i\ theor.} - y_{i\ emp.}\right)^2 = \sum_{i=1}^{n}\varepsilon_i^2 \to \min$$

Let's denote

$$\sum_{i=1}^{n}\varepsilon_i^2 = F(b_0, b_1) = \sum_{i=1}^{n}\left(b_0 + b_1 x_i - y_i\right)^2 \quad \to \min$$

Let’s denote

$$\sum_{i=1}^{n} \varepsilon_i^2 = F(b_0, b_1) = \sum_{i=1}^{n} (b_0 + b_1 x_i - y_i)^2 \quad \to \min$$

According to the necessary condition of an extremum of a function of two variables we have:

$$\frac{\partial F(b_0, b_1)}{\partial b_0} = 0 \qquad \frac{\partial F(b_0, b_1)}{\partial b_1} = 0$$

According to the necessary condition of an extremum of a function of two variables

$$\frac{\partial F(b_0, b_1)}{\partial b_0} = 0 \qquad \frac{\partial F(b_0, b_1)}{\partial b_1} = 0$$

we have:

$$\begin{cases} \dfrac{\partial F}{\partial b_0} = 2\sum_{i=1}^{n}(b_0 + b_1 x_i - y_i)\cdot x_i = 0, \\ \dfrac{\partial F}{\partial b_1} = 2\sum_{i=1}^{n}(b_0 + b_1 x_i - y_i)\cdot 1 = 0. \end{cases}$$

we have:

$$\begin{cases} \dfrac{\partial F}{\partial b_0} = 2\sum_{i=1}^{n}\left(b_0 + b_1 x_i - y_i\right)\cdot x_i = 0, \\ \dfrac{\partial F}{\partial b_1} = 2\sum_{i=1}^{n}\left(b_0 + b_1 x_i - y_i\right)\cdot 1 = 0. \end{cases}$$

Let's write down this system in the form:

$$\begin{cases} b_1\sum_{i=1}^{n} x_i^2 + b_0\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i\, y_i, \\ b_1\sum_{i=1}^{n} x_i + b_0\sum_{i=1}^{n} 1 = \sum_{i=1}^{n} y_i. \end{cases} \Rightarrow$$

Let’s write down this system in the form:

$$\begin{cases} b_1 \sum\limits_{i=1}^{n} x_i^2 + b_0 \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} x_i\, y_i, \\ b_1 \sum\limits_{i=1}^{n} x_i + b_0 \sum\limits_{i=1}^{n} 1 = \sum\limits_{i=1}^{n} y_i. \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} b_1 \sum\limits_{i=1}^{n} x_i^2 + b_0 \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} x_i\, y_i, \\ b_1 \sum\limits_{i=1}^{n} x_i + b_0 n = \sum\limits_{i=1}^{n} y_i. \end{cases}$$

Let’s write down this system in the form:

$$\Rightarrow \quad \begin{cases} b_1 \sum_{i=1}^{n} x_i^2 + b_0 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i\, y_i, \\ b_1 \sum_{i=1}^{n} x_i + b_0 n = \sum_{i=1}^{n} y_i. \end{cases}$$

Let’s divide both equations by ***n***:

$$\begin{cases} b_1 \dfrac{\sum_{i=1}^{n} x_i^2}{n} + b_0 \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{\sum_{i=1}^{n} x_i\, y_i}{n}, \\ b_1 \dfrac{\sum_{i=1}^{n} x_i}{n} + b_0 = \dfrac{\sum_{i=1}^{n} y_i}{n}. \end{cases}$$

Let's divide both equations by ***n***:

$$\begin{cases} b_1 \dfrac{\sum\limits_{i=1}^{n} x_i^2}{n} + b_0 \dfrac{\sum\limits_{i=1}^{n} x_i}{n} = \dfrac{\sum\limits_{i=1}^{n} x_i\, y_i}{n}, \\ b_1 \dfrac{\sum\limits_{i=1}^{n} x_i}{n} + b_0 = \dfrac{\sum\limits_{i=1}^{n} y_i}{n}. \end{cases}$$

Since we get **the mean** for each value

$$\frac{\sum\limits_{i=1}^{n} x_i}{n} = \overline{x} \qquad \frac{\sum\limits_{i=1}^{n} x_i^2}{n} = \overline{x^2} \qquad \frac{\sum\limits_{i=1}^{n} y_i}{n} = \overline{y} \qquad \frac{\sum\limits_{i=1}^{n} x_i\, y_i}{n} = \overline{xy}$$

Since

$$\frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \qquad \frac{\sum_{i=1}^{n} x_i^2}{n} = \overline{x^2} \qquad \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y} \qquad \frac{\sum_{i=1}^{n} x_i\, y_i}{n} = \overline{xy}$$

We get:

$$\begin{cases} b_1 \overline{x^2} + b_0 \bar{x} = \overline{xy} \\ b_1 \bar{x} + b_0 = \bar{y} \end{cases}$$

We get:

$$\begin{cases} b_1 \overline{x^2} + b_0 \overline{x} = \overline{xy} \\ b_1 \overline{x} + b_0 = \overline{y} \end{cases}$$

Let’s multiply the second equation by $\overline{x}$ and subtract it from the first equation:

$$b_1 \overline{x^2} + b_0 \overline{x} = \overline{xy}$$

$$- \quad b_1 \left(\overline{x}\right)^2 + b_0 \overline{x} = \overline{y} \cdot \overline{x}$$

We get:

$$\begin{cases} b_1 \overline{x^2} + b_0 \overline{x} = \overline{xy} \\ b_1 \overline{x} + b_0 = \overline{y} \end{cases}$$

Let's multiply the second equation by $\overline{x}$ and subtract it from the first equation:

$$b_1 \overline{x^2} + b_0 \overline{x} = \overline{xy}$$

$$- \quad b_1 \left(\overline{x}\right)^2 + b_0 \overline{x} = \overline{y} \cdot \overline{x}$$

$$b_1 \left(\overline{x^2} - (\overline{x})^2\right) = \overline{xy} - \overline{x} \cdot \overline{y}$$

Let's multiply the second equation by $\bar{x}$ and subtract it from the first equation:

$$b_1 \overline{x^2} + b_0 \bar{x} = \overline{xy}$$

$$\frac{\bar{\ } \quad b_1 (\bar{x})^2 + b_0 \bar{x} = \bar{y} \cdot \bar{x}}{b_1 \left(\overline{x^2} - (\bar{x})^2 \right) = \overline{xy} - \bar{x} \cdot \bar{y}}$$

Then

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2}$$

Then

$$b_1 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2} = \frac{\mu_{xy}}{\sigma_x^2}$$

From the second equation of the system: $b_0 = \overline{y} - b_1 \overline{x}$

The parameter **b1** is called ***a regression coefficient Y upon X*** and it is denoted as $\rho_{y/x}$

The value $\overline{xy} - \overline{x} \cdot \overline{y}$ is called ***a covariation*** or ***a correlation moment*** and it is denoted as μ_{xy}

$\overline{x^2} - \left(\overline{x}\right)^2 = \sigma_x^2$ is the variance of the variable ***X***.

The parameter **b1** is called ***a regression coefficient Y upon X*** and it is denoted as $\rho_{y/x}$

The value $\overline{xy} - \overline{x} \cdot \overline{y}$

is called ***a covariation*** or ***a correlation moment*** and it is denoted as μ_{xy}

$\overline{x^2} - \left(\overline{x}\right)^2 = \sigma_x^2$ is the variance of the variable ***X***.

The parameter **b1** is called ***a regression coefficient Y upon X*** and it is denoted as $\rho_{y/x}$

The value $\overline{xy} - \overline{x} \cdot \overline{y}$

is called ***a covariation*** or ***a correlation moment*** and it is denoted as μ_{xy}

$\overline{x^2} - \left(\overline{x}\right)^2 = \sigma_x^2$ is the variance of the variable ***X***.

$$b_1 = \rho_{y/x} = \frac{\mu_{xy}}{\sigma_x^2}$$

$$b_0 = \overline{y} - b_1 \overline{x}$$

Then $b_1 = \rho_{y/x} = \frac{\mu_{xy}}{\sigma_x^2}$ $b_0 = \overline{y} - b_1 \overline{x}$

The theoretical regression equation Y upon X has the form:

$$\tilde{y}_x = b_0 + b_1 x$$

Then $b_1 = \rho_{y/x} = \frac{\mu_{xy}}{\sigma_x^2}$ $b_0 = \overline{y} - b_1 \overline{x}$

The theoretical regression equation Y upon X has the form: $\hat{y} = \rho_{y/x} x + b_0 = b_1 x + b_0$

Like this equation we can get **the theoretical regression equation X upon Y:**

$$\hat{x} = \rho_{x/y} y + a_0 = a_1 y + a_0$$

Then $b_1 = \rho_{y/x} = \frac{\mu_{xy}}{\sigma_x^2}$ $b_0 = \overline{y} - b_1 \overline{x}$

The theoretical regression equation Y upon X has the form:

$$\hat{y} = \rho_{y/x} x + b_0 = b_1 x + b_0$$

Like this equation we can get **the theoretical regression equation X upon Y:**

$$\hat{x} = \rho_{x/y} y + a_0 = a_1 y + a_0$$

where

$$a_1 = \rho_{x/y} = \frac{\mu_{xy}}{\sigma_y^2} \qquad \sigma_y^2 = \overline{y^2} - \left(\overline{y}\right)^2 \qquad a_0 = \bar{x} - \rho_{x/y} \bar{y}$$

If data are grouped, then according to a correlation table we calculate:

$$\bar{x} = \frac{\sum_i x_i \, m_{x_i}}{n} \qquad \bar{y} = \frac{\sum_k y_k m_{y_k}}{n}$$

$$\overline{x^2} = \frac{\sum_i x_i^2 m_{x_i}}{n} \qquad \overline{xy} = \frac{\sum_i \sum_k x_i \, y_k m_{ik}}{n}$$

Example

Let's get back to the correlation table:

Y \ X	2	4	6	8	m_{y_j}	$y_j m_{y_j}$	$y_j^2 m_{y_j}$	$x_i y_j m_{ij}$
1	1	2			3			
3	1	6	3		10			
5		5	5		10			
7			1	1	2			
m_{x_i}	2	13	9	1	25			
$x_i m_{x_i}$								
$x_i^2 m_{x_i}$								
$x_i y_j m_{ij}$								

Example

Let's calculate:

Y \ X	2	4	6	8	m_{y_j}	$y_j m_{y_j}$	$y_j^2 m_{y_j}$	$x_i y_j m_{ij}$
1	1	2			3			
3	1	6	3		10			
5		5	5		10			
7			1	1	2			
m_{x_i}	2	13	9	1	25			
$x_i m_{x_i}$	4	52	54	8				
$x_i^2 m_{x_i}$								
$x_i y_j m_{ij}$								

Example

Let’s calculate:

Y \ X	2	4	6	8	m_{y_j}	$y_j m_{y_j}$	$y_j^2 m_{y_j}$	$x_i y_j m_{ij}$
1	1	2			3			
3	1	6	3		10			
5		5	5		10			
7			1	1	2			
m_{x_i}	2	13	9	1	25			
$x_i m_{x_i}$	4	52	54	8	118			
$x_i^2 m_{x_i}$								
$x_i y_j m_{ij}$								

Example

Let's calculate:

Y \ X	2	4	6	8	m_{y_j}	$y_j m_{y_j}$	$y_j^2 m_{y_j}$	$x_i y_j m_{ij}$
1	1	2			3			
3	1	6	3		10			
5		5	5		10			
7			1	1	2			
m_{x_i}	2	13	9	1	25			
$x_i m_{x_i}$	4	52	54	8	118			
$x_i^2 m_{x_i}$								
$x_i y_j m_{ij}$								

$$\bar{x} = \frac{2\cdot 2 + 4\cdot 13 + 6\cdot 9 + 8\cdot 1}{25} = \frac{118}{25} = 4{,}72$$

Example

Let’s calculate:

Y \ X	2	4	6	8	m_{y_j}	$y_j m_{y_j}$	$y_j^2 m_{y_j}$	$x_i y_j m_{ij}$
1	1	2			3	3		
3	1	6	3		10	30		
5		5	5		10	50		
7			1	1	2	14		
m_{x_i}	2	13	9	1	25			
$x_i m_{x_i}$	4	52	54	8	118			
$x_i^2 m_{x_i}$								
$x_i y_j m_{ij}$								

Example

Let's calculate:

Y \ X	2	4	6	8	m_{y_j}	$y_j m_{y_j}$	$y_j^2 m_{y_j}$	$x_i y_j m_{ij}$
1	1	2			3	3		
3	1	6	3		10	30		
5		5	5		10	50		
7			1	1	2	14		
m_{x_i}	2	13	9	1	25	97		
$x_i m_{x_i}$	4	52	54	8	118			
$x_i^2 m_{x_i}$								
$x_i y_j m_{ij}$								

Example

Let’s calculate:

X / Y	2	4	6	8	m_{y_j}	$y_j m_{y_j}$	$y_j^2 m_{y_j}$	$x_i y_j m_{ij}$
1	1	2			3	3		
3	1	6	3		10	30		
5		5	5		10	50		
7			1	1	2	14		
m_{x_i}	2	13	9	1	25	97		
$x_i m_{x_i}$	4	52	54	8	118			
$x_i^2 m_{x_i}$								
$x_i y_j m_{ij}$								

$$\bar{y} = \frac{1\cdot 3 + 3\cdot 10 + 5\cdot 10 + 7\cdot 2}{25} = \frac{97}{25} = 3{,}88$$

Example

Let's calculate:

Y \ X	2	4	6	8	m_{y_j}	$y_j m_{y_j}$	$y_j^2 m_{y_j}$	$x_i y_j m_{ij}$
1	1	2			3	3		
3	1	6	3		10	30		
5		5	5		10	50		
7			1	1	2	14		
m_{x_i}	2	13	9	1	25	97		
$x_i m_{x_i}$	4	52	54	8	118			
$x_i^2 m_{x_i}$	8	208	324	64				
$x_i y_j m_{ij}$								

Example

Let's calculate:

Y \ X	2	4	6	8	m_{y_j}	$y_j m_{y_j}$	$y_j^2 m_{y_j}$	$x_i y_j m_{ij}$
1	1	2			3	3		
3	1	6	3		10	30		
5		5	5		10	50		
7			1	1	2	14		
m_{x_i}	2	13	9	1	25	97		
$x_i m_{x_i}$	4	52	54	8	118			
$x_i^2 m_{x_i}$	8	208	324	64	604			
$x_i y_j m_{ij}$								

Example

Let's calculate:

Y \ X	2	4	6	8	m_{y_i}	$y_j m_{y_i}$	$y_j^2 m_{y_i}$	$x_i y_j m_{ij}$
1	1	2			3	3		
3	1	6	3		10	30		
5		5	5		10	50		
7			1	1	2	14		
m_{x_i}	2	13	9	1	25	97		
$x_i m_{x_i}$	4	52	54	8	118			
$x_i^2 m_{x_i}$	8	208	324	64	604			
$x_i y_j m_{ij}$								

$$\overline{x^2} = \frac{4 \cdot 2 + 16 \cdot 13 + 36 \cdot 9 + 64 \cdot 1}{25} = \frac{604}{25} = 24{,}16$$

Example

Let's calculate:

Y \ X	2	4	6	8	m_{y_j}	$y_j m_{y_j}$	$y_j^2 m_{y_j}$	$x_i y_j m_{ij}$
1	1	2			3	3	3	
3	1	6	3		10	30	90	
5		5	5		10	50	250	
7			1	1	2	14	98	
m_{x_i}	2	13	9	1	25	97		
$x_i m_{x_i}$	4	52	54	8	118			
$x_i^2 m_{x_i}$	8	208	324	64	604			
$x_i y_j m_{ij}$								

Example

Let's calculate:

Y \ X	2	4	6	8	m_{y_j}	$y_j m_{y_j}$	$y_j^2 m_{y_j}$	$x_i y_j m_{ij}$
1	1	2			3	3	3	
3	1	6	3		10	30	90	
5		5	5		10	50	250	
7			1	1	2	14	98	
m_{x_i}	2	13	9	1	25	97	441	
$x_i m_{x_i}$	4	52	54	8	118			
$x_i^2 m_{x_i}$	8	208	324	64	604			
$x_i y_j m_{ij}$								

Example

Let's calculate:

Y \ X	2	4	6	8	m_{y_j}	$y_j m_{y_j}$	$y_j^2 m_{y_j}$	$x_i y_j m_{ij}$
1	1	2			3	3	3	
3	1	6	3		10	30	90	
5		5	5		10	50	250	
7			1	1	2	14	98	
m_{x_i}	2	13	9	1	25	97	441	
$x_i m_{x_i}$	4	52	54	8	118			
$x_i^2 m_{x_i}$	8	208	324	64	604			
$x_i y_j m_{ij}$								

$$\overline{y^2} = \frac{1\cdot 3 + 9\cdot 10 + 25\cdot 10 + 49\cdot 2}{25} = \frac{441}{25} = 17{,}64$$

Example

Let’s calculate:

Y \ X	2	4	6	8	m_{y_j}	$y_j m_{y_j}$	$y_j^2 m_{y_j}$	$x_i y_j m_{ij}$
1	1 2	2 8			3	3	3	
3	1 6	6 72	3 54		10	30	90	
5		5 100	5 150		10	50	250	
7			1 42	1 56	2	14	98	
m_{x_i}	2	13	9	1	25	97	441	
$x_i m_{x_i}$	4	52	54	8	118			
$x_i^2 m_{x_i}$	8	208	324	64	604			
$x_i y_j m_{ij}$								

Example

Let’s calculate:

Y \ X	2	4	6	8	m_{y_j}	$y_j m_{y_j}$	$y_j^2 m_{y_j}$	$x_i y_j m_{ij}$
1	1 2	2 8			3	3	3	
3	1 6	6 72	3 54		10	30	90	
5		5 100	5 150		10	50	250	
7			1 42	1 56	2	14	98	
m_{x_i}	2	13	9	1	25	97	441	
$x_i m_{x_i}$	4	52	54	8	118			
$x_i^2 m_{x_i}$	8	208	324	64	604			
$x_i y_j m_{ij}$	8	180	246	56	490			

Example

Let’s calculate:

Y \ X	2	4	6	8	m_{y_j}	$y_j m_{y_j}$	$y_j^2 m_{y_j}$	$x_i y_j m_{ij}$
1	1 2	2 8			3	3	3	10
3	1 6	6 72	3 54		10	30	90	132
5		5 100	5 150		10	50	250	250
7			1 42	1 56	2	14	98	98
m_{x_i}	2	13	9	1	25	97	441	490
$x_i m_{x_i}$	4	52	54	8	118			
$x_i^2 m_{x_i}$	8	208	324	64	604			
$x_i y_j m_{ij}$	8	180	246	56	490			

Example

Y \ X	2	4	6	8	m_{y_j}	$y_j m_{y_j}$	$y_j^2 m_{y_j}$	$x_i y_j m_{ij}$
1	1 2	2 8			3	3	3	10
3	1 6	6 72	3 54		10	30	90	132
5		5 100	5 150		10	50	250	250
7			1 42	1 56	2	14	98	98
m_{x_i}	2	13	9	1	25	97	441	490
$x_i m_{x_i}$	4	52	54	8	118			
$x_i^2 m_{x_i}$	8	208	324	64	604			
$x_i y_j m_{ij}$	8	180	246	56	490			

$$\overline{xy} = \frac{\sum_i \sum_k m_{ki} x_i y_k}{n} = \frac{1}{25}\left(1\cdot 2\cdot 1 + 3\cdot 2\cdot 1 + 1\cdot 4\cdot 2 + 3\cdot 4\cdot 6 + 5\cdot 4\cdot 5 + \right.$$

$$\left. + 3\cdot 6\cdot 3 + 5\cdot 6\cdot 5 + 7\cdot 6\cdot 1 + 7\cdot 8\cdot 1\right) = \frac{1}{25}\left(2+6+8+72+100+54+150+42+56\right) = \frac{490}{25} = 19{,}6$$

Example

Let's calculate:

$$b_1 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2}$$

$\overline{x} = 4{,}72$ $\overline{y} = 3{,}88$ $\overline{x^2} = 24{,}16$ $\overline{xy} = 19{,}6$

Example

Let's calculate:

$$b_1 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2}$$

$\overline{x} = 4,72$ $\overline{y} = 3,88$ $\overline{x^2} = 24,16$ $\overline{xy} = 19,6$

$$\rho_{y/x} = b_1 = \frac{19,6 - 4,72 \cdot 3,88}{24,16 - 4,72^2} = \frac{1,2864}{1,8816} \approx 0,68$$

Example

Let's calculate:

$$\rho_{y/x} = b_1 = \frac{19{,}6 - 4{,}72 \cdot 3{,}88}{24{,}16 - 4{,}72^2} = \frac{1{,}2864}{1{,}8816} \approx 0{,}68$$

$\overline{x} = 4{,}72$ $\overline{y} = 3{,}88$ $\overline{x^2} = 24{,}16$ $\overline{xy} = 19{,}6$

$$b_0 = \overline{y} - b_1 \overline{x}$$

Example

Let's calculate:

$$\rho_{y/x} = b_1 = \frac{19{,}6 - 4{,}72 \cdot 3{,}88}{24{,}16 - 4{,}72^2} = \frac{1{,}2864}{1{,}8816} \approx 0{,}68$$

$\overline{x} = 4{,}72$ $\overline{y} = 3{,}88$ $\overline{x^2} = 24{,}16$ $\overline{xy} = 19{,}6$

$$b_0 = \overline{y} - b_1 \overline{x}$$

$$b_0 = 3{,}88 - 0{,}68 \cdot 4{,}72 \approx 0{,}67$$

Example

Let’s calculate:

$$\rho_{y/x} = b_1 = \frac{19,6 - 4,72 \cdot 3,88}{24,16 - 4,72^2} = \frac{1,2864}{1,8816} \approx 0,68$$

$$b_0 = 3,88 - 0,68 \cdot 4,72 \approx 0,67$$

$$\tilde{y}_x = b_0 + b_1 x$$

$$\tilde{y}_x = 0,68x + 0,67$$

Example

The theoretical regression equation is

$$\tilde{y}_x = 0{,}68x + 0{,}67$$

Explanation: the coefficient $b_1 = 0{,}68$ shows the increasing X by 1 unit gives the increasing Y by 0,68 units.

Example 2

The theoretical regression equation is

$$\hat{y}_x = -4{,}8x + 7$$

Explanation: the coefficient **b_1=-4,8** shows the increasing X by 1 unit gives the decreasing Y by 4.8 units.

A correlation coefficient, its properties and explanation

Since a correlation moment (a covariation) μ_{xy} is a dimensional characteristic, then one introduces an additional nondimensional characteristic (***a correlation coefficient***) as a coefficient of a **strength** (тіснота) of a linear correlation:

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y}$$

Signs of coefficients $r_{xy},\ \mu_{xy},\ \rho_{y/x},\ \rho_{x/y}$ coincide.

Properties of a correlation coefficient:

1) If random variable X and Y are linearly independent, then $\mu_{xy} = 0$ (since $r_{xy} = 0$).

Properties of a correlation coefficient:

1) If random variable X and Y are linearly independent, then $\mu_{xy} = 0$ (since $r_{xy} = 0$).

2) random variable X and Y are linearly dependent, then $r_{xy} = 1$.

Properties of a correlation coefficient:

1) If random variable X and Y are linearly independent, then $r_{xy} = 0$ (since $\mu_{xy} = 0$).

2) random variable X and Y are linearly dependent, then $r_{xy} = 1$.

3) $|r_{xy}| \leq 1$, $-1 \leq r \leq 1$

Properties of a correlation coefficient:

1) If random variable X and Y are linearly independent, then $\mu_{xy} = 0$ (since $r_{xy} = 0$).

2) random variable X and Y are linearly dependent, then $r_{xy} = 1$.

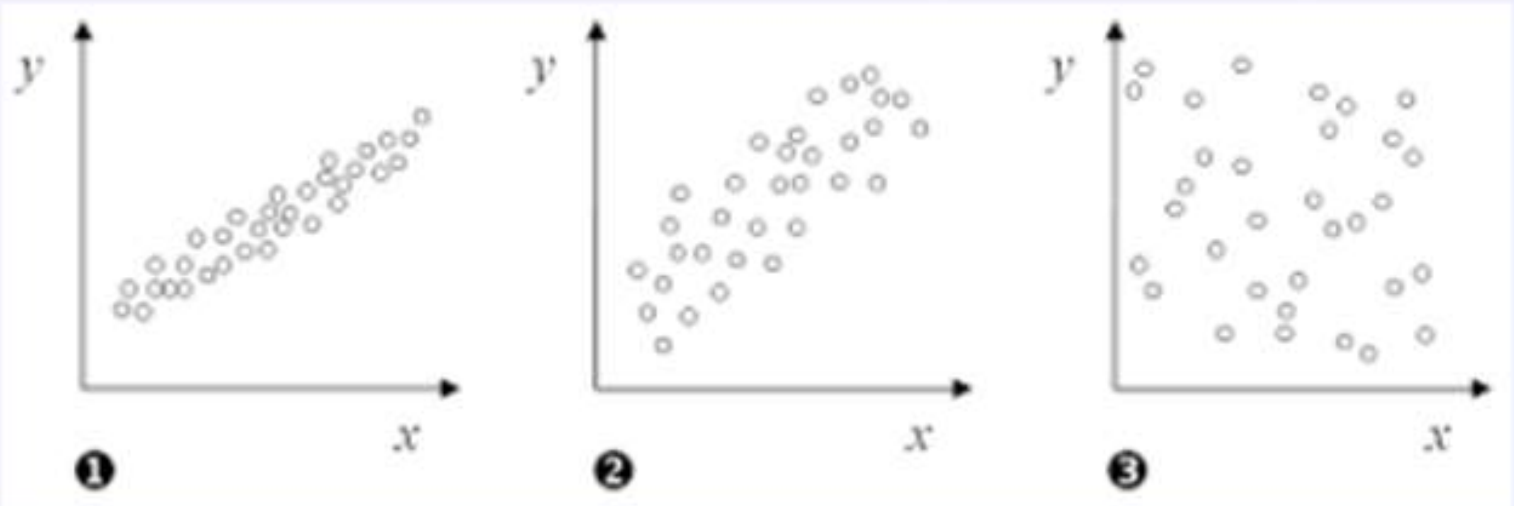
3) $\left| r_{xy} \right| \leq 1, \quad -1 \leq r \leq 1$

4) $r_{xy} = r_{yx}$ according to the definition

Values of *r*

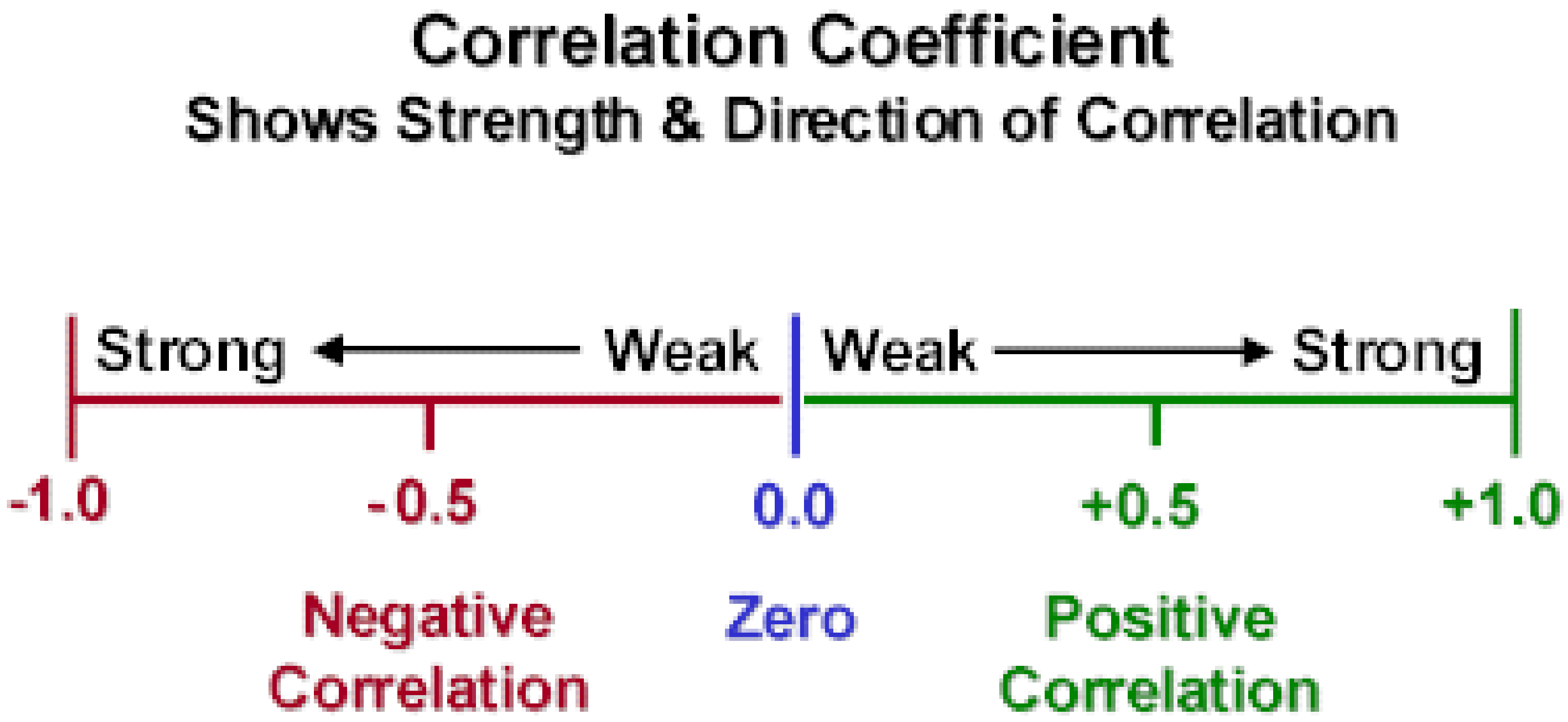
We will use this explanation:

<u>*Conclusion:*</u> If $|r| < 0{,}35$ then this correlation is **weak**, if $0{,}35 \le |r| \le 0{,}7$ then this correlation is **moderate**, if $|r| > 0{,}7$ then this correlation is **strong**.

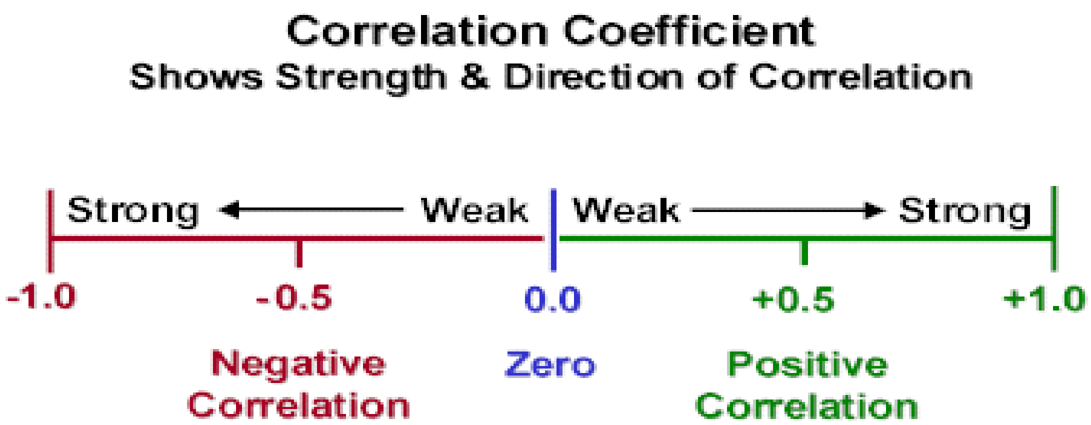


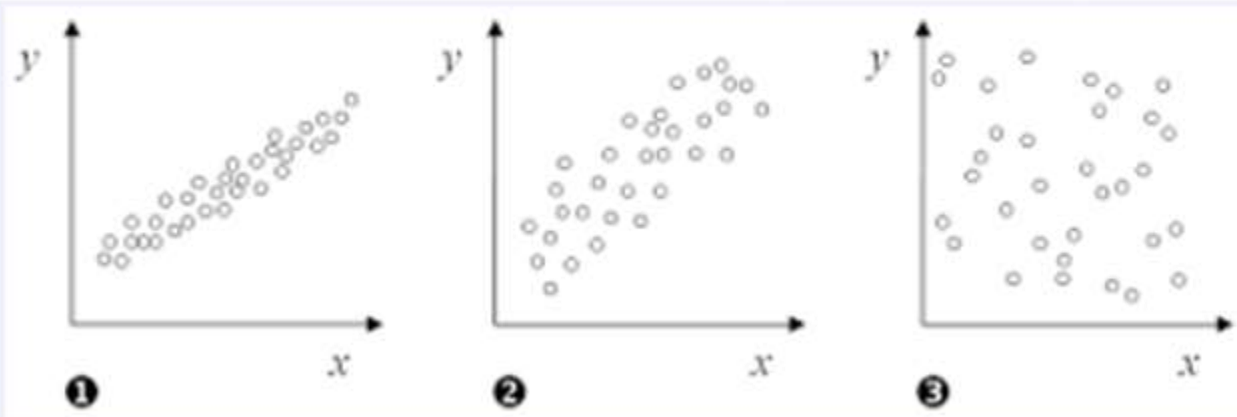
Strong, weak and no correlation

Values of *r*


Correlation Coefficient
Shows Strength & Direction of Correlation
Strong
Weak
Weak
Strong
-1.0
-0.5
0.0
+0.5
+1.0
Negative Correlation
Zero
Positive Correlation

Values of *r*



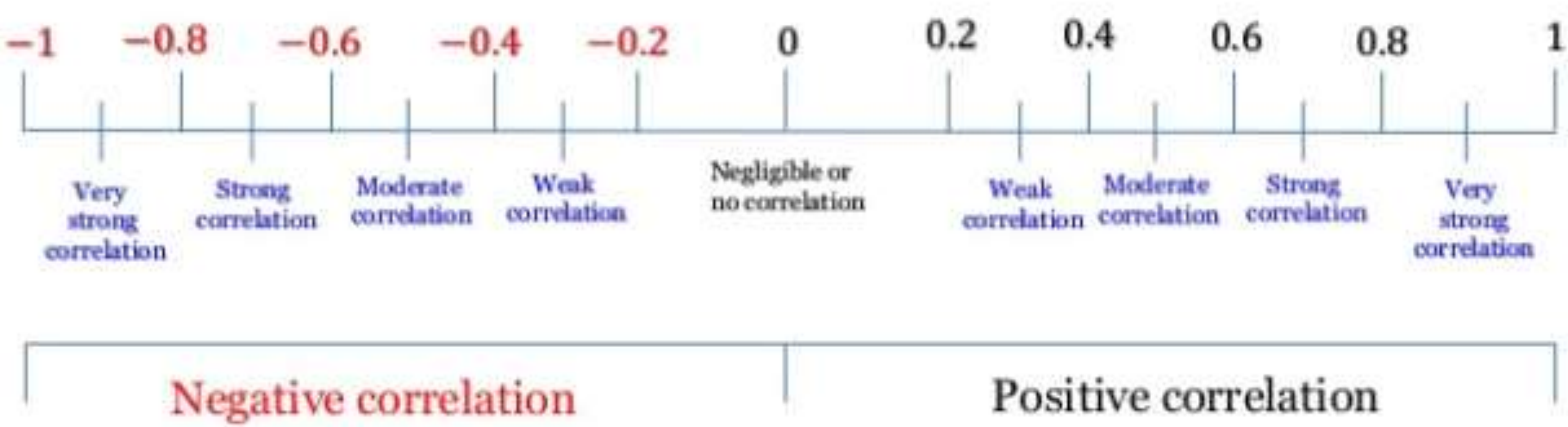


Strong, weak and no correlation

Values of *r*

Correlation Coefficient Interpretation Guideline

The correlation coefficient (r) ranges from -1 (a perfect negative correlation) to 1 (a perfect positive correlation). In short, $-1 \leq r \leq 1$.



- The value of r is such that $-1 \leq r \leq +1$. The + and – signs are used for positive linear correlations and negative linear correlations, respectively.
- *Positive correlation:* If x and y have a strong positive linear correlation, r is close to +1. An r value of exactly +1 indicates a perfect positive fit. Positive values indicate a relationship between x and y variables such that as values for x increases, values for y also increase.
- *Negative correlation:* If x and y have a strong negative linear correlation, r is close to -1. An r value of exactly -1 indicates a perfect negative fit. Negative values indicate a relationship between x and y such that as values for x increase, values for y decrease.
- *No correlation:* If there is no linear correlation or a weak linear correlation, r is close to 0. A value near zero means that there is a random, nonlinear relationship between the two variables
- Note that r is a dimensionless quantity; that is, it does not depend on the units employed.
- A *perfect* correlation of ± 1 occurs only when the data points all lie exactly on a straight line. If $r = +1$, the slope of this line is positive. If $r = -1$, the slope of this line is negative.
- A correlation greater than 0.8 is generally described as *strong*, whereas a correlation less than 0.5 is generally described as *weak*. These values can vary based upon the "type" of data being examined. A study utilizing scientific data may require a stronger correlation than a study using social science data.

Example

The correlation coefficient:

$$\mu_{xy} = \overline{xy} - \overline{x} \cdot \overline{y} = 19,6 - 4,72 \cdot 3,88 = 1,2864$$

$$\sigma_x^2 = \overline{x^2} - \left(\overline{x}\right)^2 = 24,16 - 4,72^2 = 1,8816$$

$$\sigma_y^2 = \overline{y^2} - \left(\overline{y}\right)^2 = 17,64 - 3,88^2 = 2,5856$$

Example

The correlation coefficient:

$$\mu_{xy} = \overline{xy} - \overline{x} \cdot \overline{y} = 19{,}6 - 4{,}72 \cdot 3{,}88 = 1{,}2864$$

$$\sigma_x^2 = \overline{x^2} - \left(\overline{x}\right)^2 = 24{,}16 - 4{,}72^2 = 1{,}8816$$

$$\sigma_y^2 = \overline{y^2} - \left(\overline{y}\right)^2 = 17{,}64 - 3{,}88^2 = 2{,}5856$$

$$r = \frac{\mu_{xy}}{\sigma_x \cdot \sigma_y} = \frac{1{,}2864}{\sqrt{1{,}8816 \cdot 2{,}5856}} \approx 0{,}5832$$

Example

The correlation coefficient:

$$r = \frac{\mu_{xy}}{\sigma_x \cdot \sigma_y} = \frac{1,2864}{\sqrt{1,8816 \cdot 2,5856}} \approx 0,5832$$

Then this linear correlation is moderate (середня).

Example 3

For factors x_i and y_k at the sample size n=100 we have got:

$$\sum x_i m_{xi} = 2296 \qquad \sum y_k m_{yk} = 269$$

$$\sum x_i y_k m_{ik} = 6272 \qquad \sigma_x^2 = 7{,}45 \qquad \sigma_y^2 = 10{,}21$$

where m_{xi}, m_{yk}, m_{ik} are corresponding frequencies. Estimate a strenght of a correlation.

Example 3

According to the condition of the task we get:

$\overline{x} = 22{,}96$ $\overline{y} = 2{,}69$ $\overline{xy} = 62{,}72$

Example 3

According to the condition of the task we get:

$$\overline{x} = 22{,}96 \qquad \overline{y} = 2{,}69 \qquad \overline{xy} = 62{,}72$$

$$\mu_{xy} = \overline{xy} - \overline{x} \cdot \overline{y} = 62{,}72 - 22{,}96 \cdot 2{,}69 = 0{,}9576$$

Example 3

According to the condition of the task we get:

$$\overline{x} = 22{,}96 \qquad \overline{y} = 2{,}69 \qquad \overline{xy} = 62{,}72$$

$$\mu_{xy} = \overline{xy} - \overline{x} \cdot \overline{y} = 62{,}72 - 22{,}96 \cdot 2{,}69 = 0{,}9576$$

The correlation coefficient:

$$r = \frac{\mu_{xy}}{\sigma_x \cdot \sigma_y} = \frac{0{,}9576}{\sqrt{7{,}45 \cdot 10{,}21}} = 0{,}110$$

Example 3

The correlation coefficient:

$$r = \frac{\mu_{xy}}{\sigma_x \cdot \sigma_y} = \frac{0{,}9576}{\sqrt{7{,}45 \cdot 10{,}21}} = 0{,}110$$

Since this coefficient ***r*** approaches 0, then this linear correlation is very weak (дуже слабка).

Elasticity

In economics, **elasticity** is the measurement of how responsive an economic variable is to a change in another.

Elasticity

In economics, **elasticity** is the measurement of how responsive an economic variable is to a change in another.

An ***elastic*** variable (with elasticity value greater than 1) is one which responds more than proportionally to changes in other variables. In contrast, an ***inelastic*** variable (with elasticity value less than 1) is one which changes less than proportionally in response to changes in other variables.

Elasticity

In economics, **elasticity** is the measurement of how responsive an economic variable is to a change in another.

$$\overline{E} = (b_0 + b_1 x)' \cdot \frac{\overline{x}}{\overline{y}} = b_1 \cdot \frac{\overline{x}}{\overline{y}}$$

Example. Elasticity

In economics, **elasticity** is the measurement of how responsive an economic variable is to a change in another.

$$\overline{E} = b_1 \cdot \frac{\overline{x}}{\overline{y}} = 0{,}68 \cdot \frac{4{,}72}{3{,}88} = 0{,}83\%$$

Example. Elasticity

Conclusion: The elasticity coefficient (0,83%) is a number that indicates the percentage change that will occur in one variable (*y*) when the variable *x* changes one percent.

$$\overline{E} = b_1 \cdot \frac{\overline{x}}{\overline{y}} = 0,68 \cdot \frac{4,72}{3,88} = 0,83\%$$

A determination coefficient

The **coefficient of determination**, denoted ***R^2*** or ***r^2*** and pronounced **R squared**, is a number that indicates the proportion of the variance in the dependent variable that is predictable from the independent variable

Coefficient of Determination, r^2 or R^2:

- The *coefficient of determination, r* ^{2}, is useful because it gives the proportion of the variance (fluctuation) of one variable that is predictable from the other variable. It is a measure that allows us to determine how certain one can be in making predictions from a certain model/graph.
- The *coefficient of determination* is the ratio of the explained variation to the total variation.
- The *coefficient of determination* is such that $0 \le r^2 \le 1$, and denotes the strength of the linear association between *x* and *y*.
- The *coefficient of determination* represents the percent of the data that is the closest to the line of best fit. For example, if $r = 0.922$, then $r^2 = 0.850$, which means that 85% 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 15% of the total variation in *y* remains unexplained.
- The *coefficient of determination* is a measure of how well the regression line represents the data. If the regression line passes exactly through every point on the scatter plot, it would be able to explain all of the variation. The further the line is away from the points, the less it is able to explain.

EXAMPLE

Let’s continue to solve **EXAMPLE.**

EXAMPLE

Let’s continue to solve **EXAMPLE.**

$$r = 0,5832$$

EXAMPLE

Let's continue to solve **EXAMPLE.**

$$r = 0{,}5832$$

$$R^2 = r_{xy}^2 = 0{,}5832^2 = 0{,}3401$$

EXAMPLE

Let's continue to solve **EXAMPLE.**

$$r = 0{,}5832$$

$$R^2 = r_{xy}^2 = 0{,}5832^2 = 0{,}3401$$

It means that 34,01% 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 100%-34,01%=65,99% of the total variation in ***y*** remains unexplained.

EXAMPLE

Let's continue to solve **EXAMPLE.**

EXAMPLE

Let's continue to solve **EXAMPLE 3.**

$$r_{xy} = 0,257$$

$$R^2 = r_{xy}^2 = 0,257^2 = 0,066$$

It means that 6,6% 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 93,4% of the total variation in ***y*** remains unexplained.

Thank You for your time and attention! Hope this presentation was educational and helpful to you ☺