Construction of dependence function of annual passenger flow and GPD per capita

K.O. Dohonova, Y.V. Bortnik
National Aviation University, Kyiv
Scientific advisor Y.Liashenko,
PhD in Physics and Mathematics, Associate Professor

Key words: linear regression, correlation coefficient, passenger flow, GPD per capita.

In many problems of physics, economics, medicine, engineering, etc. we have to experimentally study the dependence of a random variable *Y*, observed, on one or more other random or non-random variables $X_1, X_2, \dots, X_n$. Regression analysis is a branch of mathematical statistics that studies the dependence between random variables using regression equations, and regression is the functional dependence on average of any random variables. The main task of regression analysis is to study the dependence between the resultant variable *Y* and the variable *X*, and to observe and estimate the regression function. Correlation coefficient *r* in statistics is an indicator of the correlation between two variables X and Y. At $r = \pm 1$ variables X and Y are related by linear dependence. It is widely used in science to measure the degree of linear dependence between two variables. [1]

Example. In table are given x (passenger flow in Ukrainian airports, billions of people) [2] and y (GPD per capita in Ukraine) [3]. For the clarity of observation, we took data of previous years, where both of these factors were not affected by COVID-19 pandemic).

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
x	10686,9	8894,9	10242,5	12464,8	14107	15134,6	10896,5	10695,2	12929,9	16499,5	20545,5	24336,6
y	20502,8	19836,3	23603,6	28813,9	30912,5	31988,7	35834	46210,2	55853,5	70224,3	84192	94589,8

Firstly, is necessary to calculate sample mean of values *x* and *y:*

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{12} \cdot 167433{,}9 = 13952{,}83$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{12} \cdot 542561{,}6 = 45213{,}47$$

Then, we calculate the sample corrected standard deviation S:

$$\overline{S_x}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x^2} = 2597{,}22; \overline{S_x} = \sqrt{\overline{S_x}^2} = 50{,}96$$

$$\overline{S_y}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2 = 3{,}8;\ \overline{S_y} = \sqrt{\overline{S_y}^2} = 1{,}95$$

Let's find correlation coefficient [1]: $r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n S_x S_y} = 0{,}89$

Since *r* has the value which is close to 1 then X and Y are correlated and have a practically linear dependence. Let $y = ax + b$ be a selected linear regression equation, where the coefficients *a, b* can be found by the method of least squares:

$$a = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x^2}} = 4{,}92$$
$$b = \bar{y} - a\bar{x} = -23447{,}1$$

Thus, the equation of linear regression of the dependence of GDP per capita on the annual volume of passenger traffic has been obtained: $y = 4{,}92x - 23447{,}1$.

According to the equation of the line we construct a regression graph:

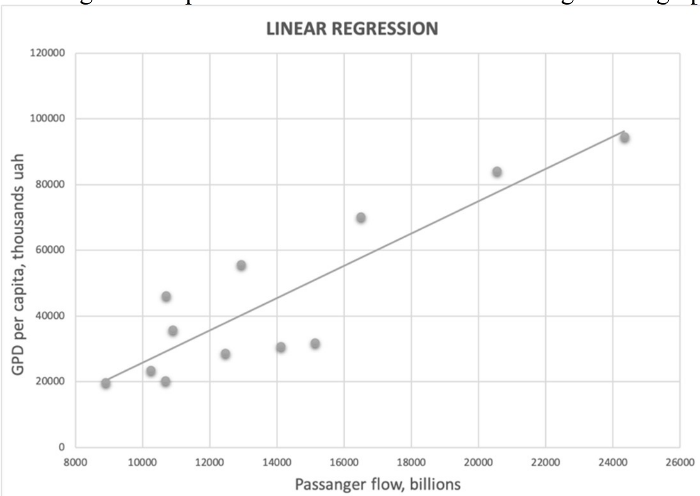

Figure 1. Graph of regression line analysis.

Summarizing all the above, we can conclude that the dependence between annual passenger traffic and GDP per capita does exists, and knowing that this dependence is defined by a linear function we can predict how changes in the passenger traffic will affect GDP per capita.

References:

1. Higher Mathematics: manual. Theory of Probability and Elements of Mathematical Statistics / V.P. Denisiuk, V.M. Bobkov, L.A. Grishina – K.: NAU, 2013. 210-216 pages.
2. Пасажиропотік аеропортів України [Electronic source] - https://www.wikiwand.com/uk/Пасажиропотік_аеропортів_України/Загальний_пасажиропотік_аеропортів_України.
3. Валовой внутренний продукт (ВВП) в Украине 2021[Electronic source] -https://index.minfin.com.ua/economy/gdp/