

# Research paper on statistics

[Business](#), [Company](#)



## **The first part of typical statistical research is descriptive statistics of the data sample.**

Descriptive statistics allows to generalize the initial results obtained by observation or experiment. All calculations descriptive statistics are reduced to group data according to their values , their frequency distribution of construction , identify trends central distribution and finally to the estimation variance of the data with respect to the central tendency found .

Presentation of descriptive statistics is usually the first step in any analysis .

The purpose of presenting data in the form of descriptive statistics - draw conclusions and make strategic ( for analysis) decisions based on available data.

### **Key indicators of descriptive statistics :**

- The average (arithmetic mean , median, mode )
- The average value
- Scattering (range scatter )
- Dispersion
- Standard ( rms) deviation
- Quartiles
- Confidence interval

For our case we have collected the following descriptive:

The next step to describe data is scatterplot. The scatterplot each observation (or the basic unit of the data set) corresponds to a point whose coordinates (Cartesian) are the values of some parameters of the two observations. If it is assumed that one of the parameters depends on the

other, it is usually independent of the parameter values is plotted on the horizontal axis and the dependent - on the vertical. Scatterplots are used to demonstrate the presence or absence of correlation between the two variables.

### **We have constructed scatter plot for Shoe\_size vs Height**

Starting from the resulting graph, it can be assumed that there is a linear relationship between the Height and Shoe\_size (as points are localized approximately on a straight line).

We claim the hypothesis about positive linear association and check it with coefficient of correlation. The null hypothesis is that there is no linear association.

According to the Pearson's  $r$  test we may reject the null hypothesis even at 1% level of significance and say, that obtained  $r = 0.864$  is a good evidence of strong and positive linear relationship between the variables.

The next step is linear regression analysis. In regression analysis, we study the connection and is determined by a quantitative relationship between a dependent variable and one or more independent variables. In our case, Shoe\_size is a dependent variable and Height is an independent variable.

### **The obtained regression equation is:**

Shoe size =  $-29.057 + 0.554 \text{Height}$

R-square is 0.747 and the adjusted R-square is 0.739. Hence, 74.7% of shoe size variance is explained by this model. This is quite good value of parameter. As p-values of coefficients of models are lesser than 0.001,

hence, the model is significant. The management of the company should use our regression equation for making forecasts and predictions.

**The second question we have to talk about is how gender may affect the size of shoes.**

The null hypothesis is: There is no significant difference between the average shoe size for males and females.

The alternative hypothesis is: There is a significant difference between the average shoe size for males and females.

Divide the sample on two new data samples: for females and for males. We use two-sampled Student's t-test to compare means.

Since p-value of t-test is lesser than 0.001, we reject the null hypothesis and state that there is a significant difference between the average shoe sizes for males and females (at 1% level of significance). The management of the Nyke company should take this information in consideration and produce separate shoes for both genders