# Traffic prediction based on cost per click

Hello, everyone! In this article, we are going to see how the regression analysis can help you to develop a predictive model. The material was written for those who already know the basics of statistics. It will be good for novices as well.

## Where to start

### Our case

We are youthful startup company which wants (as much as all other companies do) to attract users to our website so that they could worthily evaluate our product. One of the directions for traffic attracting is contextual advertising.

We have started our test advertising campaign in the USA and have changed the cost per click a lot of times in order to test and find a reasonable price.

## Data

**Error! Cannot read or display file.**

Well, after getting all necessary data, we want to develop a predictive model for traffic prediction.

## Tools that we shall use

In this article, we will use minitab express (<http://www.minitab.com/en-us/products/express/>) as the main tool for analysis (the software is paid-for, R-type analogue may be used)

We also have used a spreadsheet editor.

## Getting started

### 1. Hypothesizing

Now, we are only assuming that the cost per click affects the incoming traffic. For statistic analysis, we will have to transform our presumptions into hypothesizes:

H0 (null hypothesizes): the cost per click does NOT affect the traffic

H1 (alternative hypothesizes): the cost per click DOES affect the traffic

### 2. Setting significance point

The error of first kind is the following: the right hypothesis will be rejected. The probability of committing the error of first kind is designated as “alfa” letter, and it is called the point of significance. Typically, the point of significance is taken equal to 0.05 or 0.01. If the significance point is taken equal to 0.05, it means that we are taking the risk to commit the error of first kind (to reject the right hypothesis) in 5 cases of 100.

In even more simple terms, if the analysis shows p-value < 5%, we will reject the null hypothesis.

### 3. Performing analysis

#### A.

We enter the data to Minitab and choose regression

**Error! Cannot read or display file.**

After that, we mark the set of data on visits as Y. The cost per click will be X. We want to under stand the correspondence Y = f(x)

**Error! Cannot read or display file.**

After pressing OK, we get the following curve:

**Error! Cannot read or display file.**

Great, we have constructed Fitted Plot and acquired some important parameters:

**Error! Cannot read or display file.**

R-factor is the significance factor. More than 96%! Very significant!

p-value < 0.05

Since p-value < 0.05, we REJECT H0, and accept the alternative. The cost per click affects the amount of traffic, and at that, the effect is very significant!

#### B.

Now, we need to ensure the model accuracy. Let’s check assumptions on residuals (regressive residuals). In order to do that, we need to click on Graphs tab and tick the first item.

**Error! Cannot read or display file.**

The following conditions are important for us:

* The average and the sum for residuals of the data set = 0
* The residuals are normally distributed around 0
* The residuals have similar dispersion
* The residuals are independent from each other

Normally distributed around 0:

**Error! Cannot read or display file.**

Having similar dispersion:

**Error! Cannot read or display file.**

The average and the sum are 0:

**Error! Cannot read or display file.**

Independent and distributed randomly:

**Error! Cannot read or display file.**

Thereby, we have confirmed the accuracy of model, and now we can use the equation that has been constructed by means of regression:

**Error! Cannot read or display file.**

This formula will allow us to count the page traffic depending on the cost per click.

For example, if we want to set the limit for the cost per click (let’s say, $1.5), then we can always evaluate the amount of page traffic from such advertisement campaign.

## The total

Regression analysis may be used in many cases, for example, to understand the dependence of the expenses from selling spots, the dependence of the number of sales from the time of conversation of a sales manager, etc.

With this technique, you can easily develop a predictive model.

If you get a significant R-factor in your case, but the regression residuals do not meet the conditions – don’t give up so soon. Try to choose a quadratic approximation.

In any case, remember that data analysis always needs an individual approach.

If you want to make progress in this area of knowledge, check out our interactive courses at [Datamonkey.pro](http://datamonkey.pro). We hope, you’ll enjoy them and make use of them.

Thank you!