

Basic Statistics for Astrologers

Written by Sergey Tarassov

The statistics is the most tricky science I have ever dealt with. A decent person from the street who is not a statistician might be surprised when a lot of events of our life that appear as some laws are in reality just occasions. The simplest example is flipping the coin: we may see the heads 8 times of 10. We can assume that this is a non-occasional fact, while this is just the game of His Majesty Chaos. Try to dice the coin yourself to figure out the power of Chaos.

Let us conduct some experiment. Suppose you met someone today. You sit together at the coffee shop and chat. Then you discover that the guy was born under the same Zodiac sign as yourself. How unique is this situation? One of my friends has answered this question this way: we have 12 Zodiac signs, people are born every day, so, if we would have a company of 12 thousands people, we might have there 1,000 Aries, 1,000 Taurus, ... 1,000 Pisces. I am a Capricorn, thus my astrological twin should be among 1,000 of other Capricorns. It means that the probability to meet my astrological twin is $1:12=8.33\%$.

Does it sound scientific enough to you?

Statistically, it is not that simple. The important assumption to consider here is that we will find 1,000 representatives of each Zodiac sign among 12,000 people. In statistics, it is called "null hypothesis", and the statisticians would test this hypothesis to prove it either right or wrong. This is the most important question in statistics - to find a good (correctly formulated) null hypothesis. Our company of 12,000 people is called "a control group".

Let's consider another example. We flip the coin 100 time. Among these 100 flips, we get 65 heads. The question to consider is: what if the surface of this coin is not even? Our null hypothesis is that our coin is symmetric. Ideally, flipping it, you should get 50/50. In reality, you have got 65/35. In this case the best way is to apply the **Chi Square criterion**. To calculate Chi Square, follow this instruction:

- 1) Subtract the value you got (65) from the expected value (50) = 15;
- 2) Calculate the square of this value, $\text{Square}(15)=225$;
- 3) Divide this value (225) by average value 50 = 4.5.

This value is called Chi Square ; it shows how far the value received in our dicing experiment differs from the expected value (which is 50/50 in our example).

Now you need to use this table:

Chi Square	Probability
1.0	68%
2.0	84%
3.0	91%
3.5	94%

4.0	95%
4.5	96%
5.0	97%
6.0	98%
7.0	99.2%
10	99.8%

Find in this table the chi square = 4.5. The corresponding row gives us the probability of 96%. It means that with probability of 96% this fact is not occasional, with probability of 96% the surface of our coin is not even.

Let's continue our experiment. Suppose during the same coffee break with your friend you have found out that you two have not only the Sun in the same Zodiac sign. You have the Ascendants in the same sign. The question is: how this situation is unique?

I will form the control group another way: it will contain 1200 Aries, 1200 Taurus, etc. In total, there are 14,400 people in it. In the first sub-group of 1200 with the Sun in Aries, we are expecting to find 100 people with Aries Ascendant, 100 people with Taurus Ascendant, etc. So, within our group of 14,400 people, for each Sun sign, we have 100 persons with the same Ascendant combination. In other words, the probability to meet the person with the same Sun and Ascendant's sign is $100/14.400=0.69\%$. This is a really seldom occurring event.

These calculations are correct; however this approach does not work in practice. If we ask 14,400 people about their Sun and Ascendant's signs, we will find that the number of your double twins is different than 100. You may decide that it is necessary to perform more massive experiment and ask 14,400,000 (this number is big enough for the population of some country). But you will be surprised: the amount of your twins will be different from 100,000. It means that we have mistaken somehow.

Our wrong assumption was that, among the company of 12,000 people, 1000 should be with Aries Sun. In reality, we find that the most amount of people were born in summer and less around Christmas, this is a kind of some biological cycle. In 1998, I have got a huge database from Moscow hospitals; it contained birth data for several years. The analysis definitely showed that the probability to be born differs by the hours of the day. This is another biological cycle, a daily cycle. Thus, we deal with a not even distribution. Therefore, for the correct analysis we need to form the control group correctly; otherwise we can make a mistake.

The typical mistake #1 (not having a control group at all):

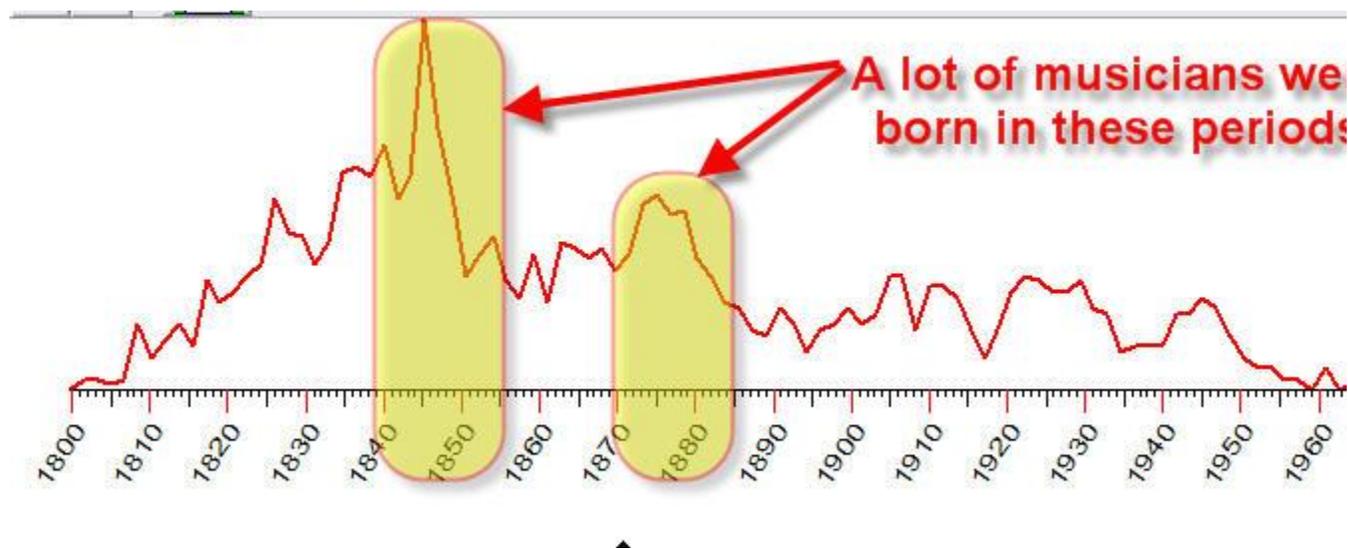
The most typical mistake in statistics is related to not correct forming of the control group. This is an example of non-correct research: "I did the analysis for 1,000 barbers and have found that the most typical Sun sign for them is Gemini". In this example, there is no control group. Let's extend our research: let us take into account the whole population of the country. Let it be England. It might happen (Ooh, a miracle!) that the most typical sign is Gemini! And we may assume that the most popular occupation in England is being a barber. You see how easy the mistake may occur. In our research, we should take into our consideration this fact, it is so called "artifact". The failure to take it into account will corrupt the results of your research.

The typical mistake #2 (the small sample size)

One more example (I take these examples from my conversations with the astrology researchers all over the world). "I did a research on barbers and collected 60 charts. I have found that 8 of them are Gemini, thus the mean value is 5". In this case, we face with a huge random error. If your friend from Ohio will collect 60 barbers in his town, he might get the totally different results. The only way to avoid this mistake is to collect as much barbers' charts as possible. So, the question of collecting data is extremely important for statistical research. We need to have at least 50 persons for each Zodiac sign, better 100.

How to do the statistical analysis using AstroInvestigator software (AIR Software)

As an example, let us do together the statistical analysis of musicians' charts. I download the database of 1,740 charts of musicians born between 1800 and 1970s. Let's look at the whole picture first. See this diagram:



This is *a histogram* - a diagram that shows the distribution of the years when these musicians were born. You see there two maximums: around the year 1845 and the year 1875. More musicians were born around these years than any other time. Now, at this stage, we may make some assumptions - based on the astrological situation at these years and some general ideas. We may start with the fact that the greatest composer Bach was born with the Sun in Aries. The peaks of the histogram are somehow related to slow planets in Aries: in 1845, Pluto and Uranus were located in Aries while in the beginning of 1870 Neptune was located in Aries. We may make these assumptions and then look for the periods when slow planets are in Aries expecting that a lot of musicians will be born. However, this is not a statistical research. To make the real statistical research, we need to form the control group. The program does it itself forming several different groups that contain exactly the same amount of random charts (1,740 in our example) within the same data range (i.e., people born between 1800 - 1970s). After forming control groups, the program performs the analysis of all these charts. It needs a lot of calculations; for example, in this particular case, the program makes about 27,000,000 astrological calculations. What is amazing is the fact that it will take only several minutes to perform this job by our program. And you will get the complete astro-statistical portrait of Musician.

The program creates a table where many different astrological parameters are considered. I would pay attention to factors like this one:

(♁)ε(♈)	107	143(132..159)	Seldom ChSq=9.1 (99.7%) r=-0.0567	✓
(♁)ε(♈)	107	143(132..159)	Seldom ChSq=9.1 (99.7%) r=-0.0567	✓
(♁)ε(♈)	181	145(133..153)	Often ChSq=8.9 (99.7%) r=0.0502	✓

It means that 181 musicians (among 1,740 observed charts) have the Sun in Aries, while only about 145 persons from the control group have the Sun in Aries as well (we have created 5 different control groups to obtain more reliable results). It means that the Sun in Aries is rather typical parameter for musicians (not an occasional one). The formal statistical analysis shows that we can rely upon this fact with the probability 99.7% (Chi square is 9.1, see the table above).

It is interesting to mention that the Sun in Libra happens seldom: only 107 times versus 143 for the control group this value is 143. I would say that this is not typical for musicians.

The program automatically compares thousands and thousands of astrological combinations; it is you who decides what combinations to consider. For example, you may like these:

♆ FIRST BEFORE ☉	148	110(98..121)	Often ChSq=13.1 (100.0%) r=0.0589	✓
♂ FIRST BEFORE ☉	149	184(172..191)	Seldom ChSq=6.7 (99.1%) r=-0.0484	✓

The program has found that a typical combination for musicians is “Neptune located before the Sun” (like Neptune is located at 10th degree of Aries, while the Sun is in the 1st degree of Aries). Also, a very non-typical situation is Mars located before the Sun.

This approach plus the application of the advanced Neural Net technology allows to use the astrology for very practical tasks. The Neural Network performs the analysis of the charts in database and creates the model of the astrological parameters for some occupation or some activity. Like in this example, after analyzing all astrological parameters, we may create a model of the Musician, a kind of the musician's portrait from the astrological point of view. Having such a model, you can compare it to the birth data of any of your client/clients. The program will rate this chart in regards to how well it fits the description of a typical musician (or a military man, or a scientist/manager/writer, etc.) chart:

Person	MusicianNN	Yes
Park Ohio Holdings PKOH	Don't Know	Yes
UMB Financial Corp UMBF	Don't Know	Yes
Herman Miller Inc MLHR	Don't Know	Yes
Bridgford Food Cor BRID	Don't Know	Yes
Cincinnati Fin Cor CINF	Don't Know	Yes
EMS Tech Inc ELMG	Don't Know	Yes
Pioneer Std Elecrn PIDS	Don't Know	Yes
Trustmark Corp TRMK	Don't Know	Yes

(Just to avoid some personal preferences or subjective opinions and to demonstrate the program's features, I used here for rating the first trade charts of some companies instead of personal charts of my friends and clients.)

This approach is very useful for human resource departments. We need just a database that contains the birth data of the people who do well something. Then the program itself does all the analysis and creates a "portrait of the ideal candidate" that should be compared to the actual data of some applicant. As you guess, for a good astrological research, the most important question is collecting the charts for different groups. And this can be done by mutual efforts of all researchers. We hope to get your help regarding this issue.

Sergey Tarassov.

Toronto, Ontario

February 19, 2007