

Statistics the science of awareness

Andrea Berdondini

ABSTRACT: The uncertainty of the statistical data is determined by the value of the probability of obtaining an equal or better result randomly. Since this probability depends on all the actions performed, two fundamental results can be deduced. Each of our random and therefore unnecessary actions always involves an increase in the uncertainty of the phenomenon to which the statistical data refers. Each of our non-random actions always involves a decrease in the uncertainty of the phenomenon to which the statistical data refers.

Introduction

This article proves the following sentence:

"The only thing that cannot be created randomly is knowledge"

A true story of a true coincidence

Ann is a researcher, is a clever and beautiful researcher, one day she decides to do the following experiment: she wants to understand if she has some special abilities that allow her to extract the number 1 from a bag containing one hundred different numbers mixed in a random way.

Day 1, Ann takes the bag and randomly pulls out a number. The drawn number is not 1, so she failed, the drawn number is put back into the bag.

Day 2, Ann takes the bag and randomly pulls out a number. The drawn number is not 1, so she failed, the drawn number is put back into the bag.

.....
.....

The days pass and Ann fails every attempt but continues his experiment.

.....
.....

Day 100, Ann takes the bag and randomly pulls out a number. The number drawn is 1, so she is successful. The probability of finding number 1 by performing a random extraction is $1/100$; this value represents an acceptable error that makes the result significant to support the hypothesis that the extraction is non-random. But Ann knows that this probability does not represent the uncertainty of her success, because this value does not take into account previous attempts. Therefore, she calculates the probability of randomly extracting the number 1, at least once, in a hundred attempts. The probability value thus calculated is 63%, since this value is very high, she cannot consider the success obtained as significant statistical data to support the hypothesis that the extraction is non-random.

Ann concludes the experiment and deduces, from the results obtained, that she has no special ability and the extractions are all random.

John is a data scientist, one day he is entrusted with the following task: he has to develop an algorithm capable of predicting the result of an experiment whose result is determined by a value from 1 to 100.

Day 1, an incredible coincidence begins, at the same time that Ann pulls a number John tests

his own algorithm. The number generated by the algorithm does not coincide with the result of the experiment, so he failed.

Day 2, the incredible coincidence continues, at the same time that Ann pulls a number John tests a new algorithm. The number generated by the algorithm does not coincide with the result of the experiment, so he failed.

.....

.....

The days pass, the coincidence continues and John fails every attempt.

.....

.....

Day 100, the incredible coincidence continues, at the same time that Ann pulls a number John tests his new algorithm. The number generated by the algorithm coincides with the result of the experiment, so he is successful. The probability of predicting the result of the experiment by running a random algorithm is 1/100; this value represents an acceptable error that makes the result significant to support the hypothesis that the algorithm used is non-random.

For this reason, John writes an article in which presents the result obtained. The article is accepted, John is thirty years old and this is his hundredth article.

Awareness breeds awareness

We call "researcher" a person who knows only his own attempts regarding the study of a certain phenomenon.

We call "reviewer" a person who does not actively participate in the study of a particular phenomenon but knows every single attempt made by each researcher.

Researcher 1: develops an algorithm that obtains the result R1 with respect to a phenomenon F. The probability of getting a result equal to or better than R1 in a random way is 1%.

Researcher 2: develops an algorithm that obtains the result R2 with respect to a phenomenon F. The probability of getting a result equal to or better than R2 in a random way is 1%.

Reviewer: defines a new result $RT = R1 \cap R2$. The probability of getting a result equal to or better than RT in a random way is 0.01%. Consequently, the uncertainty of the result RT is 0.01%.

The absence of awareness reduces awareness

We call "researcher" a person who knows only his own attempts regarding the study of a certain phenomenon.

We call "reviewer" a person who does not actively participate in the study of a particular phenomenon but knows every single attempt made by each researcher.

Researcher 1: develops an algorithm that obtains the result R1 with respect to a phenomenon F. The probability of getting a result equal to or better than R1 in a random way is 1%.

Researcher 2: develops an algorithm that obtains the result R2 with respect to a phenomenon F. The probability of getting a result equal to or better than R2 in a random way is 100%.

Reviewer: defines a new result $RT = R1 \cap R2$. The probability of getting a result equal to or better than RT in a random way is 2%. Consequently, the uncertainty of the result RT is 2%.