TOK ESSAY

**“Statistics conceal as much as they reveal.” Discuss this claim with reference to two areas of knowledge**

Session: May 2021

Word count :1595

Personal code: hzz954

"We must be careful not to confuse data with the abstractions we use to analyze them."[[1]](#footnote-1)

- William James.

Statistics are used as evidence, in arguments regarding natural sciences, because it has been established that they are a reliable source of information when relating to this area of knowledge. The quote above challenges this belief of reliability and brings the following question: To what extend is contextual understanding essential in understanding knowledge claims in Natural Science? Natural Science is defined as: “A science or knowledge of objects orprocesses [and/or numbers] observable in nature, as biology or physics, as distinguished from the abstract or theoretical sciences, as mathematics or philosophy.”[[2]](#footnote-2) In history statistics are used as evidence to support an argument, but historians are more likely to evaluate the statistics, their origin and purpose before arguing and using such evidence. History in TOK “does not simply refer to everything that has ever happened. Instead it is about the past that has been recorded by human beings. One of the major debates that you should consider in history is interested in the idea of the ‘historical fact’”.[[3]](#footnote-3) To what extend are historical claims based on statistics reliable at different times after the events?

Statistics in natural sciences conceal some information when there is a lack of context to explain the results. Experiments and their results in Natural Sciences base themselves on statistical result and the correlations of such results, but sometimes, the final results are misleading if they are put in the correct context. One example, of this is the Simpson paradox that shows that sometimes trend lines that appear in different groups of data will be reversed when the data groups are combined. The example of statistics concealing information came to me during a debate in physics class. At the beginning of the pandemic, I looked up whether people died more from COVID (per contamination) in China or in Italy, and the Simpson paradox appeared to me as every age group had a higher death rate in China, but overall, Italy had the higher death rate. This example illustrates the Simpson paradox, but does not reveal how and why the correlations contradict themselves, so I conducted research to understand, and found a document that has statistics that illustrate and explain the Simpson paradox.[[4]](#footnote-4)

Figure1: example of the Simpson Paradox

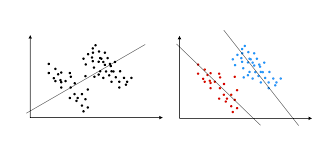
[[5]](#footnote-5)

Figure 1 illustrates graphically how the Simpson paradox work with a general positive correlation, but two distinct results with each a negative correlation. It shows that when statistics are not taken into context, the answer that they bring might come out as absurd and completely illogical, but that when the number of people in each age group for the COVID example, or number of courses taken each year by the students, then the answers are studied with reason and start to make more sense. This is remodeling the way that statistics should be seen and used in Natural Sciences because they are not true, as depending on their use they can say anything and its opposite, they can induce massive bad decisions concerning health, research and development.

It can however be argued that in the case of Natural science, when observing a result and looking at it with reason, statistics will always reveal the result, and that it is only the language, imagination or emotional response of the scientist to the statistic that will make it conceal some information that the experience has shown through the data. The idea that statistics can conceal information is therefore considered to be erroneous. For example, in physics HL, my IA consisted of measuring the effect on the refractivity of a salty water solution, depending on the amount of salt in the water. I measured angles and then used different formula to calculate the different index of refraction. The index of refraction changed from 1.21 without salt to 1.43 with the maximum amount of salt that I added. The statistics that I employed and calculated speak for themselves, there is no way to use them to say any other thing than what they explain. the interpretation of statistics for other experiments can however accidentally or purposely hide parts of the truth. If statistics concealed as much information in Natural Science as they reveal, they would not be used nearly as much to support knowledge claims and scientific consensus would not exist due to the accidental or purposeful hiding of information every time the experiment would be repeated. After looking at these arguments, when looking at statistics in natural science, it is important to ask ourselves the question of where it comes from and if the experiment was done fairly and if outliers were deleted to make the statistics fit with the wanted results. This is why using reason and research when historians are using statics is a crucial aspect as it allows a historian to use historiography and to evaluate the context and source to examine the reliability of a statistic.

No one can contest that WW2 happened, but the numbers behind the war can be applied differently depending on the point of view of the person arguing and using them. This use of statistics in history leads to the creation of different views that are orthodox, revisionist and post revisionist. The two first views are usually created by a use of statistics in opposite ways. While the orthodox historians will have one opinion using one set of statistics, the revisionists which come short after have an opposite view that they support using statistics from different origins, or interpreting them with a different view, thus using statistics around the same facts to support two opposite conclusions. After time passes, the third view, which is called post revisionism often emerges due to hindsight and it is more of a blend of the two other views. In the Western World, it was first believed that Adolph Hitler had singlehandedly brought Germany and its economy back to relevance. The orthodox historians who supported this view used the economic statistics of Germany from the economic crash of 1929 to the beginning of WWII to prove what they said. Indeed, after Hitler arrived in power, the economic sate of Germany was much better and employment numbers dropped significantly. However, the unemployment statistics did not include all the labor force and the economics statistics are taken without the statistics of the rest of the World to see if he really did it without any help and while other economies were not getting back on their feet as fast. Some years after, revisionist historians looked at the situation in Germany and said that the economy started to grow back before the end of 1932 which is why the NSDAP and Hitler had seen their scores at the elections lower a little. Then looking at statistics from other countries, the revisionist historians started arguing that Germany was just following the movement and that the economy would have gotten better without Hitler. After a longer period, post revisionists came around by mixing the two arguments. Historians such as Ian Kershaw that have been established by historiography as trustworthy, started arguing that though the world was picking its economy back up, the German economy was seeing more of an extreme rise and that Hitler was a reason for this as his short-term economic policies guaranteed a shot term growth with a large debt that would be erased if he could fulfill his goals of invading USSR. All these views are supported by historians that use different statistics as evidence and proofs, but that leave other statistics (that would weaken their claim) on the side. Historiography which is the evaluation of a source by finding its values and limitations through the author, purpose and content is very useful in these situations. When there is a clear evaluation of how reliable a source is, the historian is then able to use the said statistic, or evidence with knowledge of how much it conceals and how much it reveals, thus meaning that the statistic should be used in context and explained to understand the entire fact and appreciate the different views that emerge from it. For example, in my History IA about May 1968 events in France and their impact, I had to evaluate the reliability of people giving me different statistics about the number of protestors. Policemen wanted to reduce these numbers to make the movement seem less important and Protestors inflated these numbers to encourage people to join them and make the movement seem bigger than it actually is.

The two areas of knowledge which are History and Natural science have a lot of differences and commonalities with regards to the way statistics are used to reveal or conceal information. The arguments are usually supported by statistics in both AOKs, but the main difference is that in History, statistics become available and are used to create a different point of view with every new generation of historians. In Natural Sciences, there is much more of presence of statistics from the beginning and if there is a change, it does not occur periodically. Historiography should be applied to both areas as the context can be very useful to determine the accuracy of knowledge brought by the statistics.

In conclusion, I believe that statistics when put with the correct context and amount of explanation will reveal much more than they will conceal, however, when used in an argument, or when lacking the sample size for scientific results, they will conceal as much if not more information in both natural sciences and history, therefore becoming a liability to the creation of knowledge.

# Bibliography:

<http://www.math.wpi.edu/Course_Materials/SAS/quotes.html> accessed 20th/02/2021

<https://www.wordreference.com/definition/natural%20science#:~:text=a%20science%20or%20knowledge%20of%20objects%20or%20processes%20observable%20in,sciences%2C%20as%20mathematics%20or%20philosophy>.

https://www.lanternaeducation.com/ib-blog/theory-of-knowledge-ib-guide-part-7/

<https://towardsdatascience.com/simpsons-paradox-and-interpreting-data-6a0443516765> accessed the 15th of February

1. <http://www.math.wpi.edu/Course_Materials/SAS/quotes.html> accessed 20th/02/2021 [↑](#footnote-ref-1)
2. https://www.wordreference.com/definition/natural%20science#:~:text=a%20science%20or%20knowledge%20of%20objects%20or%20processes%20observable%20in,sciences%2C%20as%20mathematics%20or%20philosophy. [↑](#footnote-ref-2)
3. https://www.lanternaeducation.com/ib-blog/theory-of-knowledge-ib-guide-part-7/ [↑](#footnote-ref-3)
4. [↑](#footnote-ref-4)
5. <https://towardsdatascience.com/simpsons-paradox-and-interpreting-data-6a0443516765> accessed the 15th of February [↑](#footnote-ref-5)