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

1599 Words

Personal Code: jbb161

Statistics are a way of making sense of a large set of unprocessed data in an understandable way, in a numerical or visual manner. While statistics can help make knowledge claims easy to absorb or analyze, information can be misrepresented, changing the viewpoint of the prospective audience. Nevertheless, statistics are used differently depending on the area of knowledge and the type of statistic used. In a Student’s t-test for instance, the probability of an event being related to an assumption is reduced down to a single number without concealing much information, however, Real GDP in economics cannot reliably be used to compare standards of living. It is worth noting that statistics is a far wider field than can be dealt with in this essay.

To a certain extent, I agree with the statement, statistics in the human sciences tend to conceal as much they reveal when used by themselves, but in mathematics, information can be revealed in a more organized and objective manner. Contrarily, when statistics in the human sciences are used in conjunction with each other, they tend to reveal more than they conceal. Likewise, when mathematical statistical measures are used out of context, they are more likely to conceal as much as they reveal. However, almost all statistical measures do conceal information in some manner by compressing a large set of data into one, so the validity of the statement also depends on how "conceal" and "reveal" is defined.

The following argument applies to all people, whether experts or not.

A one-sample Student's t-test — a statistical method in the mathematics — is a type of statistic used to test whether the mean of a sample of data is likely drawn from a *normally distributed* population [The Editors of Encyclopaedia Britannica, 2020], and this area of knowledge contains information that is more revealing than concealing as conclusions are drawn from numerical data without any extra extrapolation needed. For example, in the medical field, a drug can only be introduced if the data shows a statistically significant difference to the original condition.

This test returns a p-value which is the probability that an observed difference happened due to random chance, and in the case of a t-test, the probability of experimental data being normally distributed around a certain value. For example, an experiment in 2020 by a student regarding comparing an optimal angle for a basketball shot to experimental data used a t-test, which found whether the average experimental release angle was normally distributed around an optimal angle. This process revealed information: without the p-value, it was difficult to glance at the data to find if it was normally distributed. Additionally, after finding the p-value, information about the relation to the optimal angle was brought to a p-value, allowing for an objective numerical comparison. In this experiment, the t-test's p-value was 0.869, meaning a high probability that that the angle was normally distributed around the optimal angle. In this sense, if "reveal" were defined as providing new information otherwise unseen, the p-value in this experiment revealed more than it concealed.

On the other hand, the p-value brought all data points into a single numerical value, concealing information about the distribution and spread of the data. If a reader looked at the p-value without knowing anything else, they would not realize that the data was positively skewed. This shows an issue that statistical measures in the mathematics have — they conceal data by bringing a large group of data into a single medium. If "conceal" is defined as hiding information that would otherwise change a conclusion, in this experiment, readers can't see the distribution using only the p-value, so there is as much "[concealing]" of whether the data is normally distributed as there is "[revealing]" when given without context. However, when the p-value is presented alongside a frequency diagram of the experimental data, readers could easily tell that the student tended to skew to shooting below the optimal angle. Therefore, with more data, statistical measures can reveal more than they conceal.

All things considered, statistics such as the t-test in mathematics conceal as much as they reveal when given without any other context, as information about the data itself is condensed and hidden away from the reader. Contrarily, the p-value by itself still gave the student a viable method of supporting his hypothesis. Ultimately, the extent of this depends on how the statistic is used. When used to provide an extra insight, many statistics are going to be more revealing than they are concealing. However, when used to claim something as objective truth, it is likely that statistics conceals as much as it reveals, if not more.

In the Human Sciences however, statistics conceal as much as they reveal due to the way these measures are calculated and what they choose to leave out, as well as what they seek to assess. In areas of the human sciences such as economics, statistics are calculated using surveys, observation, and sampling amongst many other methods, which all leave out some information and can require classification, thus concealing data from readers. One example of this is measuring standards of living, which is typically measured with Real GDP. These different factors can only be calculated to a certain degree of accuracy because each country calculates Real GDP differently, and also because it's impossible to record down every consumer transaction in an economy, meaning that the value of Real GDP recorded by each country will vary depending on the method and the demographic of the population. It is also an average, meaning that it fails to account for differences in income distribution. For example, countries in the middle east have high levels of Real GDP but also high levels of poverty, where in Scandinavian countries, the opposite is true. Moreover, Real GDP only considers factors of income and expenditure, meaning that it leaves out other factors of living standards such as the percentage of people living in poverty and education standards. Additionally, some aspects of the human sciences may not be suited for statistical analysis. For example, measuring living standards have far more non-statistical factors to consider such as hope or family unity, therefore measuring it is difficult. Leaving out these factors for living standards in this context conceals as much as it reveals — real GDP may provide misleading results for living standards among different countries.

Hong Kong and Sweden both had similar values for Real GDP per capita in 2019 — with Hong Kong having a value of USD $48,756 [macrotrends, 2020a], and Sweden having a Real GDP per capita value of USD $51,610 [macrotrends, 2020b] — so if we were to only use Real GDP to compare standards of living in these countries, we would say that Hong Kong and Sweden have similar living standards. However, we can get a better picture of living standards when given more data points. For example, if we look at poverty rates in both countries, Hong Kong had a poverty rate of 20.4% in 2019 [Census and Statistics Department, Hong Kong, 2019], while Sweden had a poverty rate of 9% in 2012 [Matthews, 2015], meaning that there was a significantly higher rate of people living under the poverty line in Hong Kong compared to Sweden, which is one factor showing that Sweden has a better standard of living than Hong Kong. By providing an extra data point to make a better judgement, more information is revealed, allowing for a stronger conclusion to be made. Therefore, in the human sciences, as long as there are enough relevant data points to make a judgement, people are able to make a better judgement of the implication of these statistics, and as a result, there is more information revealed rather than concealed.

Ultimately, statistics in the human sciences conceal as much as they reveal when used as the only metric to compare differences between samples, as these samples may have different demographics — making it harder to directly compare them — and methods of collecting data — leading to different results — and thus the same statistic in the human sciences may still paint an entirely different picture in reality than what the number may say. However, when a statistic such as Real GDP is combined with other statistics such as the poverty rate to make a better judgement, more information is revealed than concealed, and with each additional relevant statistic, more information is revealed to help make a better final judgement.

In conclusion, the extent of which statistics conceal as much as they reveal data depends on the area of knowledge, from which the methodology is derived from, the context behind how the data is used, and the definitions of “conceal” and “reveal”. If “conceal” were simply defined as hiding information, then almost all statistical methods regardless of area of knowledge would conceal as much as they reveal because almost all statistics combine a large set of data into one, thus hiding information. However, if “conceal” were defined as hiding information that would otherwise change a conclusion, it varies more depending on the area of knowledge and the specific statistical measure. In some areas of mathematics for instance, statistical measures are less likely to conceal information than those from some areas of the human sciences due to the complexity involved with some of the human sciences. Moreover, this analysis only covered a small portion of both mathematics and the human sciences — the scope of these areas of knowledge extends far beyond what is written. Regardless, in both areas of knowledge, statistics reveal more than they conceal when paired with other statistical measures in a non-malicious manner, but conceal as much as they reveal if used as the sole basis for a conclusion in most situations.

References

Census and Statistics Department, Hong Kong (2019). Poverty Situation. `https://www.censtatd.gov.hk/hkstat/sub/so461.jsp`. Visited on December 06, 2020.

macrotrends (2020a). Hong Kong GDP Per Capita 1960-2020. `https://www.macrotrends.net/countries/HKG/hong-kong/gdp-per-capita`. Visited on December 06, 2020.

macrotrends (2020b). Sweden GDP Per Capita 1960-2020. `https://www.macrotrends.net/countries/SWE/sweden/gdp-per-capita`. Visited on December 06, 2020.

Matthews, D. (2015). Denmark, Finland, and Sweden are proof that poverty in the US doesn't have to be this high. https://www.vox.com/policy-and-politics/2015/11/11/9707528/finland-poverty-united-states. Visited on December 06, 2020.

The Editors of Encyclopaedia Britannica (2020). Student's t-test. https://www.britannica.com/science/Students-t-test. Visited on December 06, 2020.