**DRAFT – PART 3**

Today we live in the Information Age where we understand a great deal about the world around us. Much of this information was determined mathematically by using statistics. When used correctly, statistics tell us any trends in what happened in the past and can be useful in predicting what may happen in the future. But as Brooks says in one of his article, “there are many things big data does poorly.” One of the problems here is correct interpretation of data and its reliability. As Brooks states in his example of use of words “I”, “me”, “mine”, people with confidence use a fewer of those words and vice versa. But we can look at it from other perspective. What if “confident” people just know the psychological effect of those words and, in order to persuade someone, avoid using inappropriate one? Data can’t register one’s intentions, thoughts or moral condition; it can only display visible events and facts.

The article, “Use and Misuse of statistics,” presented by Harvard Business school, talks that before we use any data we should know how much this is reliable. In order to accomplish that, we should be clear about purpose of using the data, what we want to discover. For example, customer satisfaction survey results in the arithmetic mean or average of a group of numbers and equal, let’s say, 3.5 on a scale of 1 to 5[4]. But in reality it could be that no one gave to a product a rating of 3.5. Instead, the responses could cluster around a group of very satisfied customers, who scored it a 5, and unsatisfied customers, who gave it a 1. In this case the mean isn’t the most helpful metric for research.

Real life is more complicated than data report, so we don’t have to take cause and effect by granted. If we do such-and-such, then such-and-such will happen. The desire if not requirement that data must be used with every decision creates paralysis. In present world we are taught to seek perfection, but sometimes we forget that the one thing more important than perfection is simply progress[3]. If everything should base on statistics, we would never have great unexpected breakthroughs in the human history.

With statistics, we can’t prove things with 100% certainty. For instance, people, recording survey results, may be dishonest or sloppy in those results. That question has emerged with survey conducted by two criminologists that has raised doubts about the integrity of the New York Police Department's highly regarded crime tracking program, CompStat. Relying on the anonymous responses of hundreds of retired high-ranking police officials, the survey found that tremendous pressure to reduce crime, year after year, prompted some supervisors and precinct commanders to distort crime statistics[1].

The biggest limit to big data is our ability to interpret it. [2] Gordon B. Drummond in his work “Data Interpretation: Using Probability” talks about principles of data interpretation. One of his key points is that we should ensure that a sample studied was right chosen and random. As Drummond states, “It’s possible that some scientists are not even clear that the word ‘‘sample’’ has a special meaning in statistics, or understand the importance of taking an unbiased sample,” we can never be certain that a sample will exactly reﬂect the properties of the entire group of possible candidates available to be studied. Drummond suggests planning study, establishing hypothesis and estimating the probabilities that the observed data could have occurred by chance, “A properly designed study that aims to answer speciﬁc questions will have deﬁned outcomes of interest at the outset, before data collection has started. These questions are then recast as hypotheses that need to be tested.” At the end, when we draw a conclusion, we should consider that absence of evidence in any study is not evidence of absence [2]. If we can’t detect or analyze something, there is no prove it doesn’t exist.

Brooks in his “What Data Can’t Do” publication states that raw data have been structured and analyzed by people who use their own values to draw a conclusion. I can’t disagree with this point. Computers can collect the data but only human beings with their own prejudices, gaps in education and sympathy to a certain things will draw final conclusions. People can’t be 100 % impartial. They will always see things through one’s life experience and can easily bend the accurate data sources by simply asking and changing the question to suit their end goals.

In spite of all flaws, everything we do now in modern world is data driven. Weather forecasts, academic success, politics, stock market, etc. – it all depends on data analysis, which is the best way to understand the present and the past. Big data has its uses, but we should remember that just because we have a lot of data doesn't mean we have the right data to answer a particular question. Data is only useful if it is honestly and thoroughly gathered. When we are summarizing and interpreting data we shouldn’t blindly rely on raw facts. In order to understand and predict the future outcomes we need to see the problem from different aspects and be maximum objective.

**Work cited**

1. Chen W. David "Crime Survey Raises Questions About Data-Driven Policy." The New York Times. February 9, 2010. LexisNexis Academic. Web. Date Accessed: 2013/05/04.

2. DRUMMOND, GORDON B., and SARAH L. VOWLER. "Data Interpretation: Using Probability." Microcirculation 18.5 (2011): 358-360. Academic Search Complete. Web. 4 May 2013.

3. Kevin Meyer. “Data, Facts, and Truth.” evolvingexcellence.com. 24 February 2013. <http://www.evolvingexcellence.com/blog/2013/02/data-facts-and-truth.html>

4. “The Use and Misuse of Statistics.” [business.nd.edu](http://business.nd.edu/). Harvard Management Update, May 2000. Pp. 6-7. <http://business.nd.edu/uploadedFiles/Current_Students/MNA_Students/2013/MisuseOfStatisticsHBR2000_SU13.pdf>