STATISTICAL INFERENCES AND CREATIVE THINKING

Syed Imtiaz Ahmad

Computer Information Systems
Eastern Michigan University[1]
Ypsilanti, Michigan 49197
cis_ahmad@online.emich.edu

"I can see nothing," said I, handing it back to my friend. "On the contrary, Watson, you see everything. You fail, however, to reason from what you see. You are too timid in drawing your inferences."

Dr. Watson and Sherlock Holmes
The Adventure of the Blue Carbuncle

Abstract

This paper discusses the use of statistical inferences in creative thinking and problem solving. It forms a subset of the larger work on developing a university-wide course on Creative Thinking and Problem Solving. The conventional works on creative thinking, critical thinking, constructive thinking, and problem solving rely primarily on techniques used in disciplines such language and logic. The work presented in this paper offers the use of statistical concepts and methods as meaningful components for creative thinking and problem solving activities.

[1] This work was initiated at International Islamic University Malaysia (IIUM) while the author was serving there as Dean of the Faculty of Engineering, and Chairman of IIUM Senate Committee for a university-wide course on Creative Thinking and Problem Solving.

1.0 INTRODUCTION:

The field of statistics provides us with concepts and tools for giving meaning to naturally observed or experimentally generated data on objects that may be of interest to us. In dealing with people, for example, we may handle data on gender distribution, age, height, weight, income, religion, family size, measurements in psychology such reaction time, indicators of intelligence, aptitude test scores, and so on. The values these data and measurements seem to follow some natural distribution. This tendency in nature was highlighted in a statistical classic, "On the Laws of Inheritance in Man," published in Biometrika in 1903. Given the weight of a person, for example, we may wonder whether this weight is below or above the mean for the population. How do we determine the mean weight for a population? Certainly not by asking everyone's weight. This is neither advisable nor feasible. Through, statistical methods we can come up with a reasonably good estimate for the mean value of the entire population by properly selecting a small sample from the population. Statistics also provides us with the means to discover patterns in data, to draw inferences from data, and to predict the outcome of events.

The use of statistics is important in view of our need to understand all kinds of data being collected and disseminated in public. Often the public statements are based on quantitative data. Without some knowledge of statistics, and an appreciation of its underlying principles, it would be difficult to understand or question these statements.

Statistics may be defined as the science of data. There are some outstanding recent works sponsored and published by American Statistical Association (ASA) on the role and history of statistics as a key area of human scientific endeavor. There are two articles to be noted in particular i.e. "Shaping Statistic for Success in the 21st Century" [Ketterning ,1997], and "A Voyage of Discovery" [Billard, 1997]. The first work presents insights into statistics and the current state of its development. The second work presents historical developments in statistics as a science, based on a review of key articles published over one hundred years, from 1839 to 1939, in the Journal of the American Statistical Association (JASA). Both of these works are available as full length articles on the World Wide Web at www.amstat.org, the ASA web site

2. CREATIVE THINKING AND STATISTICS

How do we use statistics in thinking more creatively about the things we see or record? After all, one may say that statistics is about calculating, describing, manipulating and interpreting mathematical attributes of sets or population. This may be viewed more as formalizing rather than cultivating creative thinking. Before we raise more questions about statistics, let us articulate a little about creative thinking. Let us simply not look into creative thinking but also constructive thinking, critical thinking, and so on. What is creative thinking? This question is asked and answered variously ever so often. We are raising it again to refresh our thoughts and to examine the role of statistics in creative thinking.

Creative thinking is about looking at possibilities in understanding objects or phenomena. It generally involves describing objects or phenomena, making projections on what is likely to happen in observed objects or phenomena with movements in time and space, and taking actions about objects or phenomena in order to move them in a desired direction. Here, the word objects is used to refer to things that interest us. Phenomena are what we experience about things that interest us, not necessarily derived from what those things may be within. The word creative in creative thinking implies emphasis on looking at possibilities. The word constructive in constructive thinking implies emphasis on building solutions from the possibilities, and the word critical in critical thinking implies emphasis on analysis of the possibilities and questioning the interpretations based on the possibilities. Often, we may use creative thinking, constructive thinking, and critical thinking as one and the same, keeping in mind the emphasis if there is one.

Let us do a little experiment. After all statistics is often about experiments and drawing conclusions from data gathered in experiments. In that sense, we may say that statistics is about discovering stories that numbers may tell. Consider, for example, a business is interested in knowing how its employees feel about some corporate values that it espouses. Let us say that this value is, "We take pride in our teamwork." There are twenty employees. They are asked to rate this value statement in one of the following categories:

Strongly Disagree,
Disagree,
Neither Agree nor Disagree,

Agree, or
Strongly Agree.

In order to 'simplify' matters, the above choices are assigned a numeric value ranging from 1 to 5 i.e.

Choice	Numeric Value
Strongly Disagree,	1
Disagree,	2
Neither Agree nor Disagree,	3
Agree, or	4
Strongly Agree.	5

The responses from the 20 employees in the business are recorded below as numeric scores for convenience:

5 3 5 5 2 3 3 5 1 5 4 4 2 4 5 4 1 4 2 3 (1)

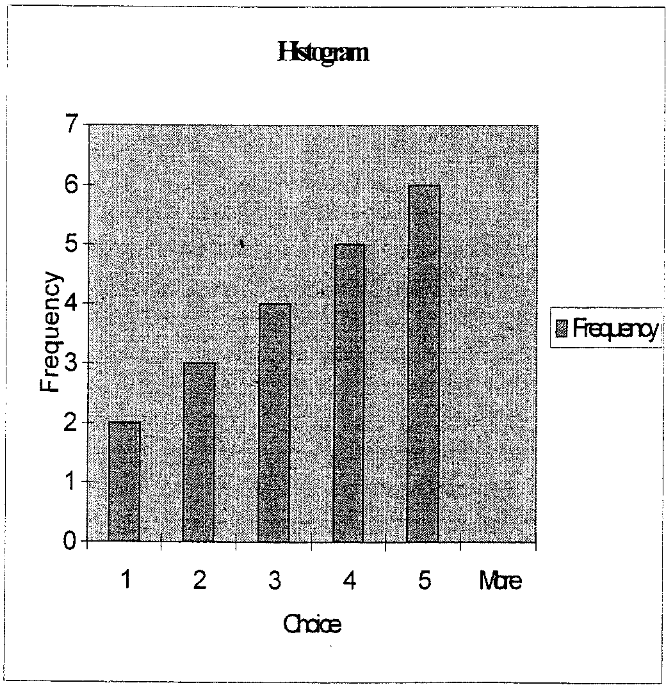
What can we discern from these responses? Do they support, "We take pride in our teamwork," the value espoused by the business? We can draw some conclusions in browsing the above scores directly. However, we can try to be a little more creative and arrange the score in a way that makes us easy to see the pattern and draw conclusions. Here are the same scores arranged according to the frequency of their occurrence:

Score	1	2	3	4	5	
Frequency	2	3	4	5	6	(2)

We have opened some new possibilities in looking at the scores. We can see easily that more than half the people agree or strongly agree with the stated value, "We take pride in our teamwork." One fifth neither agree nor disagree with the stated value, and one fourth disagree or strongly disagree with the stated value. In re-arranging the scores as shown, we reduced the number values for study into five groups. Handling five things in drawing conclusions is a lot easier than handling twenty things. This is simply in the observed nature of human beings. This observation on human beings is based on an empirical study of individuals and organizations by a team consisting of a psychologist and a mathematician. Their study produced a heuristic called "7 plus or minus 2" rule. It means that mental skills work best when we deal with five to nine things that require our attention at the same time. Attending to less than five things at a time, our mental skills may be under utilized, and attending to more than nine things at a time we may not be able to attend to all of them well. The numbers five to nine presumably are related to the capacity of short term memory in humans – that part of our memory where we retain things for ready recall when engaged in a task. Creating piles of things and categorizing things into objects is what we do all the time. For example, we may categorize fruits into apples, oranges, and so on rather than talking about them as a single category of all fruits, just as we have tried to do in the above example of scores. Rather than working with twenty individual scores, we arranged them into five categories corresponding to the five categories of responses. It then became easier to talk about what the scores may tell us regarding the espoused corporate value, "We take pride in our teamwork."

We could keep on analyzing the scores further in their numeric form. However, it is often helpful to visualize what is happening, or the implied patterns of things, when we show them in a pictorial form. We may graph the numeric data into a picture. One such pictorial representation is shown in Figure 1, corresponding to the numeric data in (1).

Figure 1: A histogram for scores in (1). This is a bar graph in which the area (height) of each bar is proportional to the frequency of values for a particular category of data. The horizontal scale shows the categorization scheme.



The graphical view often allows us to perceive the patterns contained in numeric data more easily. For example, looking at Figure 1, the staff members in this business are not divided evenly between those who strongly disagree with the value statement, "We take pride in our teamwork," and those who strongly agree. Also, one may wonder why as many as four persons out of twenty, a significant number neither agree nor disagree with this statement. Is it because they do not quite understand what the statement says, or that they are simply not interested in the state of affairs? There is some food for further thought. Five out of twenty persons disagree, and two of them strongly, that the statement, "We take pride in our teamwork," reflects the true state of affairs in the business.

Taking the raw data shown in (1) and arranging it as shown in (2) or Figure 1, did open the possibilities to deal with the value statement, "We take pride in our teamwork," for this business. It opened the doors to creative thinking, it allowed critical thinking by facilitating analysis, and in at least some small ways, it allowed the business to constructively think about the possibilities.

The data in (1) includes all employees. Therefore, we can be quite confident that our conclusions reflect the true state of affairs for the entire business as seen from the perspective of employees. What if the number of employees was very large, and it was considered neither necessary nor practical to engage all employees in responding to the value statement, "We take pride in our teamwork." What if we simply took the responses from a reasonable sample of employees and drew conclusions as if they were based on the responses of all employees? This is a new possibility worthy of consideration. However, would we place a lot of confidence in our conclusions if we work with this new possibility? Well, it all depends on how representative of all employees is the sample of employees we chose? To raise questions when something new enters in what we may have been doing unquestionably before is part of creative and critical thinking. We will deal with the issues of taking a **sample** and making conclusions about the **population** from which the sample is derived, in later sections. For now, we will say that the sample of values represents a random selection of values from the population i.e. it is free of bias in favor or against any specific segment of population.

Sometimes, we may be interested in very simple categorizations. For example, which way did the majority respond? Did they tend to disagree and strongly disagree or agree and strongly agree? If we assume a center between those who agree and those who disagree then which side weighed more heavily? We can learn this and other interesting features in the scores by simply arranging them in increasing or decreasing order of magnitude. The sorted scores in increasing order are:

1 1 2 2 2 3 3 3 3 4 4 4 4 4 5 5 5 5 5 5 5
(3)

Separating at the mid point, we may rewrite (3) as follows:

1 1 2 2 2 3 3 3 3 4   4 4 4 4 5 5 5 5 5 5 5
(4)

More than half the persons responded with agree or strongly agree. The value at this mid point is 4. In statistics, it is called the **median**. For an even number of observations, the median is the average of two values in the middle. The median of a set of observations is an indication of central tendency. When these observed values are arranged in increasing order of magnitude, half of the observations will be less than this value and the other half will exceed this value. The median value of 4 for the scores means that the majority agreed or strongly agreed with the statement, "We take pride in our teamwork."

There is another statistic, called **mode,** noting something of interest about observed values. It is a singular categorization picking up the value that occurs most frequently. For the observations listed in (1), the mode is 5, meaning that more persons responded by strongly agreeing with the statement, "We take pride in our teamwork" than those responding in any other single category.

Both median and mode provide insights in what we observe. They are statistics and they require calculations or manipulations in order to gain insights into what we observe. The focus is not on calculations or manipulations. They have to be done. The focus is on gaining insights into what we observe, and this in some small ways is always likely to assist in thinking creatively, constructively, or critically about what we observe.

The sorted arrangement also helps us spot two other simple statistics in the given data i.e. the **minimum** and **maximum** values.

Now turning to another possibility. What if the recorded values were not from a discrete set of possible integers such 1, 2, 3, 4, and 5 in the preceding example? Rather the values may take on fractions such as 1.3 or 2.5. Consider the following example of height measurements in feet and fractions of feet for a sample of twenty people:

5.5, 6.5, 6, 5.8, 6.1, 6.2, 5.6, 6.6, 5.4, 5.9, 5, 6.2, 6, 6.7, 4.8, 7.1, 5.3, 5.8, 6, 5.9
(5)

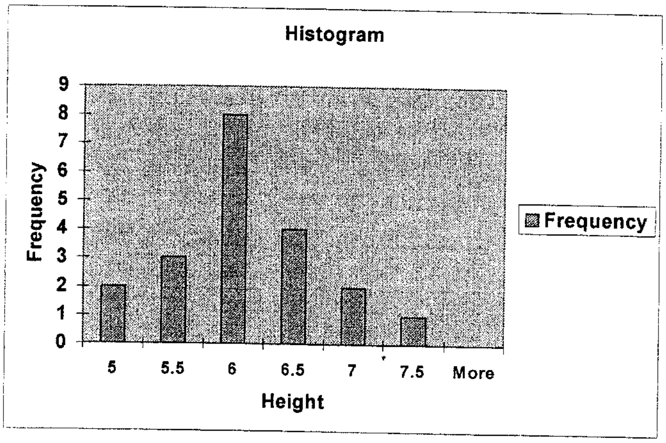
How do we put these values into a small number of categories? As an example, we may define them as follows:

Category #:	1	2	3	4	5	6
Height	5	5.5	6	6.5	7	7.5

(6)

First category gathers all values less than or equal to 5 in height. We may say that this covers the range from 4.5 to 5, with mid point at 4.75. The second category gathers values over 5 and those less than or equal to 5.5, with mid point at 5.25, and so on.

Figure 2: Graphical representation of data listed in (5) using categories (6)



The values listed in (5) and categorized according to (6) produce the following results:

Height	5	5.5	6	6.5	7	7.5
Freq.	2	3	8	4	2	1

(7)

The graphical representation of these results is shown in Figure 2.

In evaluating the results shown in (6) and Figure 2, we may draw many conclusions as before. But these conclusions may apply only to the sample. Their generalization to the population from which the sample is drawn is well warranted for a good sample. However, these generalizations may have to be qualified with the degree of confidence we may be able to associate with them. For example, the sample indicates clearly that majority of people measured are nearly six feet or taller. How confident are we that this represents the true state of affairs for the population from which the sample measurements have been taken? This leads to questions about the sample itself. For example, does the sample appear to be typical or normal for the kind of thing we are measuring? We know that for a measurement like the heights of people, there some perceived middle or average value that typifies most people. As we move away from this middle or average value, the number of people having those heights would diminish. The pattern of diminishing values away from the middle value represents a tendency in nature, through some mysteries in nature. The resulting pattern is called a bell curve.

The frequency distribution data in (6) and Figure 2 is redrawn in Figure 3, showing a curve that results from connecting the top points of all frequency value bars in Figure 2. It is indicative of a bell shape, although
It is not a very smooth bell curve. The 'quality' of this bell shape determines, in ways, the quality of our generalizations. A better understanding of this issue will be developed in later sections.

3. RAW DATA AND DATA AGGREGATIONS BY CATEGORIES:

Sometime the raw data is simply overwhelming. We may not be able to deal with it without suitable aggregation by categories even if those categories are not obvious, as may be seen in (1) and (5) described previously. As before, these aggregations can help us to see underlying patterns more easily. Consider the example of 40 data values shown in Table I. One may say that 40 values of some data need not be overwhelming in many situations. That is very true. We are not looking at the number 40 but whatever number is found to be overwhelming in a given situation. Sometimes, it is also a question of effectiveness. If a reduced set of things is likely to lead us to the same conclusions as a large set then why not conserve our resources in measuring, handling, and computing by using the reduced set. We may also be more efficient in our work and more effective in conveying it.

Figure 3: Frequency distribution of heights shown as a curve indicating the peak and tapering of the values away from the peak.

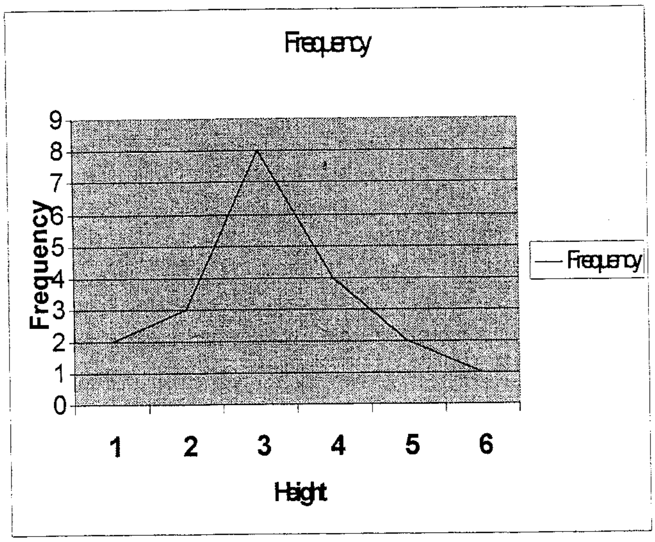


Table 1: Raw Sample Data Values

67 53 83 69 33 17 39 74 85 41 35 70 58 49 42 43 62 40 63 21

66 79 76 27 60 51 70 96 60 71 90 37 56 45 48 54 25 75 55 44

This data represents examination scores in a class. However, the same data may represent a random sample of annual income of people (in thousands) for some population, or their ages, or the number of highway accidents in various regions or over a period of time, or the number of people with aids disease in various cities, etc.

Let us summarize this data into categories or classes as shown in Table 2.

Table 2: Summarization of Raw Data into Intervals and Frequencies (Frequency Distribution)

No. i	Class Interval	Midpoint m_i	Frequency f
1	11-20	15	1
2	21-30	25	3
3	31-40	35	5
4	41-50	45	7
5	51-60	55	8
6	61-70	65	7
7	71-80	75	5
8	81-90	85	3
9	91-100	95	1

This summarized data, or the frequency distribution of raw data values, appears to be more meaningful in form. It shows that most values fall within the range 31 to 80, the highest occurring values are around 51-60, with very few values at the low or high end.

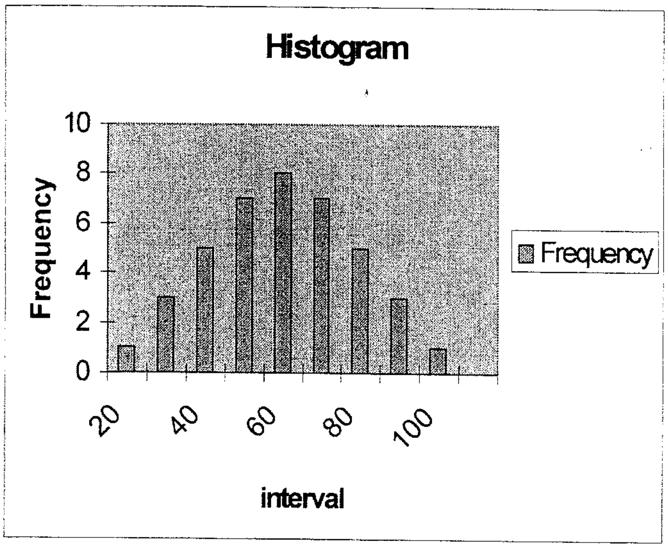
The frequency distribution is often displayed pictorially as shown in Figure 4.

The frequency distribution graph in Figure 4 shows the pattern of distribution for data values in a more noticeable form. This point has also been discussed in the preceding section. In Figure 4, we see that most of the values are clustered at the middle, and that the midpoint value, in this case, is in the interval 51-60 (despite any apparent misalignment of the horizontal scale in the figure as drawn). The value at this central point is the midpoint of 51-60 i.e. 55. This is called the **mean** value. What is the significance of this mean value? How can we use it to think and articulate about the characteristics of the measured scores for the sample and the general population? We will deal with these questions in the next section. For now, we may make the following observations. If the frequency data plot takes the, form of a sharp bell curve then the sample mean value is a very good estimator of population mean value. If the bell is stretched out then the mean value is likely to change considerably from sample to sample and thus the mean value by itself cannot be a very good indicator of the population mean. If the bell is not symmetrical on its sides then the sample median may become a better indicator of the population mean. We say that the bell curve is negatively skewed if it is slow in rising from the start to its peak and positively skewed if it falls slowly down from the peak to its right.

As mentioned earlier, the shape of frequency distribution for many phenomena takes the form of a bell curve i.e. a frequency distribution with a peak in the middle and tapering off on both sides symmetrically. This is called a normal distribution.

The question is how normal is a situation that looks quite normal? We may visualize the top points of the bars in Figure 4, when connected together, produce a good bell shape, although in this case it appears to be a little stretched out. We will also examine how the shape of this bell may influences us in making abstractions or generalizations from the given values of data.

Figure 4: Frequency Distribution Graph for Data in Table 1 and 2



The bibliographic references on statistics as well as the Qur'an and Sunnah provide very useful material on discerning patterns, and reflecting on patterns to develop insights on situations that may appear to be different on the surface but largely rest on the same basic foundations. This is discussed elsewhere in the work related to this paper[Ahmad et al, 1997].

Returning now to the data in Table 2, we may notice that the reduced form of data lacks precision seemingly contained in the original data as listed in Table 1. However, in most practical cases this loss of precision does not affect resulting statistics in any significant way. The question is whether not expressing something precisely is undesirable. The answer in many cases is no. First, the precision used in recording measurements may itself be misleading. For data in Table 1, when a number such as 67 is recorded, how sure are we that it could not be 65 or 69 or something like it. For this particular set of score values, there may be have been limitations which make the score of 67 no more a true reflection of students performance than say 65 or 69. We may be on much safer ground if we make the score somewhat fuzzy, indicative of a range of values rather a single value. This possibly what happens when we assign letter grades to numeric scores recorded for students. Let us say that we are recording temperature values in a particular place in our home. The measuring instrument may not be very precise so that any value we record is suspect but we may be more certain if we record a range given the known limitations of the measuring instrument used. The problem is further compounded by that fact that the same type different instances of measuring instrument may not record exactly the same value. These remarks aptly to a student's work being graded by different instructors. It is highly unlikely that different instructor, even teaching the same material to different sections of a course would give the same numeric score for a given piece of work.

Making fuzzy statements about situations may often be more accurate and meaningful than precise statements. Fuzzy values and how to draw inferences from fuzzy values is an area that we will take up elsewhere.

An equally important consideration is the issue of judicial use of the resources. Let us say that we are given a thousand items, and somehow we are able to select a representative sample of say 20 or fewer items. If we are able to work with this small sample to draw whatever inferences we wished as well as we could do one thousand, then working with a few makes both efficient and effective. Efficient because less resources and energy will be used in working with a small sample. Effective because it is a lot easier to grasp and articulate with a small number rather than a large number of items. Small may turn out to be quite beautiful in many situations.

4. MEASURES OF CENTRAL TENDENCY:

We begin by raising issues and asking questions about phenomena in nature - things that are not necessarily man made although man may have contributed positively or negatively to the phenomena. For our purpose, phenomena are things that we observe in time and space e.g. the amount of rainfall, people dying of accidents, incidence of certain diseases, the performance of students in classes, and so on. A thoughtful question to ask whether there is an observable pattern or perceived nature of things in the phenomena. Why should we look for patterns or perceived nature of things in phenomena? From the thinking perspective, it is simply to gain insight into what happens. For example, it may be simply comforting to know the pattern that rainfall follows in a given region as we move in time through the year, or to know the pattern of rainfall as we move in space from region to region. Knowledge brings comfort. Knowing what may happen next, in time or space, is at least comforting even if we are not able to do much about it. Knowledge is also a potential source of power. We can use the knowledge to be in the right place at the right time, improve the quality of life or make financial gains or avoid financial losses. In addition to discovering useful features of phenomena and making use of these features in time of need, we may also examine whether certain phenomena show a cause and effect relationship.

Once a pattern is discovered, it can be improved further with repeated observations, and used as a basis to make inferences about new observations.

Experience has shown that the observed values of features in many phenomena have a certain tendency i.e. the values tend to cluster around what may be called a mid-point and gradually disperse away from this mid-point in decreasing number. This is called the **central tendency**. Of course, we assume that what we are observing is random i.e. it is based on the nature of things that happen on their own rather driven by some systematic force aligning the values in its favor. The average of all observed values, also called the **mean**, is a measure of central tendency in that it identifies a mid-point. However, without knowing how the values disperse away from this mid-point, the mean value itself may not serve much useful purpose. The dispersion or the variability of the values is calculated and named as **variance**. Mean and variance are common measures of central tendency and variability in observed values. These are statistics of common choice when the observed values behave as discussed in this paragraph, and the sample of values is randomly drawn so as to make it representative of the population at large.

Let the random sample of data values be denoted as x_1, x_2, x_3, ..., x_n for n values of data. Then the mean value is calculated as:

mean = $(x_1 + x_2 + x_3 + \ldots\ldots\ldots\ldots x_n)/n$ (8)

We may also denote mean as

mean = $1/n\sum x_i$ for I=1,2,3, ..., n (9)

Applying equation (1) or (2) to the data in Table 1, the calculated value is:

mean = 55.723, or

mean = 55.73 (approximate but a better choice) (10)

Equations (1) and (2) can be easily adopted for the summarized (frequency distribution) data in Table 2, as described below.

mean $=(f_1x_1 + f_2x_2 + f_3x_3 + \ldots f_kx_k)/(f_1 + f_2 + f_3 + \ldots (f_k)$
$= (f_1x_1 + f_2x_2 + f_3x_3 + \ldots f_kx_k)/n$ (11)

or

mean = $1/n\sum f_j x_j$ for j=1,2,3, ...,k (12)

Using (11) or (12), the calculated value is:

mean = 55 (13)

Note that the value in (13) differs only marginally from the value in (10). Generally, classifying raw data into frequency distribution form has only marginal effect on the calculated statistics.

What is important here is not to simply calculate the mean but to determine how usefully it characterizes the sample values and the population that the sample represents. Computational aids for statistics are within easy reach via a calculator or through software packages in Microsoft Windows environment. However, the relative ease with which these statistics can be derived should not lead to their thoughtless and misleading usage.

For the data in Table 1, the calculated mean in (10) is 55.73. Let us simply use mean=55 in our discussion. How well does this mean value characterize data shown in Table 1. Can we use this mean value to say that majority of students are scoring or near 55? Of course, we can answer this question by using the methods discussed in the previous section. Here, we would like to draw inferences by using the means and the related statistic of variance. In addition to using the mean in characterizing the sample, we are interested in finding whether we can do some generalizations. After all, the purpose of statistics is not simply to calculate a statistic. It is more to see if we can draw broader inferences from what we have calculated. In addition to asking whether a sample mean characterizes the population from which the sample is drawn, we should also ask whether the mean we have calculated would repeat its value or very nearly repeat its value if we recalculated it by taking many more samples.

We know that for data on the entire population, the calculated mean value is precise. For sample data from the population, the calculated mean is only an estimate of the population mean. It is a measure of the population mean if the sample is random in character. A sample is called random if every member of the population has an equal chance of being included in the sample and each selection of data is made independently of all others.

How good is a sample mean in terms of the likelihood that most values encountered in the population would be in proximity of this mean? In order to find an answer, we develop a measure of variations from the mean value. Let us start by determining how sample values deviate from the calculated mean. We calculate the difference of each sample value from the mean. The differences or deviations will take both positive and negative values depending on whether a sample value is larger or smaller than the calculated mean. If we sum up the differences, the positive values would cancel the negative values producing a result that does not represent the accumulated difference from the mean but something quite different altogether. The accumulated difference can be found by squaring each difference from the mean, summing them up, taking the average, and finally taking the square root to have a value that can be compared to the mean. This is called the standard difference or deviation. Squaring the differences and summing them up is easily understood intuitively. The mean value of sample variance is obtained by dividing the accumulated difference with the number of values, n, in the sample. However, the mean value of the variance for the population is obtained by dividing it with n-1. Why use n-1 and not n? We may view it as compensating the answer for using sample data as opposed to data for the entire population. Given n sample data values, n-1 denotes the degrees of freedom. The degrees of freedom are derived by taking the total number values in the sample and reducing it by an amount equal to the number of restrictions placed on calculations from sample data values. For calculating the mean value of variance, the degrees of freedom is one less than the number of calculated differences, assuming that once we have found n-1 differences, the nth difference cannot be freely calculated. It is fixed for a population. The calculation of standard deviation may now be described as follows:

Variance (of differences from the mean),
$$var = 1/(n-1)\sum(x_i - mean)^2 \quad for\ I=1,2,3, \ldots, n \qquad (14)$$

The variance is often calculated more easily and with lesser potential impact of round-off approximations in calculations as:

varx =(sumsq-sum*sum/n)/(n-1) (15)

where

sumsq is the sum of the squares of given data values,
sum is the sum of given data values, and
n is the number of data values

Equations (14) and (15) can be easily modified for data represented in the form of frequency distribution.

Standard deviation is then calculated by taking the square root of variance i.e.

stdev = sqrt (var) (16)

The calculated value of standard deviation for the above data is 19.28. Standard deviation is an indicator of how much the sample values may vary from the mean. How do we use this value of standard deviation in speaking about the mean value we have calculated? Can we say that the mean values we may find when we take more and more samples from the population would be near the one we have calculated? In other words, would the mean value remain stable from sample to sample? Can we come up with a statement of confidence about the calculated mean value versus the mean values from other samples? In order to answer these questions, we introduce a measure of variation for the sample mean. It is called **standard error of the mean** and calculated as follows:

standard error of the mean,
stderr = standard deviation/sqrt(n) (17)

A small value for standard error of the mean implies that the calculated mean will not differ much from means calculated with other samples from the given population. This is a significant advance in that we can generalize the mean calculated from the given sample to mean values from other samples drawn from the same population. If we assume a very large, almost infinite, number of such samples and their mean values being near the same as the one just calculated then this calculated mean may in fact be claimed as very nearly the true mean of the entire population.

The calculated value for the standard error for the sample in Table 1 is 3.05 using the formula in (17) and the calculated value of standard deviation equal to 19.28.

How do we use this value to locate the population mean? We may never succeed in finding the actual value of population mean because it may be impossible to take a measure of every single item in a population. What we can do is to determine the limits within which the population mean resides and to state it with certain degree of confidence.

For the sample in Table 1, the calculated sample mean is 55. Suppose that the population mean is 54. How does it lead to defining limits and confidence for the calculated sample mean of 55? This question can be transformed to the following question. For population mean of 54 and the sample mean is 55, how many of the sample means, for samples like the one in Table 1, would be as high as 55? Conversely, for population mean of 56, how many of the sample means, for samples like the one in Table 1, would be as low as 55? These are questions of probability. We would like to determine the probability that the population mean lies within the limits of mean values we are selecting. Rather than taking some arbitrary limits for the mean value and determining the confidence level or probability of the population mean lying within these limits, it may be more meaningful if we select some desired confidence level and then determine the corresponding limits for the mean. We may also examine the impact on the limits if we lower or raise the confidence level. For example, what are the limits for desired confidence level of 95% for the calculated mean of 55 for the sample shown in Table 1? For a population characterized by normal distribution, the confidence levels can be determined by mathematically defined formulas. For ease of reference, these are readily available in a tabulated form. Within the context of sampling distribution, we can use a corresponding distribution known as the Student's t distribution. This distribution allows us to define confidence levels even when the sample size is not too large. For the sample in Table, the number of observations is 40 and the degrees of freedom equal 39, as explained in the preceding paragraphs. Lookin at a table of t-distribution, we locate the entry for 3 degrees of freedom and 95% confidence level. It i 2.02.

We now calculate a value that sets limits for the mea at a given confidence level. Let us call it lmean calculating as follows:

$$\begin{aligned} lmean &= t*stderr \\ &= 2.02*3.05 \\ &= 6 \text{ (approximately)} \end{aligned} \quad (18)$$

The limits on the mean are, therefore, plus and minu 6. Applying it to the mean value of 55, we get th limits of 49 to 61. We can now state with a confidenc level of 95% that the mean value of the population, t which the sample shown in Table 1 belongs, will li between 49 and 61.

The sample size for data in Table 1 is 40. Generally this may be considered a large sample size. We may therefore, apply the normal probability distributio function in determining limits for the mean. The z value, corresponding to the above t-value, for a norma probability distribution is 1.96. The resulting limits o the mean are 1.96*3.05=5.98, very nearly 6 agair Thus for large sample size, we may simply use th table of normal probability distribution values.

In order to facilitate applying the method describe informally in expression (12) to (18), we may use th following formulation:

Symbols	Associated Meaning
n	sample size,
s	sample standard deviation,
c	range of values for the mea at a given level of confidence,
α	probability that the populatio mean is outside the specifie range, and
x'	sample mean

Therefore,

$$x' - t_{(\alpha/2,\, n-1)} * s / sqrt(n) >= c <= x' - t_{(\alpha/2,\, n-1)} * s / sqrt(n)$$

(19)

Referring to t-table for α=.05, we get

x' –2.02* 19.28/sqrt(39) >= c <= 2.02* 19.28/sqrt(39)

x' – 6 >= c <= x' + 6

49 >= c <= 61 (20)

In t-table, $\alpha/2$ represents the probability distribution value at one tail end. The total area under the two tails represents the probability value of $2 * \alpha/2 = \alpha$. The probability value for confidence level that the range is not reaching these tail end is $1 - \alpha$. With α=.05, $1 - \alpha$ = .95 or 95% in terms of percentile units.

What will happen if we wish to raise or lower the confidence level. Here are some examples following the calculation described in (18)

Confidence level	t value	lmean	Limits on the mean value
99%	2.7	8	47 to 62
98%	2.42	7	48 to 62
90%	1.68	5	50 to 60
80%	1.3	4	51 to 59
60%	0.85	3	52 to 58

(21)

It will be easily seen from the above values of confidence levels and corresponding limits that our claim for population mean being close the sample mean can only be stated with marginal confidence.

It may all sound very convincing but there may still be some questions about the basis for all this. We will make some final remarks about it. Consider that we are considering here statistics about phenomena that follow normal pattern of frequencies with which observed values occur in population. Placing the mean value of the population in the middle of a horizontal line, and plotting the frequencies of occurrences of values at the mean and away from the mean, we naturally get a bell shape curve. The area under this curve represents the entire population. Now if we take numerous, almost infinite, number of randomly selected samples of data, calculate all the means, and plot these means as a frequency distribution, we should get the same bell shaped curve from the samples as for the entire population. Given a sample and its frequency distribution curve, we try to determine the extent to which population distribution may fail to envelop the distribution corresponding to the sample.

The essence of preceding discussion is that statistics is not merely about calculating and characterizing. It is more about drawing inferences and questioning what we infer to see how well it applies to a given situation. Statistics is about understanding what the world may tell us based on statistics and to question what we are told because the statistics presented may or may support the claims being made. Statistics is about discovering patters and not fabricating patterns where none exist. Statistics is about quantifying what cannot be easily handled qualitatively based on certain natural human limitations.

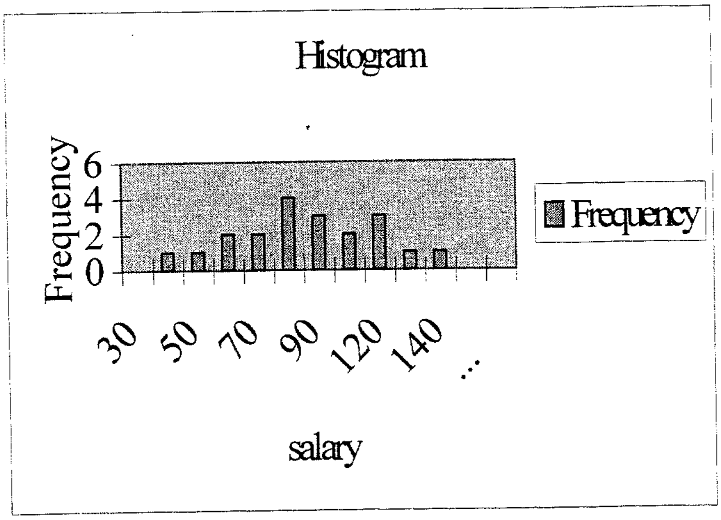
We started this section by taking a fictional sample of student scores in some class. When plotted in the form of a frequency distribution curve, it gave us a fairly good appearance of a normal distribution. Should this good-looking pattern be extended to student scores in general. Many teachers and even schools think that this is in fact the right thing to do. This is called grading on the curve. Grading on the curve, however, does not discover or characterize the actual performance of students in class but simply customizes, or shall we say homogenizes, the performance to the some expected standard. This hardly seems natural. Naturalizing something that may not in fact be natural hides discovering the phenomena causing unexpected patterns. This is not necessarily an honorable goal for use of statistics.

Before we leave this section, it is important to note that not all phenomena are characterized by well balanced normal distribution or normal distribution at all. We have focused on normal distribution because it may occur more frequently in nature and also because there are more possibilities for drawing and characterizing inferences on phenomena that follow normal distribution.

5. ASSESSING SAMP[LE VALUES ON THE BASIS OF SAMPLE STATISTICS:

Given a sample of values, how do we assess a specific value with respect to sample distribution. For example, given the sample of student scores in Table 1, and its mean value of 55, how do we rate someone who scores 65? How good it is with respect to rest of the students? More importantly, how do we rate the student's performance relative to other students in more than one class? Can we simply use the mean value and difference from the mean to compare performance? The problem may be further compounded if the maximum scores were quite different in the two cases, coming up with mean values having a meaning within the context of measuring scale. We can generalize this situation by saying that the two cases under consideration may have a different point of reference and unit of measurement for recorded values. We, therefore, need some kind of a 'standardizing' process to compare the two. We already know that any characterization using the mean is not very meaningful without taking into account the variations that the individual scores have with respect to the mean. This has been expressed in terms of variance, standard deviation and standard error for the mean value.

In order to make this discussion less abstract, let us consider a person, named Zia, who is considering making a move to another country. Zia salary in country A, where is residing now, 62,000. The mean value of salary for a sample in country A is 56,000, with a standard deviation of 12,300. Zia has been offered a salary of 99,000 in country B where is considering to move. The corresponding values of mean and standard deviation are 84,000 and 28,600, respectively. We know that the currencies in the two countries may not be same and that there cost of living might be different. Zia associates the quality of life he may have by comparing his own salary with the rest of the population.



Here is a superficial comparison by looking at Zia's salary and the corresponding means.

	Country A	*Country B*
Zia's Salary	62,000	99,000
Mean Salary	56,000	84,000
Difference from mean	6,000	15,000

(22)

The difference from the mean in Zia's salary in country B is two and a half times that of country A, whereas the mean value of salary for country B is only one and a half times that of country A. Should Zia expect a better quality of life in country B with respect to rest of the people?. Inferring this would be both superficial and misleading. In the above comparison, we are neither taking account the points of reference for salary values nor their unit or measurement. In order to standardize the statistics to a common point of reference and unit of measurement, we take into account the variance and standard deviation for sample values.

We convert the value we are examining to common standard as follows:

$$z = d / \sigma \qquad (23)$$

Where

z is the standardized value, called z-score,
d is the deviation from the mean for the value being evaluated, and
σ is the standard deviation of the distribution where d occurs.

Substituting the values in (23) for country A and B in Zia'a case, we arrive at the following standardized scores:

	Country A	Country B
z	.49	.53

(24)

This is approximately .5 for either case. It would, therefore, be wrong for Zia to assume that he would have a better quality of life looking at his salary versus the rest of the people. We may interpret the results in more precise statistical terms by saying that in either case, whether country A or B, Zia's salary is only .5 of standard deviation higher than the mean i.e statistically speaking there is no difference.

Figure 5: Frequency distribution of salaries in Country A

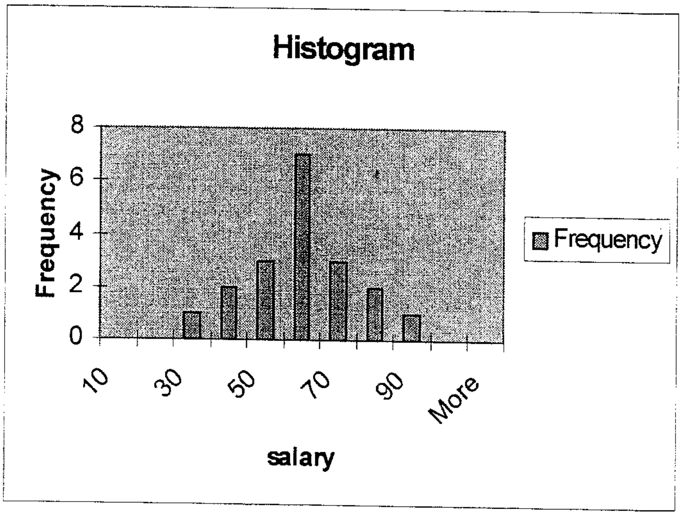
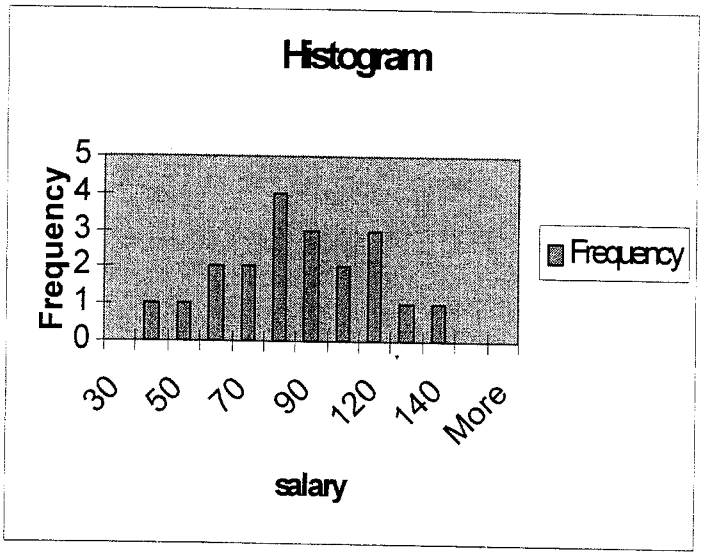


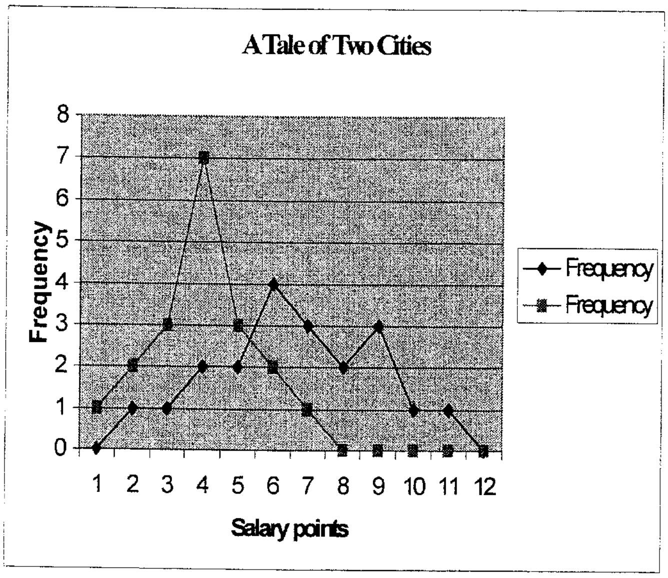
Figure 6: Frequency distribution of salaries in country B



A clearer view of the above situation may be obtained by looking at the graphs of frequency distributions, as shown in Figures 5, 6 and 7. The salary values are expressed in thousands. The title shown for Figure 7 is 'A Tale of Two Cities' whereas Zia was looking at two countries. Since a person living in a country actually resides in a city, we have chosen the title as shown, and it sounds a little more appealing.

We see that the salary values, shown in the form of frequency distribution in Figure 5 are more clustered around the middle or the mean, whereas in Figure 6 they more spread out. This comparative assessment becomes more apparent in Figure 7.

Figure 7: Comparison of frequency distributions of salaries in country A and B



Of course it would be contrary to creative thinking to just take a single measure for something as complex as the quality in making a decision where to live. There are likely to be many other, more important, factors affecting the quality of life. Just because we have a scientific measure of something does not make it a superior measure in making decisions.

6. CONCLUSIONS:

The preceding paragraphs provide a brief introduction to the subject of statistics. The introduction does not cover many other useful methods of statistical analysis, and what it presents is not necessarily fully explained. The purpose is to simply provide a motivation for the study of statistics. The readers are encouraged to pursue the subject further regardless of their field of study. Statistics are being put to good use in almost all areas of human endeavor. Without meaningful exposure to statistics, great many statements based on statistics may simply be baffling or understood incorrectly.

For further insights and access to methods of statistics, the following readings are recommended:

Canavos, G.C (1984). Applied Probability and Statistical Methods. Toronto: Little, Brown & Company (Canada) Limited.

Freund, J.E. (1979). Modern Elementary Statistics, 5e. Englewood Cliffs: Prentice-Hall.

Phillips, J.L., Jr. (1996). How to think About Statistics, 5e. New York: W.H. Freeman and Company.

7. Cited References

Ahmad, S.I. (1996). Creative Thinking and Problem Solving, Course Pack, First Draft. Kuala Lumpur: International Islamic University Malaysia.

Billard, L. (1997). A Voyage of Discovery. Journal of American Statistical Association. 92(437), 1-12.

Ketterning, J.R. (1997). Shaping Statistics for Success in the 21st Century. Journal of American Statistical Association. 92(440), 1229-1234.