Introduction

Statistics is the crucial process behind how we make discoveries, make predictions, and make decisions, all based on data. Proper statistical analysis assists by providing the information that companies require to make more accurate decisions driving productivity, quality, and ultimately sales and profits. Since such decisions are backed up by data, they tend to be much more reliable compared to decisions made based on opinions. Analysis reduces the risk of making wrong decisions that can lead to company losses.

In my experience with numerous students and business employees over the years, the words "*Statistics-101*" have instilled fear in many hearts. "*Statistics Anxiety*" has been defined as the feeling of anxiety when needing to take a statistics course (ref. Cruise, Cash, & Bolton, <u>Psychology Teaching Review</u>, Vol. 24, 2, 2018). If you are not planning to become a statistical theorist or statistician, fear not, an elementary short course in Applied Statistics may be all that is needed.

Mathematical Statistics vs. Applied Statistics

The discipline of statistics relies on probability and decision theory. Two key subsets of the discipline of statistics are *Mathematical Statistics* and *Applied Statistics*. Mathematical Statistics courses are concerned with statistical issues and problems, while Applied Statistics courses address the use of statistics for solving problems encountered in fields such as basic research, engineering, manufacturing, and medicine.

Mathematical Statistics courses involve studying the origin and derivation of formulae for statistical theories, the information provided by the formulae, and the improvement of statistical procedures using probability and decision theory. Applied Statistics, on the other hand, provides course material necessary for those interested in applying statistical tools for the purpose of data analysis. For Applied Statistics, interest does not lie in researching statistical theory, but rather understanding its applications. Mathematical Statistics is concerned with statistical formulae, while Applied Statistics is about using statistics for collecting appropriate data and problem solving.

Applied Statistics

Applied Statistics is the science of conducting studies to collect, organize, summarize, analyze, interpret, present, and draw conclusions from data. The statistical methods of data analysis facilitate making inferences about a population by analyzing a small sample chosen from it. Statistics is a crucial process behind how we make discoveries, make decisions based on data, and make predictions in the face of uncertainty. Statistical knowledge helps to use the proper methods to collect the data, employ the correct analyses, effectively interpret, and present the results. Industry professionals use statistical models to gain understanding and control of many aspects of business operations. Understanding the statistical methods that make up the various analytical tools used in manufacturing is key to correctly applying, running, and analyzing the results of these statistical tools.

Components of Applied Statistics

The *Population* is the complete collection of items being studied. A *Parameter* describes the population's characteristics. A *Sample* is the subset of a population and contains items on which the measurements are to be taken. A *Statistic* is a number that describes a sample's characteristics. Statistical methods of data analysis make inferences regarding the population by studying small samples selected from the population.



Descriptive Statistics & Statistical Inference

A basic understanding of the difference between Descriptive Statistics and Inferential Statistics is needed for the deployment of statistical analysis. Descriptive Statistics describes sets of data measurements that were actually made for analysis in the form of tables, charts, and graphs. The three main types of Descriptive Statistics are the Frequency Distribution, Central Tendency, and Variability of a Data Set. Inferential Statistics allows one to make predictions ("inferences") from the Descriptive (sampling) data taken. With Inferential Statistics, one uses data from samples that were measured, and draws conclusions about a population.

Sampling

Establishing the correct sampling parameters is crucial in statistics. Improper sampling will lead to incorrect results, and even worse, reaching misleading conclusions. Random Sampling is a part of the sampling technique in which each sample has an equal probability of being chosen.

A sample that is chosen randomly is meant to be an unbiased representation of the total population. An unbiased random sample is important for drawing conclusions. A sampling method is called "Biased," if it systematically favors some outcomes over others.

A Sample is defined as a percentage of the total population in statistics. One can use the data from a sample to make inferences about a population as a whole. The Ideal Sample size computed is a function of the Confidence Level and Margin of Error that one is willing to accept. Selecting a sample size is a challenging task that often requires concessions due to external forces such as costs and meeting delivery schedules. These constraints can greatly affect the number of samples that can be measured.

The Confidence Level is the probability that a result is significant. A high confidence level means that there is only a very small probability that the result (e.g., correlation or independence) happened purely by chance. Margin of Error is how many percentage points your results will differ from the real population value (e.g., a 95% confidence interval with a 4 percent Margin of

Error means that your statistic will be within 4 percentage points of the real population value 95% of the time).

Classifying Random Variables

Before proceeding, one must be aware of the difference between Quantitative and Qualitative Random Variables.

Quantitative (Numerical) Random Variables are data that represent counts or measurements. They are numerical in nature and can be ordered or ranked. This data can be characterized as Discrete or Continuous Variables. Discrete Variables are data values that can be counted and are finite. For example, measurements such as counting parts and the number of employees absent from work in a given day must be a discrete number like 2, (i.e., the data point cannot be 2.3). Continuous Variables, however, take on all the values in some interval of numbers. For example, weight, age, salary, height, temperature, time to complete a task, etc.

Qualitative (Attributes) Random Variables are data that refers only to a name classification. They can be placed into distinct categories according to some characteristic or attribute. There is Nominal Data that cannot be ranked, e.g., gender, race, citizenship, Zip code, etc., and there is Ordinal Data that can be ranked, e.g., school grades (A,B,C,D,F)

Summarizing and Analyzing the Data

Once you understand what kind of data you are working with and the type of information that you are seeking, the task becomes that of selecting the appropriate statistical tools (formulae) to apply to analyze your data that will lead to resolving your problem. Firstly, you should determine whether your data is normally distributed or not, since your tool selection will depend on this. A Histogram or Box Plot are good choices to determine this. When data is not normally distributed, the cause for non-normality should be determined and appropriate remedial actions should be taken.

Histograms		
Probability Dist	tributions	
Standard Devia	tion	
Variance		
Central Tenden	cy Data Ar	nalysis:
Mean	Percentiles	
Median	Quartiles	
Mode	IQR	
Range	Outliers	
Exploratory Da	ta <mark>Analy</mark> sis	
Pareto Diagra	m	Stem & Leaf Plots
Scatter Plots		Radar Charts
Correlation Co Box Plots	efficient	SPC Charts

As shown in the Table, there are several statistical

analysis tools that can be easily learned and applied to the data collected. There are also more advanced tools such as Hypothesis Testing, Chi-Square Goodness of Fit, ANOVA, and Design of Experiments (DOE).

In Conclusion

Statistics helps in providing understanding various phenomenon of nature. Statistics does this by the proper and efficient planning and execution of a statistical inquiry in various fields of study. Knowledge of Applied Statistics provides for the use of proper methods to select and collect the data, employ the correct analyses, effectively interpret the results, and make decisions to take the

best course of action. It can also deal with uncertainties by forecasting based on current data, helping with sound decision making by providing accurate estimates on the basis of sound predictions and assumptions.

The applications of statistics are many and varied. For example, statistical analysis is crucial in engineering and manufacturing disciplines, and can result in significant quality improvement, productivity enhancement, and cost reduction.

I sincerely hope that this brief article will not only put you at ease to learn more regarding the uses of Applied Statistics, but will encourage you to take a short course that addresses the applied statistics techniques discussed in this article. Doing this should provide any business employee with the ability to be able to understand and execute to a statistical plan of collecting, organizing, analyzing, summarizing, and presenting data to drive implementation of the statistical analysis performed.

INTERESTED IN LEARNING MORE?

Please contact me if you are interested in learning more about a reasonably-priced four-hour, online Applied Statistics training class. This class is available as synchronous event where the instructor is connected "live" with the class participants via the Internet. Being in direct contact with each other, both the instructor and the class members have the opportunity for interactive learning which enables the asking and answering questions in real-time.

Dr. Al Pucino Principal Instructor & Consultant essentialMFGconcepts Web Site: essentiaMFGconcepts.com email: apucino.emc@gmail.com