

Tour to the Statistician's Workshop

Mulugeta Gebreselassie*

A king going to war wishes to know what reserves of man power and resources he can call upon. How many men need be put in the field to defeat the enemy? How much food could be enough? and what is the cost of all the resources needed? etc. We are also reminded about ancient statistician's every Christmas when we read the Bible about the decree of Caesar Augustus that the whole world should be enrolled. Had it not been for the lousiness of the statisticians, to crowd the enrollees in that little village, Christ would have been born in the modest comfort of a cottage in Nazareth. Such events were the earliest problems to apply statistics to. At this point it is wise to ask one self what statistics and what a statistician is? What tools a statistician has in his workshop?

It could be very difficult to give an exact definition of a field which is outbound to grow like Statistics. The definition given by R.A.Fisher, however, can be considered as a standard one. Fisher defined statistics as a science of decision making applying mathematics to observational data. When more elaborated, it is a science which deals with the techniques of data collection, data organization, data analysis and interpretation. That is, techniques which constitute the main tools that are found in a statistician's workshop. Similarly, the name statistician is given to a person who does, apply and use statistics. But note the fact that, it is not the intention of the writer here to explain what statistics is? and what a statistician is? and bela bela.... Rather the intention is to familiarize the reader with the main tools that can be visited in a statistician's workshop. Not only visiting, the tour could also serve another purpose. To know the purpose of each tool and on how to operate the tools in the workshop.

Now a days the need of statistically sound results for decision making is at the peak. Most publications are decorated by statistical findings. You can boldly say that including these results, in any form, in one's publication has become more than a fashion. It is not amazing to see a person using scissors for drilling than the driller.

* Ass. Lect, Department of Statistics, Addis Ababa University

What do you think is wrong? To me it is only lack of a short tour to the statistician's workshop.

The first step in any statistical investigation is to make a clear statement of your objectives. That is to vividly put your hypothesis. After having this one needs to answer among others the questions, what type of data to be collected? What is the type of sampling technique to be employed? What optimum size of a sample is needed? What should the methodology of data collection be employed? What is the extent of clarity of the instrument in data collection? and what else?

The toxicity of failing to apply correct techniques at this level is very critical. But how? The answer is very easy. Just resort a reasonable amount of time and effort to visit a statistician's workshop. The shop is full of tools which ameliorate the toxicity. Tools of sampling, tools of developing a correct data collection instrument and ... lasting verification of objectives.

Once you developed a good instrument to collect data selected by a correct sampling technique for an optimum number of samples, the next stage is to feed your data into a computer. It is very advisable to check the data for possible errors that may occur during data entry, however. As the next stage is organization of the data into a manageable form, a prerequisite to generation of results which describe the situation under study. There are different techniques of data organization, tabulation, constructions of graphs and diagrams and etc. At this level it is wise to prefer a correct instrument of data organization. Not doing so could result in distorting out the reality. It can not be a problem to do so. These tools are abundant in the statistician's workshop.

It may not be difficult to guess the next step. After data organization, relevant information should be extracted from the organized data. At this juncture it suffices to mention one of the articles written by Mekdim in the 1992/95 graduation bulletin. The paragraph long article starts by urging the reader to become keen to learn statistical techniques. Especially to those who are used to apply statistical techniques, he advises them to abide by the statistical rules if they want to get amenable results. Indeed, he continues by saying that he had no light to lead him to a future bright in his earliest days of his second year, except to accept the misconception that statistics is a lie.

Thanks to his relentless effort he found the mean and variance to give him their hands and lead him to the seat of wisdom. Do not forget that these (the mean and variance) are the priceless tools in the statistician's workshop. But not only these, if you commit like Mekdim to never be partial and never treat figures for your own sake, the workshop is full of extraction instruments. The absolute and relative descriptive measures, the correlation and regression coefficients, and so on.

The whole story in the preceding paragraphs is for nothing but to make a sound decision based on a limited information about the population under study. That is about the hypothesis set at the beginning. In the statistician's workshop the biggest room full of vital devices is the room of statistical inference. I would doubt the existence of statistics had it not been for the concept behind a sample, which is key in the theory of statistical inference. Estimation (both point and interval) and hypothesis testing are its major aspects. Since this is the stage which decides the fate of your decisions, it is important to visit every device carefully.

You can imagine how big the risk of making an incorrect decision could be especially in medical science research which involves the human life itself. One of the most important tools, which did not get the minimum attention, in inference are assumptions. The reason could be different. But the author believes that the advent of statistical packages which can easily be used by laymen contributed to the aggravation of the problem. For instance most commercial statistical packages generally concentrate on testing hypothesis about various parameters with out providing a means for overall diagnostics of the model assumptions. Hence, it is worthy for the data analyst to advise him/her that it is wise to focus on the important assumptions which must be examined to ensure that the analysis is sound and appropriate. You do not have to go far to obtain these, they are the valuable tools of the statistician. It is just to give a tour to his/her workshop!