Training and Teaching Statistical Methods for Quality

Matthew Barsalou, Germany, email matthew.barsalou@gmail.com

Pedro Saraiva, Portugal, email pas@eq.uc.pt

The purpose of this white paper is to discuss how critical it is for organizations and societies to have many individuals with knowledge of basic data-driven approaches to quality and the ways currently available to conduct such teaching and training, which are even more crucial in the current era of big data, society 4.0, and quality 4.0. Both quality professionals and citizens in general will be much better prepared for making sound decisions if, and only if, they are able to understand and apply data-driven principles and approaches.

It is the seventh paper in a series of thoughts collected, organized, and promoted by the Quality in Education Think Tank (QiETT) of the International Academy for Quality (IAQ).

The first paper addressed a broader scope of topics and put into perspective the overall field of "Quality in Education,", which set a common ground for further reflection and guidance of QiETT activities. The forthcoming papers, such as this one, focus on more specific topics and delve deeper into particular topics based upon the collection of international inputs from quality and education experts.

To date, this collection of white papers comprises the following titles:

"Quality in Education: Perspectives from the QiETT of IAQ"
"Large Scale Training of Quality Professionals"
"Inclusive Quality of Education"
"Continuing Education in Quality Improvement for Healthcare Professionals and its effects on organizational improvement"
"Current Societal Challenges to Quality and Quality Management in Higher Education"

“Applying Quality Theory to Educational Systems”
“Training and Teaching Statistical Methods for Quality”

1. Introduction

Statistical methods are both a necessary and powerful additional component of the quality professional’s tool set as has pretty much always been the case. Consider the historical roots of the discipline of quality where many leading figures of the quality movement in the twentieth century had a strong statistical background (e.g. Walter Shewhart, W. Edwards Deming, or George E. P. Box), or the statistical nature of many quality tools and improvement methods, such as Statistical Process Control (SPC), Design of Experiments (DoE), or Six Sigma.

However, teaching statistical methods is not as simple as explaining how to create a flowchart. Statistical methods can be perceived as dry and uninteresting if not well taught and this can end up removing the joy of learning from the topic (Neumann, Hood, and Neumann 2013), as many of us may have experienced. But this does not need to be the case and this white paper presents approaches that can be used to make statistical methods more interesting to learn and to help teachers and learners with some alternative ways to present and introduce basic statistical concepts and methods. Therefore, this white paper pertains to both teaching statistical methods in an academic setting such as a university classroom and in industry under the scope of some more specific training contexts such as Lean Six Sigma.

2. Statistics and Examples

Examples are helpful in explaining statistical concepts and simple examples should be used for introducing topics. Examples must be simple enough for the learners to understand them and should focus on what really matters, which is learning statistical principles and tools from such examples. For instance, using an example that requires a detailed understanding of chemistry could result in some confusion due to a lack of understanding of the example details for people without a strong chemistry background. Furthermore, under such a setting, learners might lose the main focus of learning statistics by spending too much of their time and focusing too much of their attention on the chemistry related to the example as opposed to the data-driven knowledge that could be acquired from the corresponding example’s data set. Therefore, whenever that is possible, simple but realistic examples are best for introducing statistical concepts and approaches.

It is also essential for the examples used to be the ones that can be easily understood by the learners. One university statistics exam provided the time of three competing skiers and asked what type of data it was. Time is ratio data as it has both a zero and equidistant measure, with four minutes being twice as long as two minutes. However, the correct answer on the exam was ordinal data. When challenged, the professor explained that it should be obvious that the data pertained to who came in first, second, and third place because such data in skiing only pertains to placement of competitors. But what if not every student had been skiing before or knows about such skiing competition rules? This exam question required completely unrelated outside knowledge and the correct answer was therefore not unambiguous.

The value in graphing data prior to analysis can't be mentioned enough as shown by simple univariate control charts for close to a century. In addition to explaining why graphing is essential, an example should be provided to illustrate the necessity to graph data prior to any analysis as data visualization is indeed very powerful and insightful. The value of this concept can be shown by explaining a scenario under which data is collected. For example, you can provide the learners with measurement data for the height in meters of a sample of 27 people and ask them to calculate the mean. The learners should create a graph of the results for this set of data, but let´s make one value in the sample be 0.52 meters. The mean may sound plausible, but a look at the corresponding graph of the data will quickly show that something is wrong due to the presence of the outlier. Perhaps there was a typo, the measuring device malfunctioned, or somebody failed to enter a 1 before the value. We can illustrate this point with examples of how things have gone wrong due to not looking at the data or have been prevented from going wrong by looking at the data.

Another example is a Lean Six Sigma Master Black Belt who was given thousands of values to use in a regression analysis. A quick look at a graph indicated a problem as one ambient temperature value was below absolute zero. Although such a discovery, if verified, maybe worthy of a Nobel Prize, the Lean Six Sigma Master Black Belt discovered that such a value is what happens if a temperature sensor is not connected. No Nobel Prize was awarded since negative temperatures have yet to be found and would contradict fundamental laws of thermodynamics!

Teaching statistics requires more than just explaining how to properly apply a formula. Yeh *et al*. provide a case study in which the researchers determined there is no statistically significant difference in the number of injuries and deaths reported by people who jumped from an aircraft with a parachute versus those who jumped from an aircraft without a parachute. The statistics show no difference and the math is correct, but the most important detail for understanding these results is not in the math. Rather, it is related to the research design and data collection. The aircrafts used in the study were parked and stationary (2018). The numbers alone don't tell the whole story if you do not understand the context under which they were collected, how good they are, and what goals one aims to achieve with them.

Presenting such examples can drive home the need to understand the context of the data that is statically analyzed. In addition, Bisgaard recommends teaching details such as why student's t-distribution approaches the normal distribution as sample size increases and what is the intuitive difference between a paired t-test and a t-test of two means (1989). The two t-tests have different data requirements and different formulas and most textbooks should explain this, but learners who are expected to apply the concept on the job should also understand why there are two different formulas.

Learners should also be provided with an opportunity to apply the concepts they are being taught and it is even better if they can bring their own date sets for doing so. A simple example that uses multiple statistical methods is a company with community coffee in a kitchen for all employees. Suppose one employee had the impression that the one liter milk container was often left with just a little bit of milk and this person suspected that somebody was initially avoiding using the last drop to avoid switching out milk containers. Let the learners offer suggestions on how this could be investigated statistically.

Groups of learners could be assigned the task of planning such an analysis. One solution would be to track the number of times the milk is almost empty versus the number of times the milk is not almost empty. A hypotheses test of proportions could be performed on these results. But first how often the milk could be expected to be almost empty by random chance must be determined. To do so the amount of servings of milk in one liter must be determined.

Ideally, the amount of milk used in a cup of coffee by each employee should be determined, but attempting to do so may result in being reported to human resources. The person performing the study could determine how much milk they use. Assuming the amount of coffee dispensed is consistent and the amount of milk poured into the coffee cup is consistent, a mark could be placed in the cup at the point the cup is filled to when containing both coffee and milk. The next time the cup is filled, the coffee should be poured into a measuring cup to determine how much coffee is used. The coffee would then be poured back into the coffee cup and milk would be added. Then the contents of the entire cup should be poured back into the measuring cup. The difference between the new result and the previous result is the amount of milk used.

Let us further assume that 0.075 liters of milk are used as one serving. That means there are 13.3 servings in a one liter container of milk. Therefore, the milk container would contain less than one serving after the thirteenth serving. The chance of using the milk when it needs replacing is thus 1 out of 14 times. The chance of milk needing to be replaced is 0.071, which is 7.1 percent of the time. Suppose two cups a day of coffee are consumed over a period of ten working days for a total of 20 cups of coffee with the milk needing to be replaced six times. This information could be used to perform a hypothesis of one proportion to compare the actual proportion of 0.3 to the expected proportion of 0.071.

This example can be worked by the learners if data is provided to them. In addition to providing practice with performing a practical and easy to follow hypothesis test, the example just described also demonstrates that there is far more to statistics than just performing a test. Careful planning was required to determine what to evaluate and how to evaluate it.

The example should be as close as possible to the learners' experiences whenever realistic examples are used. In industry, examples should be tailored to the type of product or process the learners are involved with. Unfortunately, this is not always possible such as when teaching university students or learners from multiple organizations and industries. In such situations, simple generic examples are needed as an alternative; alternatively, each student can be asked to perform and present individual analysis of their own data sets.

Real world case studies using actual data can be useful in teaching statistics (Viles and Martin 2008). Examples could come from projects in which the educator was involved in or out of case studies presented in the literature. Multiplying all values by a random variable is an option for educators who can't share potentially proprietary or confidential data; this results in the relative positions of the values will remain the same while avoiding the sharing of actual values.

Fortunately, there are many resources available such as free access online to many interesting real life examples together with the corresponding data sets that can be used both by teachers and learners.

3. Generating Data with a Random Distribution Generator

A random distribution generator can be used for generating both normally distributed data and data that follows other distributions and this data can be analyzed by the learners. The number of values used should be lower if the learners are doing calculations by hand, but learners using statistical software or even a spreadsheet can be given larger data sets.

Table 1 provides the mean, standard deviation, sample size, and descriptions of possible tasks to be conducted. The same basic idea can be used for generating many different data sets for practicing statistical analysis or to complement other available real-life data sets and examples. This approach should not be underestimated in terms of their learning potential, as stated above.

Table 1. Simulated examples that can be created using a random distribution generator for learning about the given statistical tasks and approaches.

Data set	**Mean**	**Standard deviation**	**Sample size**	**Use**
Bore diameter machine A	12.283	0.049	35	One sample student's t-test and one sample Z-test to compare the mean to a hypothesized mean of 12.15 and a Chi-square test to compare the standard deviation to a hypothesized standard deviation of 0.040 as well as constructing box plots and histograms
Bore diameter Machine B	12.4	0.044	35	Compare to bore diameter day 1 with a box plot, hypothesis test of two means, and compare standard deviations with an F-test as well as comparing the data sets in box plots and histograms
Rod outer diameter A	22.5	0.008	100	Capability study with capable data using a target value of 22.50 with a tolerance of +/- 0.04
Rod outer diameter B	22.5	0.016	100	Capability study with non-capable data using a target value of 22.50 with a tolerance of +/- 0.04
Shaft Length	44	0.05	100	Control charts using variable data

A large quantity of available data is useful for understanding and practicing various statistical methods. However, experience in planning experiments is also necessary and training aids can be useful for both generating data as well as letting the learners plan statistical studies. Once learners can apply basic concepts, they should move onto performing their own experiments and collect their own data sets, as argued in the previous Section.

4. Training Aids for Statistical Methods in Quality

Textbook examples often offer a simplified view of the world. This may be justified by pedagogical reasons; however, real world data can be messy, with typos slipping into the results or unexpected problems arriving during the study. Under real word conditions, much time and resources need to be allocated to the planning and collecting of data appropriate for knowledge extraction and decision-making.

For example, paper helicopters may be used for experimenting and this method become a classic way to conduct Design of Experiments (DoE) and has been promoted by George E.P. Box. But what happens if the helicopter hits a wall and crashes? Should the result be used or should the experiment be repeated? The proper use of training aids provides learners with an opportunity to navigate through such problems before they attempt an actual experiment in industry.

To be useful, training aids should help us to simplify concepts or keep discussions focused on the most relevant topics. Alternatively, training aids can also help to make topics memorable for learners or take them to places they could not otherwise go to, such as watching a video of a process in operation. Training aids are useful for creating variety or save time (Mitchel 1998).

An easy to prepare and low-cost option for letting learners collect and analyze their own data is to use poker chips or bingo chips with values written on them. For example, Tables 2 and 3 each provide 100 values that can be written on a poker or bingo chip. Alternatively, a random distribution generator can be used to create values which can be written on additional chips.

Table 2. Sample size 100 with mean 20.005 and standard deviation 0.0178 for a normal distribution.

20.023	19.988	20.020	20.001	19.997	20.007	20.023	20.009	19.991	19.986
19.977	20.035	19.995	19.968	19.993	19.987	20.000	19.985	19.982	20.030
20.021	19.995	20.023	20.017	19.993	20.012	19.988	20.008	20.016	19.959
20.012	19.982	20.020	19.977	19.989	19.986	19.992	20.024	20.029	20.017
20.015	19.990	20.018	19.992	19.990	20.011	20.036	20.033	20.009	20.001
19.984	19.992	19.988	20.028	20.019	19.993	20.032	19.976	20.034	20.037
20.019	19.996	20.004	19.974	20.029	19.994	20.011	19.995	20.025	19.989
20.011	19.983	20.021	19.988	20.009	20.000	19.986	20.006	20.004	20.004
20.009	19.995	20.017	20.021	20.014	20.030	20.025	19.997	20.001	20.024
20.001	20.040	20.003	20.003	19.999	19.980	20.015	20.003	19.982	20.029

Table 3. Sample size 100 with mean 20.008 and standard deviation 0.0415 for a normal distribution.

20.029	19.984	19.926	20.006	19.971	19.999	20.008	20.055	20.030	20.015
20.035	20.022	19.940	19.995	20.009	20.010	19.974	20.047	20.058	20.051
20.008	20.043	20.001	20.044	20.013	19.960	20.035	19.943	20.028	19.980
19.973	20.093	19.970	19.941	20.007	20.060	20.060	19.979	20.013	19.961
20.084	19.958	19.978	20.015	19.946	20.020	20.005	20.076	19.996	19.985
20.042	20.044	19.975	20.032	19.987	20.026	20.059	20.050	20.046	20.048
20.017	20.058	19.967	19.982	19.937	19.997	19.977	19.949	19.978	19.990
19.991	19.986	20.013	20.088	20.060	19.972	20.014	20.033	20.012	19.988
20.039	20.037	20.014	19.954	19.988	19.918	19.937	19.998	19.921	20.050
19.974	20.049	20.035	20.048	19.942	20.100	20.077	20.021	20.054	19.996

Many scenarios can be created based on these two sets of chips. For sample, the chips can be stored in two separate containers with Table 2 values written on one color and Table 3 values written on the other color. Learners can be tasked with taking a sample of a given size from one or both containers to complete tasks such as creating a histogram, comparing the means of the two samples, comparing the standard deviations of the two samples, or comparing the mean or standard deviation of one sample to a given value.

The chips can also be useful for quality related statistical methods in addition to basic statistical concepts. For example, you can inform the students that the chips are for a product with a specification of 20.000 mm with a tolerance of +/- 0.005 mm and ask them to perform a capability study on such data. The same chips can also be used for creating an individual value and moving range chart, Xbar and R chart, or a Xbar and S chart.

Alternatively, unmarked chips can be used for attribute charts by mixing a larger quantity of blue chips with a smaller quantity of red chips, with red chips representing defective parts. Or, a marker can be used to create one or more blemishes on selected chips with these chips representing a part having one or more defects.

A container with a large quantity of white balls and small quantity of red balls could be used to recreate Deming's famous red bead experiment. For example, 100 white balls and 20 red balls could be placed in a bin and mixed. A small scoop could then be used to take a sample from the bin. In addition to performing the red bead experiment, this could also be used for creating control charts for attribute data. One of the balls could also be used together with a funnel to perform Deming's funnel experiment (Deming 1994).

These training aids can be quite useful both for top managers to understand statistical thinking as well as for quality professionals to learn from a more technical point of view which statistical tools to use as well as how and when they should be applied.

There are many possible training aids available for teaching DoE (Santy and Einwalter 1997). A typical DoE training aid is the catapult, which can be purchased or constructed and Deane and Burges provide both training scenarios for a catapult and instructions for building a catapult (1998) and many variants of the catapult can be built and then used for learning purposes. Table 4 shows a parts list for assembling a homemade catapult. Heavy duty rubber bands are listed and these can be purchased from a model shop. The types used for model airplane propellers come in different sizes and are robust. Small balls are also listed, with the type known as "hacky sacks" being ideal as they roll less after landing and removing some stuffing material helps to ensure that they roll even less.

Table 4. Parts list for a catapult that can help to understand and learn about DoE.

Quantity	Material
2	2 Pieces of 2.4 x 3.8 x 26 cm wood
4	2.4 x 3.8 x 10 cm wood
2	2.4 x 3.8 x 11 cm wood
2	2.4 x 3.8 x 5 cm wood
10	Wire rod 20 cm long
16	Wood screws
4	Eye hooks
1	Small nail
1	Small plastic cup or bottom of plastic bottle
3	Heavy duty rubber bands of different diameters
3	Small balls of different weights

Figures 5 through 9 depict catapult assembly instructions. Figure 5 depicts a side view, Figure 6 depicts a front view, Figure 7 shows where holes need to be drilled, Figure 8 shows where a hook must be attached, and Figure 9 shows where additional hooks need to be attached as well as where the cup should be attached.

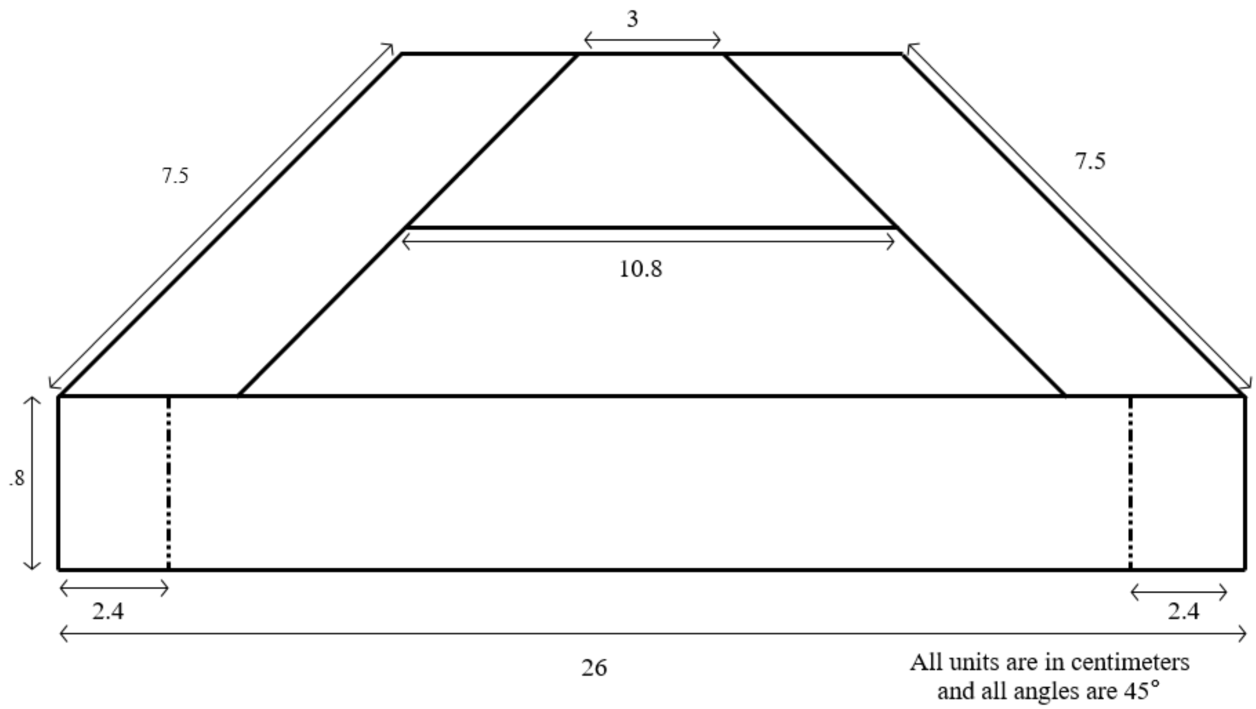

Figure 5. Side view of catapult assembly

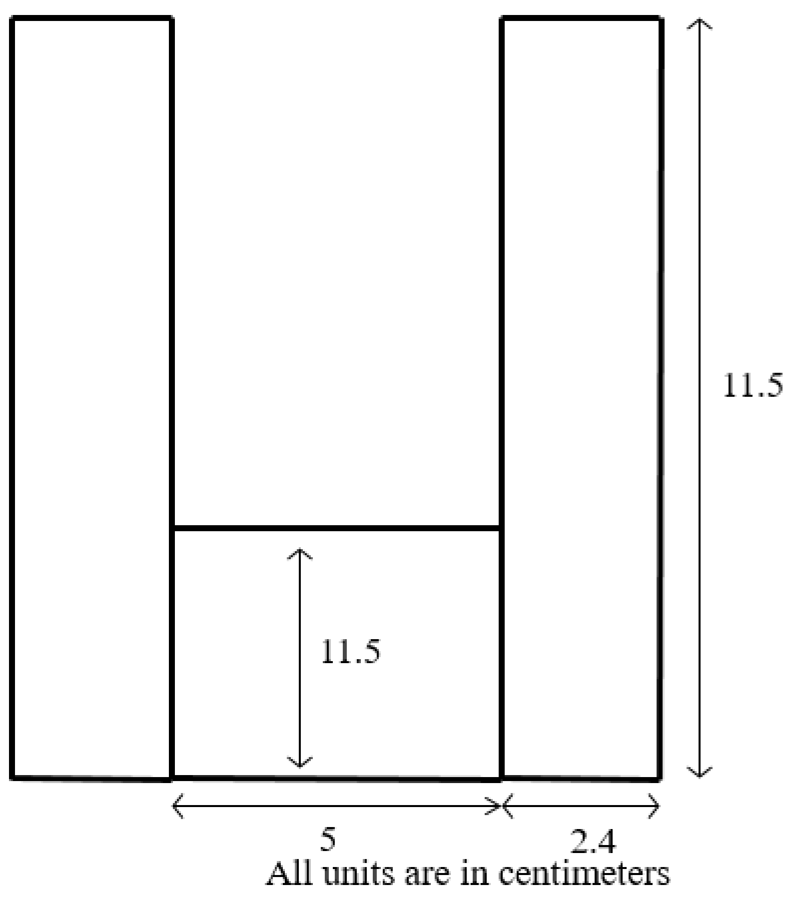

Figure 6. Front view of catapult assembly

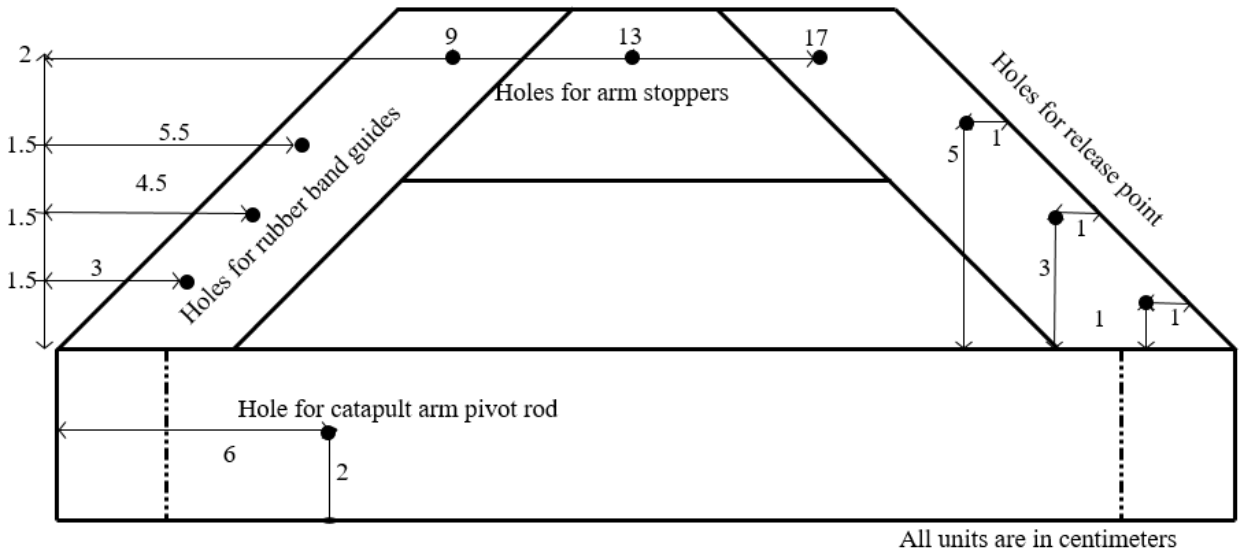

Figure 7. Side view of catapult drilling points.

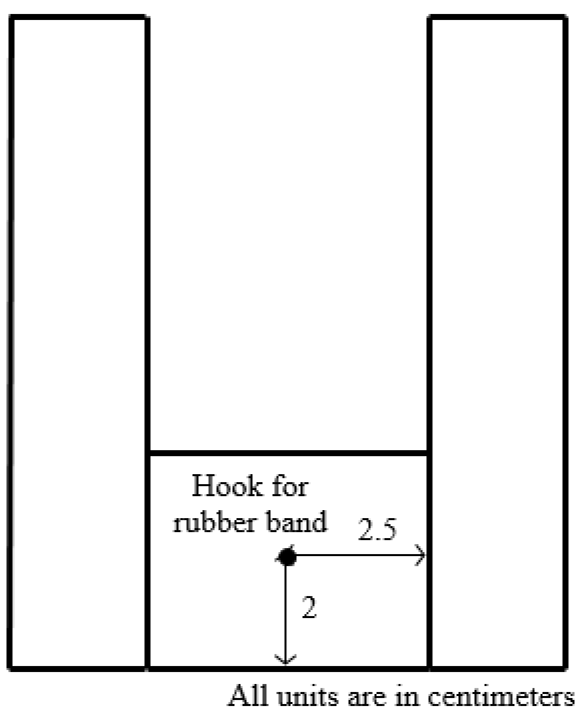

Figure 8. Front view of catapult hook attachment point.

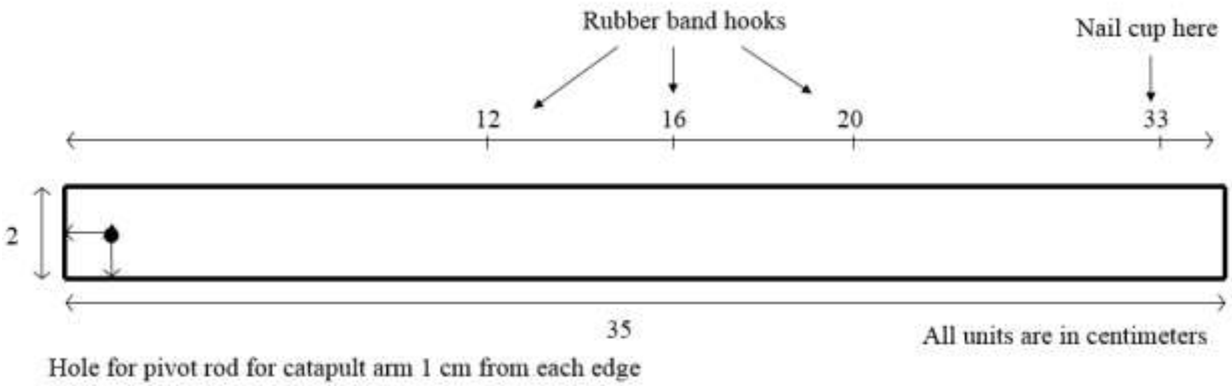

Figure 9. Catapult pivot arm with drilling point and hook attachment points.

This type of catapult provides an opportunity to do more than just analyze data using statistics. The experiment must be well planned and the learners can use quality tools such as an Ishikawa diagram to help them plan the experiments to be conducted (Lunder, 1994). A catapult can also be used, without changing the factors, to generate data for creating control charts and performing capability studies.

A low-cost method for performing a DoE is the paper helicopter (Box and Liu 1999). The response variable would be flight time with leg length, rotor width, and with or without a paper clip as possible factors. Figure 10 provides the design for paper helicopters. With these helicopters, every variant should have one paperclip attached to the leg to add stability. A second paper clip can be added as an additional factor.

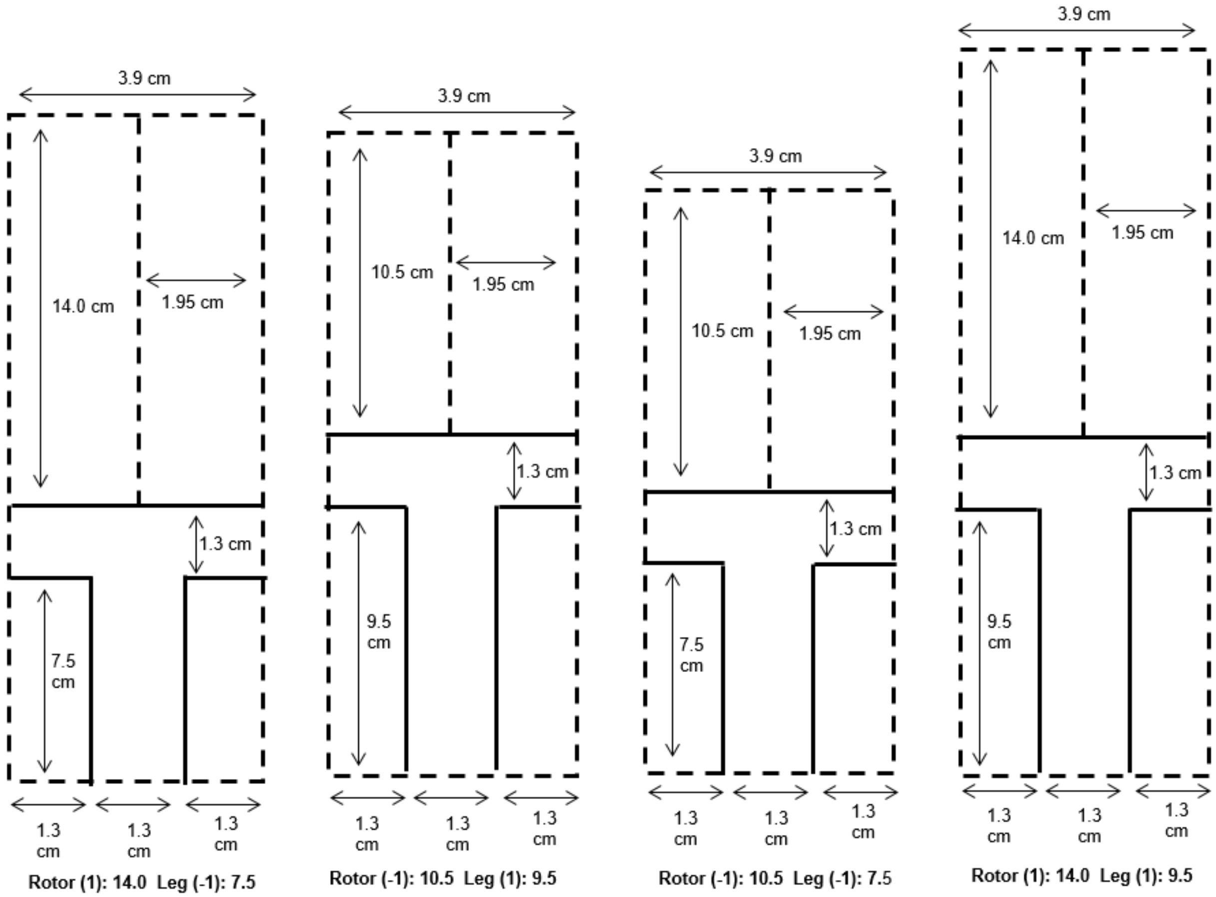

Figure 10. Paper helicopter plans (not to scale).

Not all data are normally distributed and a coin can be used to generate non-normal data with just a ruler. Provide the learners with a coin and place a marking on the floor, three meters from a wall. The task is to toss the coin as close as possible to the wall. Measure the distance between the wall and the coin after each toss. Such data can be used to represent a process where lower values are better and there is a natural limit such as in the case of surface roughness, where low values are often better, but values less than zero are not possible.

5. Conclusions

Teaching statistics requires more than just teaching formulas and the requirements associated with various statistical tests. Simple examples should be used to introduce concepts and then data should be provided to gain practice and insights with statistics as drivers of value creation and good decision-making. Learners should also have to perform their own experiments to generate data to then be analyzed. In this regard, training aids such as catapults and paper helicopters can be useful for teaching learners how to plan an experiment in addition to just analyzing data.

Figure 11 depicts a flow chart for steps to take into consideration while teaching or learning statistical concepts or tools. The illustration provides a good summary of the main messages that this paper attempts to convey.

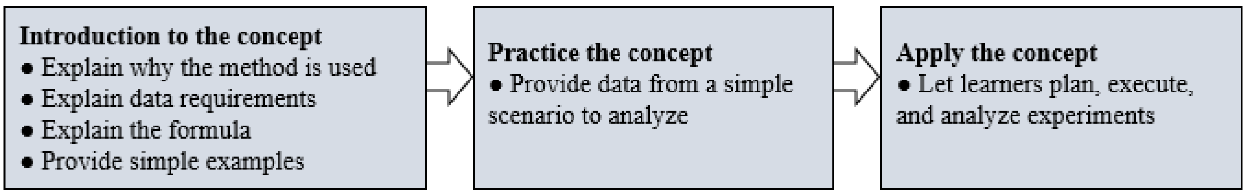

Figure 11. Flow chart of teaching and learning statistical concepts and tools.

For example, suppose that the statistical topic being covered was capability studies with normally distributed data. The educator or trainer should explain why capability studies are performed and then explain the corresponding requirements and assumptions, such as a minimum amount of normally distributed data. The formula should then be introduced and explained, while the entire concept could be illustrated with simple examples. This would be a good time to start explaining how things could go wrong, such as when a process is not in a state of statistical control prior to a capability study being conducted.

Data should then be provided so that the leaners can practice performing the corresponding calculations. This could be done by hand or using software deepening up the course. For both approaches, leaners must become comfortable with both performing the operations required (i.e. doing the math or selecting the correct analysis options in the software) and interpreting the results that are obtained.

Once proficient in calculating a capability value, either by hand or using software, the learners should start planning and carrying out their own capability studies. Ideally, a production process will be available for collecting data. However, this may not be always the case. Training aids can therefore be used as an alternative to generate the required data such, as by firing a catapult 100 times or dropping a paper helicopter 100 times. Another option may consist of using public domain data sets that are accessible online and include real world data, which learners could use to explore.

As we move to an increasingly data-intensive and data-driven world, it becomes even more important for quality professionals, top managers and also citizens to understand the tremendous value that can be extracted from data, namely through statistical tools and methods. In this paper we refer to a number of simple principles, approaches, tools, and resources that can help people learn more about applied statistics, in rigorous yet also attractive ways, and help teachers to promote such learnings in more efficient ways, making the corresponding classes moments of hard work coupled with fun as well.

References

Bisgaard, Søren. 1989. "Teaching Statistics to Engineers." *The American Statisti*cian. 45. No. 4: 274-283

Box, George E.P. and Patrick Y. T. Liu. 1999. "Statistics as a Catalyst to Learning by Scientific Method Part 1-An Example." *Journal of Quality Technology*. 31. No. 1: 1-15

Deane, Rochard H. and Robert H. Burgress. 1998. "Experimental Teaching Techniques in Quality Management: The Roman Catapult." *Quality Management Journal*. 5. No. 2: 58-66

Deming. W. Edwards. 1994. *The New Economics: For Industry, Government, Education*. (2nd ed.). Cambridge, MA: The MIT Press

Lunder, Jeffery J. 1994. "Archiving Continuous Improvement with the Dual Approach: A Demonstration of the Roman Catapult." *Quality Engineering*. 6. No. 4: 691-705

Mitchell, Garry. 1998. *The Trainer's Handbook: The AMA Guide to Effective Training*. 3rd ed. New York: American Management Association

Neumann, David L, Michele Hood, and Michelle M. Neumann. 2013. "Using Real-Life Data when Teaching Statistics: Student Perceptions of this Strategy in an Introductory Statistics Course." *Statistics Education Research Journal*. 12. No. 2: 59-79

Santy, Woody and Bob Einwalter. 1997. "Comparison of Classroom Toys for Teaching Experimental Design." *Quality Engineering*. 10. No. 1: 75-83

Viles, Elisabeth and Cristina Martin. 2008. “Educating Engineers.” 2008. *Quality Progress*. 41. No. 2: 45-51

Yeh, Robert W., Linda R Valsdottir, Michael W Yeh, Changyu Shen, Daniel B Kramer, Jordan B Strom, Eric A Secemsky, Joanne L Healy, Robert M Domeier, Dhruv S Kazi, Brahmajee K Nallamothu. 2018. “Parachute use to Prevent Death and Major Trauma When Jumping from Aircraft: Randomized Controlled Trial.” *British Medical Journal*. 363: 1-6