SAND99-2178C

Paper ID #126

DATA TORTURING AND THE MISUSE OF STATISTICAL TOOLS

Marcey L. Abate

Sandia National Laboratories

This work was supported by the United States Department of Energy under Contract DE-AC04-94AL85000. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy.

CATEGORY: Statistical thinking & methods

FORMAT: Participative

SUMMARY

Statistical concepts, methods, and tools are often used in the implementation of statistical thinking. Unfortunately, statistical tools are all too often misused by *not* applying them in the context of statistical thinking that focuses on processes, variation, and data. The consequences of this misuse may be "data torturing" or going beyond reasonable interpretation of the facts due to a misunderstanding of the processes creating the data or the misinterpretation of variability in the data. In the hope of averting future misuse and data torturing, examples are provided where the application of common statistical tools, in the absence of statistical thinking, provides deceptive results by not adequately representing the underlying process and variability. For each of the examples, a discussion is provided on how applying the concepts of statistical thinking may have prevented the data torturing. The lessons learned from these examples will provide an increased awareness of the potential for many statistical methods to mislead, and a better understanding of how statistical thinking broadens and increases the effectiveness of statistical tools.

KEY WORDS: Statistical methods, Statistical thinking, Variability

INTRODUCTION

Statistical thinking is a philosophy of learning and action that focuses on processes, variation, and data. The philosophy is based on the principles that all work is a series of interconnected processes, variation exists in all processes, and that reductions in variation provide the means for improvement. The successes in applying the principles of statistical thinking to real problems is well documented (Britz et al. 1996) and can be traced back to the fundamental concept that reducing variation leads to improved quality. Because statistical tools are used to quantify and better understand variability, they are often an integral part of the implementation of statistical thinking.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Unfortunately, statistical concepts, methods, and tools are all too often applied in the absence of statistical thinking. This misuse of statistical tools usually results in the misinterpretation of variability, the misunderstanding of the associated processes, and can have momentous consequences.

The consequences of misusing statistical tools by not applying them in a proper framework, such as statistical thinking, can be fatal as illustrated by Tufte (1997). He examines the statistical and graphical reasoning used in making two life-and-death decisions; how to stop a cholera epidemic in London during 1854; and whether to launch the Space Shuttle Challenger in 1986. By using statistical tools to properly understand the process by which cholera was spreading, Dr. John Snow was able to discover the cause of the epidemic and bring it to an end. In contrast, by applying statistical tools that did not properly represent the variability in the data, those who decided to launch the space shuttle made a decision that cost the lives of seven astronauts. In both cases, the consequences resulted directly from the quality of the framework within which the statistical tools were used to understand processes and variability.

STATISTICAL TOOLS AND DATA TORTURING

Statistical tools are often used to quantify and better understand variability in the implementation of statistical thinking. Unfortunately, when common statistical tools are misused, they may fail to properly explain variability and subsequently encourage faulty decisions and actions. Even worse, statistical techniques may be misused to find a desired result. This practice is sometimes referred to as "data torturing", and even though the conclusions from such analyses appear intuitively correct, the conclusions are in reality based on incorrect interpretations of variability. One of the earlier uses of the term "data torturing" is in Mills (1993) who reviews data torturing by summarizing some poor statistical practices, that although well known, are still frequently applied. Mills comments on the meaning of data torturing by stating:

> *... study data, if manipulated in enough different ways, can be made to prove whatever the investigator wants to prove. Unfortunately, this is generally true. Because every investigator wants to present results in the most exciting way, we all look for the most dramatic, positive findings in our data. When this process goes beyond reasonable interpretation of the facts, it becomes data torturing.*

The manipulation and misuse of statistical tools that leads to data torturing may be intentional or not. In either case, the reason that many common statistical tools do not adequately perform is that they are used out of context without a proper understanding of the process that created the data. Because statistical thinking is about processes, variation, and data, understanding statistical thinking concepts increases the effectiveness of statistical tools and helps to prevent data torturing. However, because one may be forced to make a decision or take action on an analysis without the process knowledge that statistical thinking would provide, it is also important to be aware of practices that encourage incorrect interpretation of variability. Among common statistical tools that may lead to improper interpretations of variability are:

- Averages
- Percentages
- Smoothing
- Cross tabulations
- Trend analysis

Although each of these methods are valid when properly used, the potential for each to mislead has been documented through observation in the relevant literature. In the hope of averting future misuse of these statistical tools, the following presents examples where the application of each, in the absence of statistical thinking, provides misleading results by not properly representing the variability in the process under study.

AVERAGES

A common method for analyzing data is to compare a current value to an average, specification, or target value. The conclusion of any such analysis will simply be the binary characterization of the current value as "above average" or "below average". In essence, this method compares an individual value and treats any difference between it and the average, specification, or target as something special.

An example of the dangers associated with this type of analysis is shown in Figure 1 where the percentage of failed monthly inspection results for an organization has been plotted over time. A "red flag" was given if a monthly percentage unfavorable exceeded twice the peer group average (denoted by the solid line on the graph). By ignoring the time progression, and only comparing to averages, the monthly variation in the values is not taken into account and no indication is ever given that the values may be exhibiting a downward trend. This form of data torturing may lead to acting on a perceived difference when none really exists, and potentially wastes time, resources, and effort. If one had instead recognized the statistical thinking principle that all processes are variable, the simple comparison to an average would have not provided a satisfactory basis for decision making. Rather, the importance of the time progression may have been recognized, and an analysis of the data would have attempted to quantify the monthly variability, and possibly recognized a trend in the data.

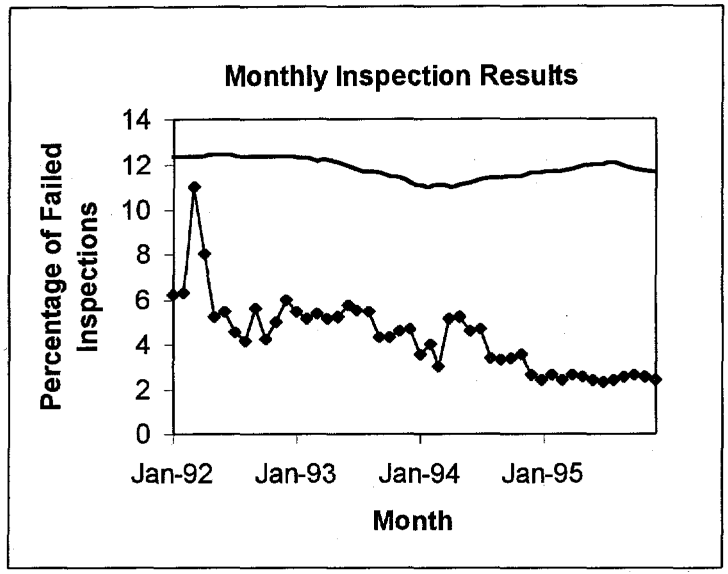

Figure 1: Comparing monthly values to an average

PERCENTAGES

Analysis and display methods using percentage data are also susceptible to data torturing. Because many of these methods do not incorporate a context of common variation, they may encourage improper interpretations of variability. To better understand, consider a typical analysis using percentage data. Figure 2 presents the results of a material aging study where samples of different ages (in years) were obtained, tested, and the percentage of failures in each age category was recorded. Inspection of Figure 2 may cause one to conclude, from a reliability standpoint, that the end of lifetime has been found with the critical age for increased failures at approximately twenty-six years of age.

However, if one examines the number of available samples for each age category, a different conclusion may be reached. Table 1 provides the number of samples tested in each of the age categories of Figure 2. This additional information about the total number of samples tested in each age demonstrates that percentages, especially when presented graphically, can be extremely deceptive because the variability induced by different sample sizes is not taken into account. This form of data torturing, brought about by not considering the denominators of the percents, may lead to taking action on the process with no real justification. If instead the process that created the data had been considered, as advocated by the principles of statistical thinking, then the inappropriateness of calculating and comparing the percentages without consideration of the sample size would have been apparent.

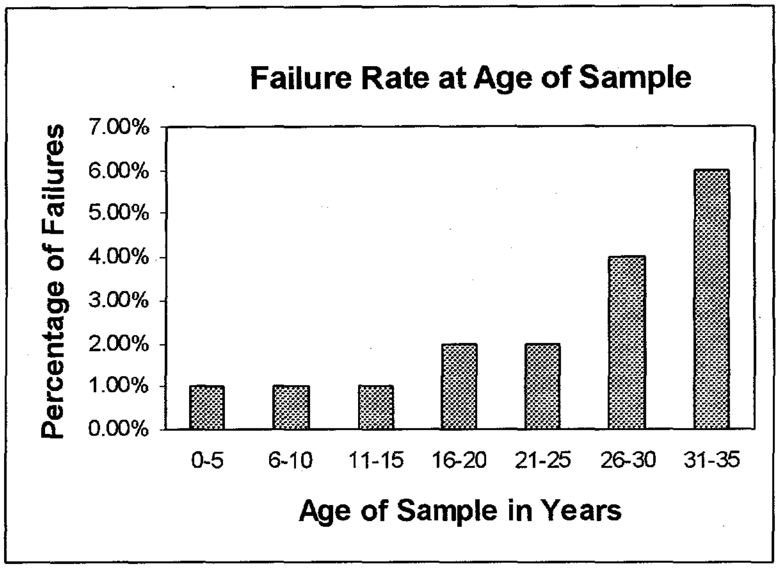

Figure 2: Results of an aging study

Table 1: Age and number of samples

Age in years	Number of samples tested
0-5	7626
6-10	2755
11-15	1773
16-20	1112
21-25	631
26-30	308
31-35	50

SMOOTHING

Smoothing is a technique frequently used in the display and analysis of data. Although it is possible to properly use smoothing to isolate trends or eliminate high frequencies in a time series, it is also possible to apply smoothing procedures and falsely introduce periodicities. Unfortunately, such injudicious practices are becoming more common as smoothing is used to make a corresponding graph look "nice" or "friendly" to the user by hiding the natural variability in the data. An example of the dangers of smoothing is given in Balestracci and Barlow (1997) and is reproduced in Figure 3. The figure shows run charts of data from four processes. A standard analysis may indicate that the process in Figure 3a is stable and contains only noise, while the other three processes (Figures 3b, 3c, and 3d) indicate the presence of trends.

However, as Balestracci and Barlow reveal, the four graphs show the exact same data! The difference is that Figure 3a displays the raw data while the other three figures present the data as various smoothed versions of the raw data. To construct the charts, raw data points were generated from a normal distribution. The raw data are plotted in Figure 3a and exhibit stable behavior, Figure 3b was created by taking rolling averages of four raw data points, Figure 3c consists of rolling averages of twelve raw data points, and Figure 3d was constructed using a rolling average of fifty-two. This example demonstrates that smoothing can result in creating the appearance of something special in the process that does not actually exist. This form of data torturing may result in taking action when no justification exists. If instead, the second principle of statistical thinking, that all processes vary, is recognized then arbitrary smoothing endeavors become pointless.

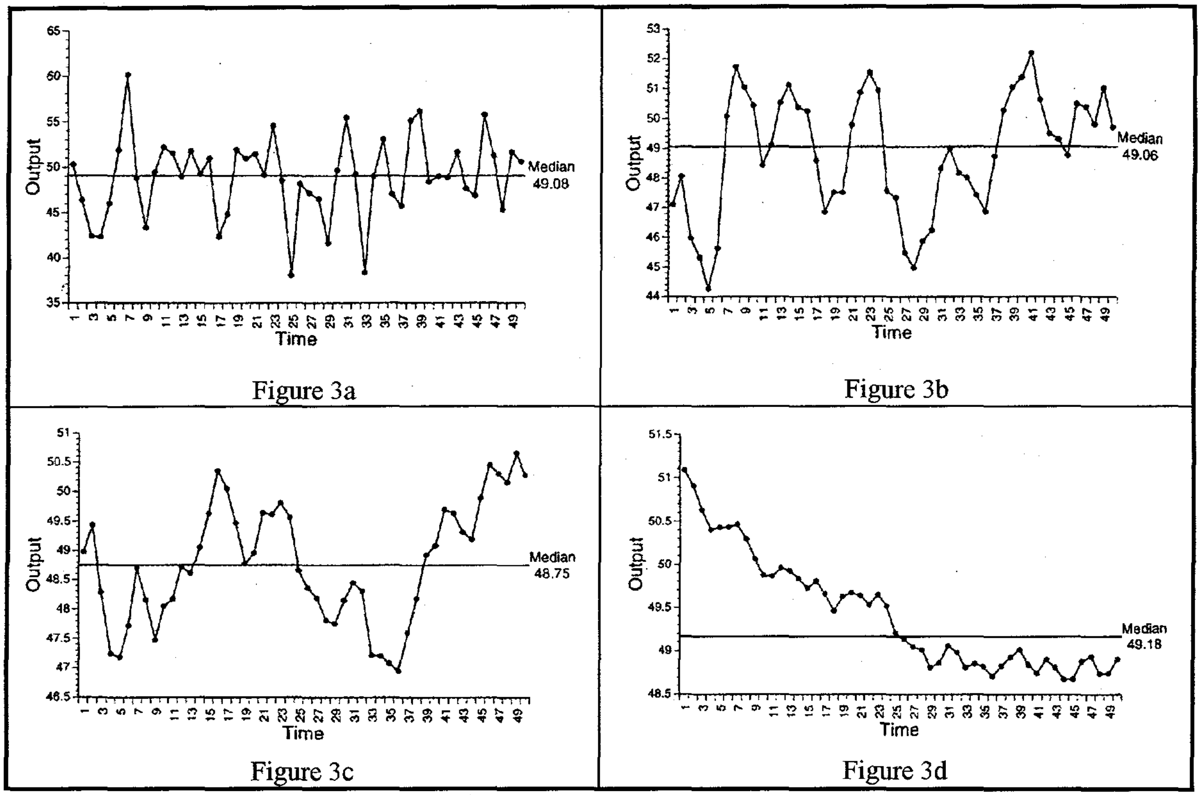

Figure 3a

Figure 3b

Figure 3c

Figure 3d

Figure 3: An example of smoothing a process

CROSS TABULATIONS

A common statistical tool used to summarize count data is cross tabulations. The degree to which the data are aggregated in cross tabulations often does not properly consider the process that created the data. However, the nature of the relationships observed in cross tabulations can differ depending on whether or not the data have been adequately stratified. That is, an observed association between variables can differ depending on whether or not other hidden variables have been taken into account. If data have not been stratified down to an appropriate level, the inherent variability may be hidden, often resulting in a misleading interpretation of relationships. Consider the following example that demonstrates this potential for data torturing. It was desired to compare the performance of two airlines, Airline N and Airline S, on a specific type of safety inspection. The hypothetical results of the inspections are provided in Table 2 and would typically be analyzed by computing the inspection pass rate for each airline.

Using the data in Table 2, the pass rate for Airline N can be calculated as 25/60, or 42%, and the pass rate for Airline S to be 15/50, or 30%. Analyzing the data at this level of aggregation indicates a substantial difference in inspection results and may cause one to expend further time and resources to investigate. However, consider stratifying the data as in Tables 3a and 3b by a third variable, season, that denotes the time of the year when the inspection was recorded. These tables show that the airlines had identical pass rates of 50% during the summer and 25% during the winter! That is, stratifying the data by season revealed that the previous conclusion was not valid and demonstrates that an observed association between two variables can be misinterpreted when a variable that interacts strongly with both variables is omitted from the analysis.

This particular example could be explained by a natural or necessary tendency to avoid inclement weather. For example, the patterns in Tables 3a and 3b could be explained if Airline N was located in the north and the majority of inspections took place in the summer to avoid the harshness of the winter weather, while Airline S was located in the south where summer inspections are often avoided due to the stresses of the summer heat. Because these tendencies are reflective of the natural inspection recording process, the observed differences are attributable to common variability. Concluding anything more is a form of data torturing and may result in expending unnecessary time, resources, and effort. If the process that created the data was well understood and thoroughly investigated by applying the principles of statistical thinking, the possibility for excessive aggregation (as in Table 2) will be reduced.

Table 2: Cross tabulation of inspection results by airline

	Airline N	Airline S
Pass	25	15
Fail	35	35
Total	60	50

Table 3a: Classification of inspection results from the summer

Summer		
	Airline N	Airline S
Pass	20	5
Fail	20	5

Table 3b: Classification of inspection results from the winter

Winter		
	Airline N	Airline S
Pass	5	10
Fail	15	30

TREND ANALYSIS

It is a natural tendency to seek explanations for patterns in data. The statistical methods and tools for pattern detection are collectively referred to as trend analysis and are often misused to make decisions with either limited data points or with inadequate knowledge about the process creating the data. This practice may result in data torturing by wrongly identifying the type of trend or by leading one to conclude that a trend exists when in fact it does not.

An example of the misuse of trend analysis is given in Balestracci and Barlow (1997) and is partially reproduced in Figures 4 and 5. Consider the data points shown in Figure 4. While it is difficult, if not impossible, to formulate a meaningful interpretation of this data by only observing three data points without a broader contextual basis, it is all too common that this data would be labeled an "upward trend". To those that would be tempted to suppose a trend in the data of Figure 4, consider Figures 5a and 5b that present different contexts for the "upward trend" data points. In Figures 5a and 5b, the last three data points in each run chart are the same points as those given in Figure 4. For the process represented in Figure 5a, the behavior of these three data points is atypical compared to previous values. However, the same increase in the last three points when part of a process such as that of Figure 5b, is not unexpected.

Making decisions and taking action on perceived trends from limited data points will almost surely result in data torturing by either underreacting or overreacting. If instead of presenting only the data of Figure 4 to decision makers, the principle of statistical thinking that all processes are variable was applied, then data torturing may have been avoided by providing an appropriate context of variability.

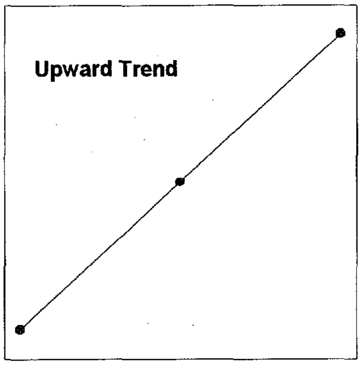

Figure 4: An upward trend

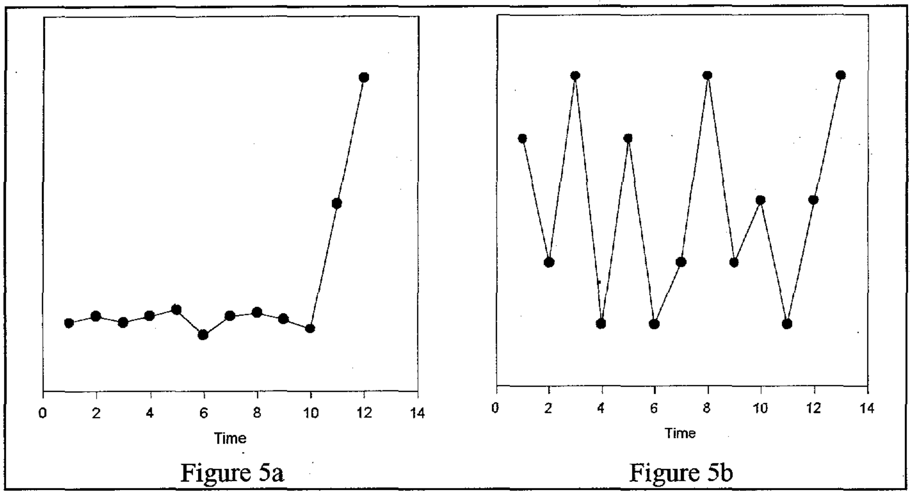

Figure 5a

Figure 5b

Figure 5: An upward trend in two different contexts

As a second example of the potential for misuse of trend analysis, consider fitting a trend line to a given set of time series data. Often the conclusions of the analysis present only the resulting trend line with no reference to the original data. However, this practice can very easily lead to attributing a linear trend to data that contain no trends or contain a different type of trend. As an example, consider the scatter diagrams in Figure 6 from Anscombe (1973). Although the four sets of data in these diagrams all differ, fitting a linear trend line to each yields the same regression results!

In these cases, either different types of trends or no trend at all is influencing the data. Clearly, considering the trend lines outside the context of variability provided by the raw data may lead to data torturing. If instead, the principles of statistical thinking had been applied, the process that created the data would have also been considered and the possibility for faulty decisions or actions based on trend analysis would have been reduced.

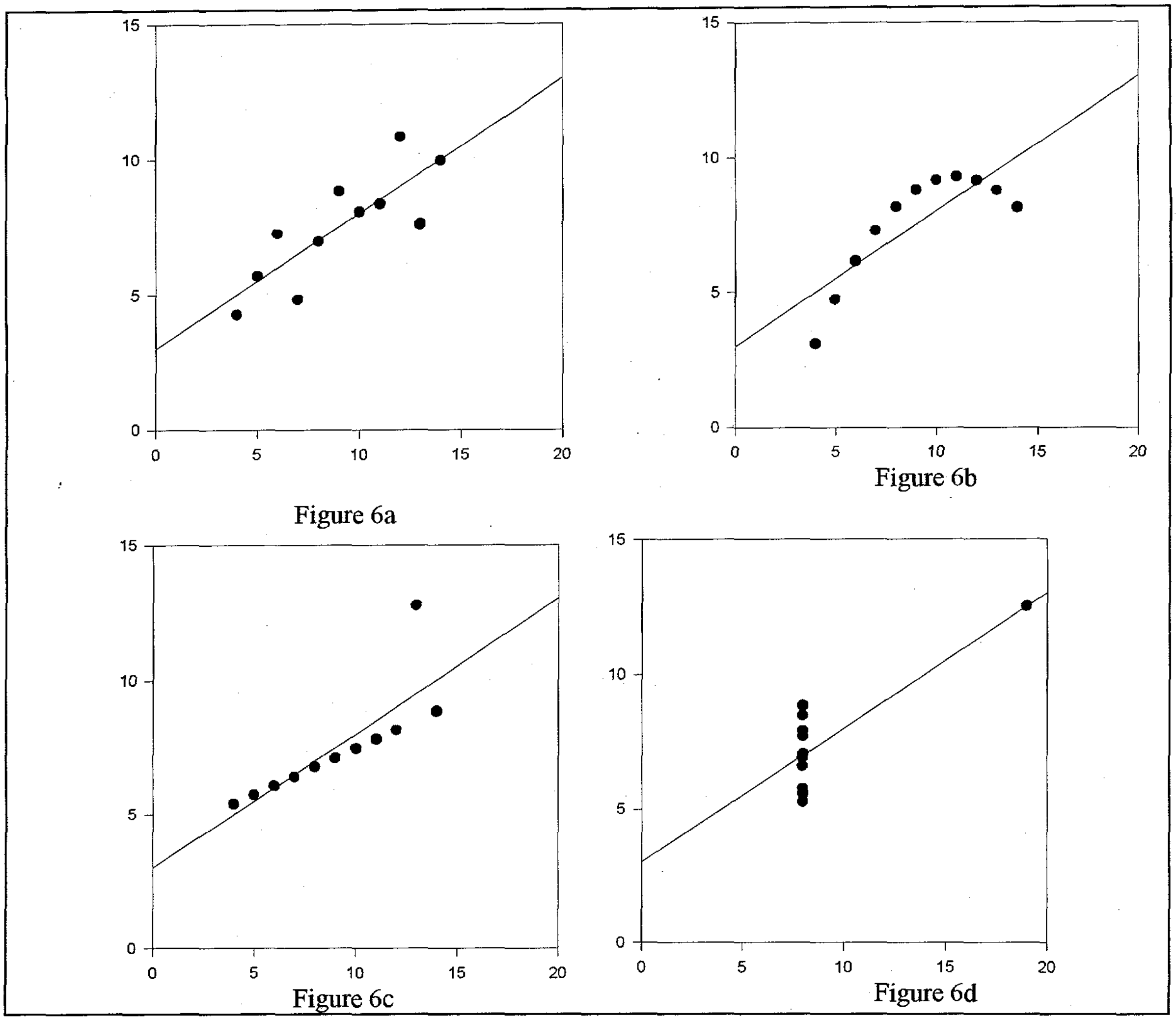

Figure 6a

Figure 6b

Figure 6c

Figure 6d

Figure 6: The same trend line resulting from different data

CONCLUSION

The previous examples demonstrated that common statistical tools and methods are capable of data torturing. Specifically, the following key points were highlighted.

- Comparisons to averages, specifications, and targets ignore common variability and treat every fluctuation as something special.
- Analysis and display of percentage data need a context of common variability for proper interpretation.
- The smoothing process can create the appearance of something special that is not in reality exhibited by the data.
- Stratification is necessary in cross tabulations so that excessive aggregation is not allowed to hide the inherent variability.

- Limited point comparisons automatically make a special interpretation of the variability, even if no trends have been observed.
- Fitting inappropriate trend lines may result in attributing a specific type of trend to data that contain no trend or a different type of trend.

The potential for common statistical tools to torture data stems from an inability to adequately account for the variability in the underlying processes. Unfortunately, it is easy to misuse the most straightforward of statistical tools with possible fatal consequences. These shortcomings make evident the importance of applying statistical thinking even when using basic statistical tools. As repeatedly shown, failure to consider the processes, variation, and data within the mindset of statistical thinking can result in faulty decisions and actions. In summary, because statistical thinking requires a focus on the process, the application of the associated concepts will increase the effectiveness of statistical tools and help to prevent data torturing.

REFERENCES

Anscombe, F.J. (1973), "Graphs in Statistical Analysis", *The American Statistician*, Vol. 27, No. 1, pp. 17-21.

Balestracci, Davis, Barlow, Jeanine L. (1997*), Quality Improvement: Practical Applications for Medical Group Practice*, Second Edition, Center for Research in Ambulatory Health Care Administration, Englewood, CO.

Britz, Galen, Emerling, Donald, Hare, Lynn, Hoerl, Roger, Shade, Janice (1996), "Statistical Thinking", *Special Publication of the American Society for Quality Statistics Division*, Spring 1996.

Mills, James L. (1993), "Data Torturing", *New England Journal of Medicine*, Vol. 329, pp. 1196-1199.

Tufte, Edward R. (1997), *Visual Explanations: Images and Quantities, Evidence and Nature*, Graphics Press, Cheshire, CT.