

IE 361 Module 9

Simple Principles in the Collection of Industrial and Engineering Data
and Simple Statistical Graphics for Quality Assurance

Reading: Sections 2.3, 2.4 *Statistical Quality Assurance Methods for
Engineers*
(Sections 1.4 and 1.5 of Revised *SQAME*)

Prof. Steve Vardeman and Prof. Max Morris

Iowa State University

August 2008

The Habitual Collection and Display of Process Data

Step #3 in the 6 Step Process-Oriented Quality Assurance Cycle of Table 1.1 of *SQAME* urges the regular collection and summarization of data on process performance. Data from a single period serve to give one a snapshot of process performance. Comparison of data sets from multiple periods allows one to see and act on trends in process performance. In this module we make some observations about real world data collection and then note how effective some of the simplest tools of statistical graphics can be in giving one a quick picture of the main features of a process performance data set.

Simple Principles of Engineering and Quality Assurance

Data Collection

Useful and informative data don't just magically appear in a data file or on a data collection form. They must be gathered by real humans (or "automatic" data collection systems designed and implemented by humans) and are only as useful as the wisdom, care, and genuine good will that go into their collection. There are technical matters that impact the usefulness of QA data (like exact sample size choices and particular choices of structure for a data collection plan). But whether data are to be used to monitor process stability/performance or to guide changes aimed at process improvement, there are also some simple qualitative guidelines that are relevant. Among these are the following:

- There must be usable operational definitions of the quantities on which data are to be gathered. Where measurements are to be taken, the measurement equipment itself must be stable/well-calibrated. Technicians must be properly trained in the meaning of the definitions and the use of any equipment involved.

Simple Principles of Engineering and Quality Assurance

Data Collection

- A small or moderate amount of carefully collected and immediately used data will almost always be worth much more than even a huge amount of thoughtlessly collected or never used data.
- It is the absolute size (rather than the relative size) of a sample and the basic process/population variability that determines the information content of a sample. (For example, a blanket "take a 10% sample" rule will sometimes over-sample and sometimes under-sample.)
- The closer that data are taken (in time and space) to an operation whose performance they are meant to reflect, the better. (The ideal is probably data collection by well-trained process workers who have adequate time for the task.)

Simple Principles of Engineering and Quality Assurance

Data Collection

- Routine data collection should be made as convenient as possible, and where feasible, any form used for data collection should make them immediately useful (without transfer, e.g., to another form or medium). The point is to get data used, not to make presentation quality displays.
- In order to be useful in indicating sources of variation in a data set, care needs to be taken to keep track of conditions surrounding each observation (e.g. machine number, operator, etc.).
- One must take into account psychology and politics when assigning data collection tasks. He or she who is to collect data should be convinced that their production is a help rather than a threat, and that faithful representation of a situation (rather than "good numbers") is the goal.

Simple Statistical Graphics for Quality Assurance

Once data are collected, it is important to get them immediately used. Often, very simple statistical graphics (like those taught in the first couple of lectures of an elementary statistics course like Stat 231) can be amazingly effective in conveying the most important features of a data set and suggesting what is going on in a process (and, sometimes, how it might be improved).

Histograms provide very simple summaries of distribution center, spread, and shape, that are understandable by even very non-quantitative persons. Usually, a (not necessarily symmetric, but) fairly smooth "unimodal" (one hump/"up and then back down again") shape is what one expects to see in a snapshot of process data. When something else appears, there is typically an interpretable cause for that shape that provides valuable information for process understanding and improvement.

Simple Statistical Graphics for Quality Assurance

The next figure is a **bimodal** (two-mode) histogram of measured axle diameters (in units of inches). Its shape suggests that there are two versions of some process element upstream ... two fundamentally different raw materials sources, two (differently adjusted) lathes, two machine operators with different understandings of how to run a single lathe, or ... This possibility can produce variation in diameter that might be avoided (thereby improving quality). Sliding the two "humps" below together and would reduce the overall spread in measured diameter.

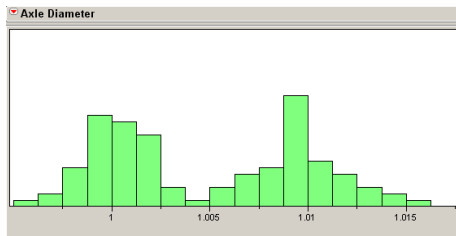


Figure: Bimodal Histogram of Some Axle Diameters (in)

Simple Statistical Graphics for Quality Assurance

By the way, this notion of eliminating multiple versions of a single process element is part of what motivates modern trends to reduction of supplier bases. People even go so far as to consider "single sourcing" for parts and components that must be very uniform. Reducing numbers of vendors helps an organization reduce variability, time and hassles associated with switch-overs in production between sources, and the overhead involved with dealing with more vendors than absolutely necessary. (There are also, of course, dangers associated with depending upon a single source for critical parts and materials.)

In a first course in statistics, it is common to provide nomenclature for shapes of histograms with little consideration of how different shapes might arise or how one might use information about origins. In a QA context, what a graphic or numerical summary tells one about a process and what kind of action it suggests is paramount. Take, for example, the pair of histograms in the next figure. These are net contents (in ml)

Simple Statistical Graphics for Quality Assurance

delivered to pop bottles by two different filling processes, where the nominal fill level was 591 ml.

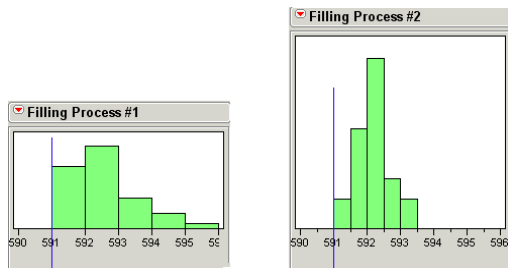


Figure: Fill Levels Produced by Two Different Filling Machines (in ml) Where the Nominal Content is 591 ml

The story suggested by these two histograms is this. Process #1 is apparently fairly imprecise. The apparent upshot is that guaranteeing that

Simple Statistical Graphics for Quality Assurance

bottles have at least the nominal contents requires sorting and discarding/reworking under-filled bottles. This would explain the *left-truncated* or *right-skewed* appearance of the first histogram. Filling process #2 appears to be more consistent/precise and sorting/rework may be largely unnecessary with this machine. Notice too that the figure on panel 9 also suggests that the mean of the final Process #2 distribution is closer to 591 ml than that of the Process #1 distribution. (This is possible because of the smaller spread associated with Process #2 fill levels.) That has the additional happy consequence (beyond reducing or eliminating the need for rework) that one needs to give away less product in order to meet the nominal fill level (a nice additional bonus of small variation).

The figure on panel 9 already begins to illustrate the important point that marking engineering specifications on a histogram is an effective way to communicate whether a process is currently meeting engineering

Simple Statistical Graphics for Quality Assurance

requirements or is even capable of doing so if one could adjust the process *aim* (the process mean). The four possibilities in this regard are illustrated in the next figure.

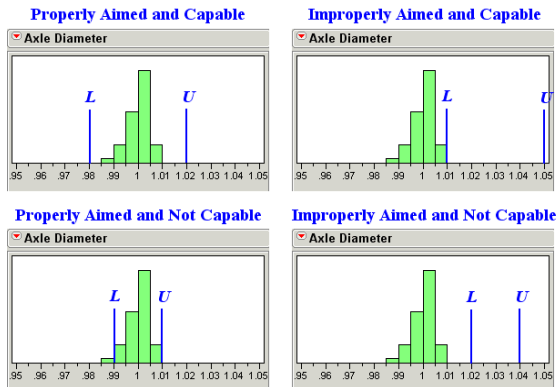


Figure: The Same Process Distribution Compared to Different Engineering Specifications (Axle Diameters in in)

Simple Statistical Graphics for Quality Assurance

Proper process *aim* is a question of whether the process mean is on target at the mid-point of engineering specifications. Process *capability* is a matter of whether process variability is small enough that if properly aimed, essentially the entirety of the process output would fall within engineering specifications.

A kind of bar chart that has experienced popularity in QA contexts is something called a **Pareto Diagram**. One represents some measure of the importance (like frequency of occurrence or dollar impact) of different quality problems by bars of corresponding heights, placed in order largest to smallest, and often augments this with a line plot of cumulative importance. The next figure is an example of a Pareto diagram that might represent what is found when a bank auditor checks a number of auto loan files looking for irregularities.

Simple Statistical Graphics for Quality Assurance

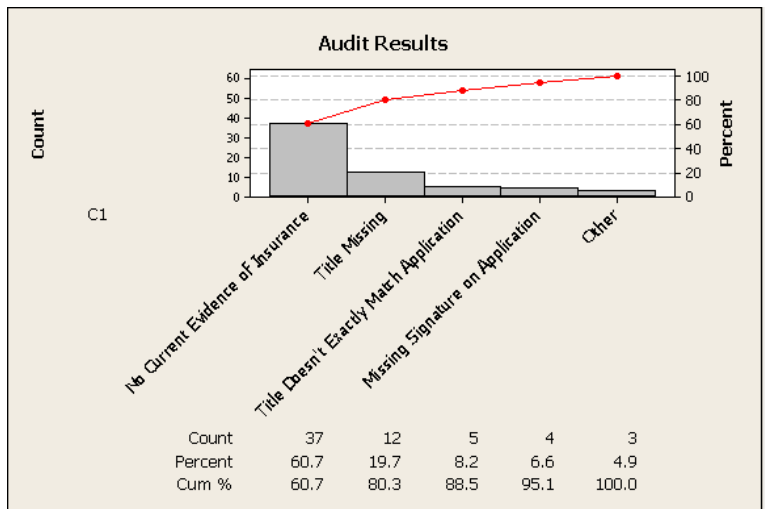


Figure: Pareto Chart for Bank Audit of Auto Loans

Simple Statistical Graphics for Quality Assurance

The idea of a Pareto diagram is to get people to focus on "big" problems first. It is fairly standard QA doctrine that most problems are traceable to a very few causes (and if one can track down and eliminate them, one can achieve results big results from fixing a relatively few things). In fact, the "Pareto principle" is that 80% of anything is traceable to 20% of its sources. (80% of all books read are read by 20% of people. 80% of hot dogs consumed are consumed by 20% of people ... etc.) Pareto diagrams function as effective political tools to get people to focus first on big problems.

The simple statistical device of making a **scatterplot** is an effective way of portraying relationships (or lack thereof) between quality variables. Where there is an apparent relationship between variables, one can consider the possibility that there is some common "cause" that stands behind them both, while if there appears to be no obvious relationship, it is unlikely that such a common source exists.

Simple Statistical Graphics for Quality Assurance

The figure below suggests that upon counting defects of types A,B, and C in several lots of product and plotting counts per lot, it seems possible that a common cause exists for types A and B, but that C likely has a separate source. This can be valuable information as one attempts to narrow a large list of potential problems to shorter ones for subsequent careful study.

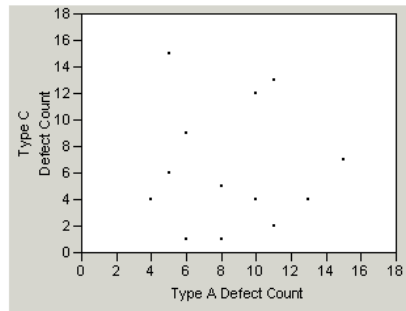
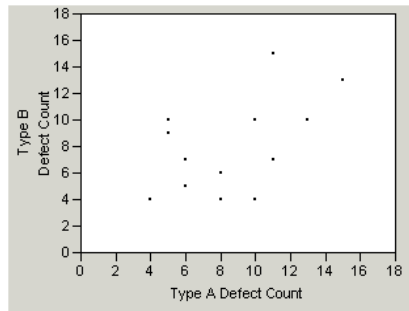


Figure: Two Scatterplots of Defect Counts on 13 Lots of Product

Simple Statistical Graphics for Quality Assurance

Plotting measures of process performance against time is one of the most basic and effective habits that a process improvement analyst can develop. Products and services are produced and delivered over time. Changes over time create variation and "unquality." Plots can show trends and make the timing and size of changes easy to "see" (and thus allow one to begin to track down and eliminate corresponding sources).

The next figure shows measured diameters of 40 consecutive axles turned on a particular lathe. It is very easy to see on this plot that over time (and presumably as the cutting tool used to make the axles wears) the diameters are both increasing in size and becoming less consistent. This kind of a plot (commonly termed a **run chart**) and economic considerations would be important input for the development of both an adjustment schedule for the lathe and a tool change schedule. (The upward trend in the picture could typically be to some degree mitigated by an appropriate tool position adjustment policy—see *SQAME* Section 3.6 for some discussion

Simple Statistical Graphics for Quality Assurance

in this direction—but it is likely that decreased diameter consistency over time could only be addressed with a tool replacement.)

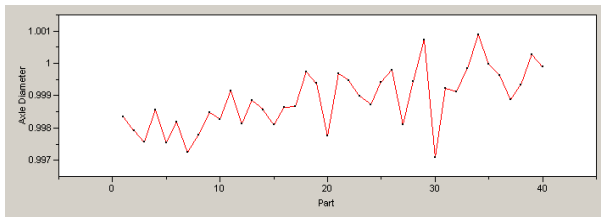


Figure: A "Run Chart" of Axle Diameters (in)

The most important type of "plot against time" for QA purposes is the Shewhart Control Chart. The next Module opens our discussion of the use of this important and famous tool. Our purpose here has been to present the most elementary ideas of displaying process data, and to emphasize how effective that even these tools can be in QA efforts.