STATISTICAL GRAPHICS - A GREAT TOOL-KIT

Edward R. Hansen The Gillette Company, Boston, Mass.

Using any one of many available computer packages, it is possible to analyze data and prepare reports, using graphs to communicate the most important information. Graphs allow readers to understand relationships, data centering and variability very readily. Well-designed graphs contain tremendous amounts of information and our minds interpret such "pictures" very easily. But graphs are not a substitute for judgement and experience; like any other tool they must be understood and used correctly. They are particularly useful in "exploratory data analysis," where one is in the initial stages of studying a problem, to help plan more formal experiments, etc. This exploratory process is like peeling an onion; as each layer is removed you gain more insight into the process factor(s). Using graphics, the accompanying tears as you peel the onion will be minimized.

Most current texts on statistical methods emphasize the value of using simple graphs and plots to get an overall view of the data, look for errors, outliers, non-normality, etc. We will see how to do this as we proceed.

I have used data from a plastic injection molding process on a characteristic, "Housing," to illustrate a graphical approach to data analysis and reporting. Our objective is to run experiments to judge the capability of the process and to prepare a report with our conclusions and recommendations. Graphs are particularly useful in this situation because of the complexity of the molding process due to the multiple sources of variation.

The first stage in our plan is to take a sample of 10 complete shots taken over one shift of production. (A "shot" is one piece from each cavity taken during one cycle of the molding press.) Figure 1 shows measurement data from this sample of the housing dimension, displayed as a "Stem-and-Leaf" diagram. (This technique is explained in the summer issue of the Statistics Division Newsletter.) Two tentative conclusions can be drawn; one piece at -.0019 may be an "outlier," and the data looks distinctly bi-modal. We have peeled one layer from the onion. Figure 2 is a histogram of the same data so you can compare the two graphical approaches. My preference is the Stem and Leaf diagram because it displays every number. The Histogram abbreviates the amount of information by using class intervals based on arbitrary criteria.

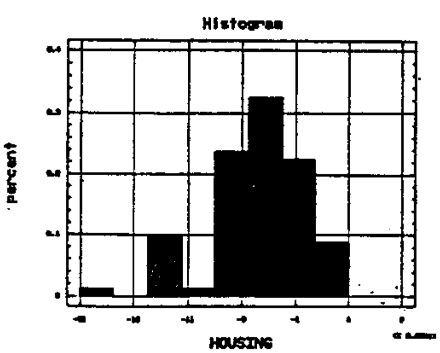
Figure 1

```
      Stem-and-leaf display for
      HOUSING: unit = 0.0001

        LO|-0.0019

  6     -1|44444
  9     -1|222
 10     -1|1
 14     -0|8888
 22     -0|77777777
(23)    -0|55555555555555544444444
 35     -0|3333333333222222
 19     -0|1111111100
  9      0|001111
  3      0|222
```

Figure 2

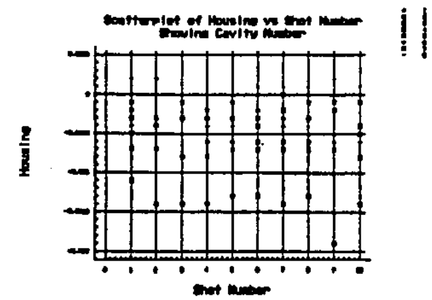


Next, let's use a general purpose tool called "stratification," in which we break down the data into levels or strata, to obtain a different perspective. Fortunately, the data is classified by time of sampling and mold cavity number, two obvious ways of stratifying molding data.

A simple "Multi-Vari" plot gives a "snapshot" of the differences in dimensions of parts from the 8 cavities and 10 shots as we can see from Figure 3. Further deductions can be made from this chart. The possible outlier is from cavity 4, shot number 9.

Cavity 4 output consistently runs well below the other cavities and cavity 8 is usually the highest. This fact explains the bi- modal appearance of the histogram and the Stem-and-Leaf Leaf diagram. Now we have peeled the "onion" down to a more useful level.

Figure 3



It is evident from Figure 3 that the averages of the 8 cavities varied somewhat, but the variability within each cavity was about the same. An idea of the variability over one shift of the process can be derived from this graph.

The dimensional variation of parts produced by a plastic molding process has two primary causes.

1. Variation between cavities, as we have seen in Figure 1. These differences should remain stable, as they are the result of the way the steel in the mold was cut and the mold/process design. This source of variation is visible as the differences in the cavity averages. Thus each cavity is a "mini-process" with it's own average and variation.

2. Variations from cycle-to-cycle of the molding process. This is the result of many small changes in temperatures, pressures, timing and a whole host of other factors. This source of variation effects all cavities on each cycle, although not always in the same amount, due to possible interactions. This variation can be measured by changes in the averages of complete shots of the 8 cavities.

Figure 4 is a very useful type of graph, helpful in stratification, called a "Box Plot," used here to show the variability between cavities explained in #1. Each cavity has its own box, based on the 10 samples measured for the cavity. The explanations for the features of this type of plot are as follows:

1. The center line in the box is at the median value, which splits the ordered data values in half.

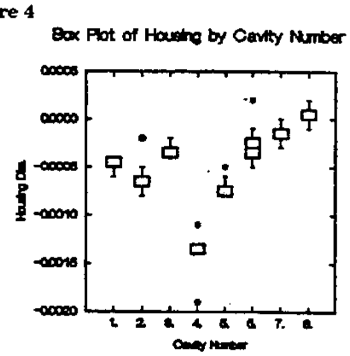
2. The rectangular box ends are placed at the 25%ile and the 75%ile of the measurements, called the "Interquartile Range" (IQR), thus splitting each half in half again.

3. The two lines from the box, called "whiskers," extend out to a distance of 1.5 times the IQR beyond the box ends, to the "fence."

4. Points marked with an "*" are beyond the fences and those marked with an "o" are beyond 3 times the IQR and show probable outliers. A "well-behaved" (normal) distribution would have all points inside the fences.

The Box Plot shows at least two and possibly five outliers, which need further investigation. Since all ten shots had been saved, it was possible to recheck these pieces. Three of the outliers were errors in recording and two were measurement errors. Another layer has been peeled away.

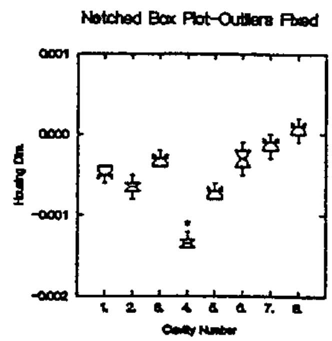
Figure 4



You will note the analysis so far has calculated no averages, standard deviations or other formal statistics. It is my experience that one should always review the raw data from a skeptical viewpoint using graphs, common sense and process knowledge, before one proceeds to more formal methods. There may be errors or other problems that require "purification" of the data before continuing.

Figure 5 is a "Notched" Box Plot of the data after purification. The notches around the medians allow us to make multiple comparisons between cavities. If the notches of any two cavities do not overlap, we can be approximately 95% confident that the medians are different. We still have one outlier "*," but this one was not a measurement or recording error.

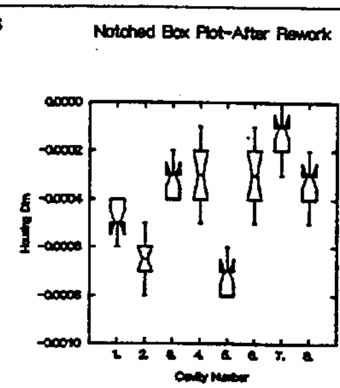
Figure 5



Now we can safely confront a potentially more costly and difficult question. What shall we do about cavities 4 and 8? Both medians are away from the target of -.0004 in., and also from the other cavities.

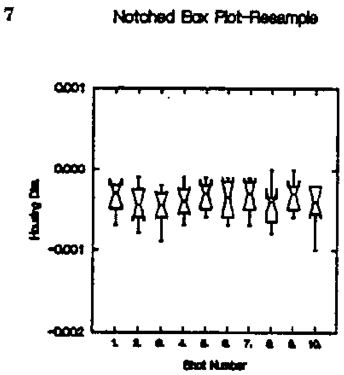
Let us assume that the decision is to rework these two cavities to bring them closer to target. After the rework, a further sampling of 10 shots is selected over one shift. Figure 6 is a Notched Box Plot that shows the result of this action. Notice that even after reworking cavities 4 and 8 we still have significant differences between cavities, but the range of these differences is much less. Assuming a specification of -.0004 .001, the capability of the process <u>may be</u> acceptable.

Figure 6



To gain further assurance, we resampled the process with 10 complete shots randomly taken over a one week period; Figure 7 shows the result. This confirms the earlier sampling. We find no outliers and the notches overlap on all 10 samples. The process appears well centered and capable of meeting customer requirements as reflected in the specifications.

Figure 7



In investigations of this type, it is important to determine if the data can be classified as "normally distributed," since the use of most statistical techniques requires normality. For instance, if we wanted to make predictions about the capability of the process with probability limits, normality would be required. Figure 8 is a "Normal Probability Plot" that shows a graphical approach to an evaluation of the normality of the data from the sample used in Figure 7. A normal probability plot is made as follows:

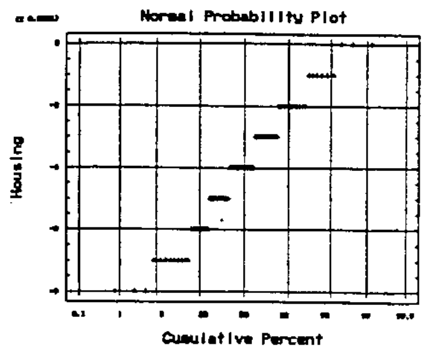
1. The data is ordered and for each point an associated percentile value is calculated.

2. The data is plotted against the Housing dimension on the Y-axis and the cumulative normal percentiles on the X-axis.

3. Some PC programs convert the percentiles in #1 above to Z values and then plot them against a linear axis of Z values ranging from -3 to +3. The end result is the same in that a straight line fit indicates normality.

Fortunately, PC statistics software does all this heavy labor with a click of a mouse. If a straight line fits the data "reasonably well," then the we judge the data to be normally distributed. From a graphical standpoint, this approach is superior to superimposing a normal curve to a histogram. The eye can judge the fit of the cumulative distribution of the data to a straight line more accurately than the bars of a histogram to a curved line.

Figure 8



Unfortunately, our data distribution seems to have heavy tails, and is not normally distributed. This is not surprising, considering that we have only eight cavities and one high or low would result in 1/8th of the values high or low, giving us the heavy tails. This problem could exist in any multi-stream process, in which the number of streams is small (less than 25).

Another problem which contributes to the lack of normality, is that our measurement procedure resolution (.0001in.) is large compared to the spread of the data (.0008in.). This raises the question of whether our distribution is "continuous." While it would have been "cleaner" if our work had resulted in a nice, normal distribution, this is more like real life "in the trenches." I hope our readers will not be too disappointed, but the techniques used here are valid in any case.

More work would have to be done on the measurement system and on "rationalizing" the cavities to center them if we needed to assure normality. Future decisions about the process regarding control procedures and estimates of output capability must consider these factors, so clearly shown by Figure 8.

Figure 7 shows a process which meets customer requirements; Figure 8 says "problem!" What should be done? This is a typical question which must be answered. Graphs can help to illustrate the problems; people must make decisions.

Results of analysis of data from the study of multi-stream processes, such as this example, are difficult to explain in reports. A table of statistics, cavity numbers, percents out of specification, becomes almost overwhelmingly complex, whereas graphs such as Figures 6, 7 and 8 portray the important information very effectively. These three graphs, with appropriate commentary, could form the basis for the final report with conclusions and recommendations.

This brief example has shown only a few of the many powerful graphical techniques available. In recent years, graphical approaches to statistical data analysis and reporting have come to the forefront, spurred on by the many PC software packages that are available to do the "grunt work." I hope all our readers use these techniques to support their analyses.

Execustat (Copyright 1990 Strategy Plus, Inc.) and SYSTAT (Copyright 1990 SYSTAT Inc.) were used to create the graphs used in this study.

Reference

Chambers, J.M., Cleveland, W.S., Kleiner, B., and Tukey, P.A., Graphical Methods for Data Analysis, Duxbury Press, Boston, Ma., 1983.

ASQC

STATISTICS DIVISION MEMBERS — OPPORTUNITY TO UPGRADE STATUS TO SENIOR MEMBER

If you are currently an ASQC member who meets the following eligibility requirements you should apply for Senior Member status. Eligible Members must:

1. Be at least 30 years of age, and actively involved in the quality profession for at least 10 years. Graduation in an approved engineering science, mathematics, or statistics curriculum is considered the equivalent of 4 years of professional experience.

and,

2. Qualify under one or more of the following:

a. be responsible for important engineering or inspection work involving quality control for at least 2 years.

b. be an instructor of quality control, engineering or statistical methodologies as applied to quality control for at least 2 years; should be capable of teaching a variety of courses in the quality field.

c. be a professional engineer or member of a technical society of national status in any country for which the qualifications require a standing equivalent to that required for a senior member of ASQC.

All eligible Members interested in upgrading their membership status should contact the Statistics Division Examining Chair, Bob Perry, at Grand Metropolitan Technology Center, 330 University Avenue S.E., Minneapolis, MN 55414, or at (612) 330-8916.

Call for Videotape Reviewers

Quality Press is actively seeking authors who might have insight into quality as it relates the health care industry; transportation; food, drug, and cosmetic industries; retail sales, utilities and other service industries. For more information, contact Deborah Dunlap, Acquisitions Assistant, Quality Press, American Society for Quality Control, P.O. Box 3005, Milwaukee, WI 53201-9488 or call 414-272-8575, extension 7292.

QA DAY AT ARGONNE NATIONAL LABORATORIES

Quality Assurance Day at Argonne National Laboratories will take place in April, 1992. Statistics Division sponsors a session at this conference. Basic tools or applied industrial papers are needed for this session. Presentations should be 45 minutes in length. If you are interested, contact Jed Heyes (708) 291-4229, FAX (708) 291-4280.