Is it difficult to use statistical tools properly? Some think not. As Lynne Hare, 1988-1989 Chair, has remarked, "There are those who would have statistics oversimplified to the point of following ten easy steps as one might follow a recipe."

Let's follow up on this idea. How would you react if your co-worker, gone from work for a few days, came back and announced that, based on a three-day short course in Chemistry, he was now applying for the job as a chemist in your plant's metrology lab? Or your co-worker had just gotten a job as a Mechanical Engineer based on a nutrition course completed 10 years ago?

This seems insane. Yet all too often I talk with people in industry who are doing this in statistics. They have been given either a title or a responsibility that requires a good working knowledge of key statistical methods, yet their knowledge of these methods can only be traced back to a short course or two, or to an old college course in statistics.

Why does this state of affairs exist? Consider the following two points. First, in the last 5-10 years, SPC, and DOE (including Taguchi's contributions), have been seen as key parts of solutions to many companies' problems; in response, many companies decided to do "mass training." Many short-courses ads indicated, implicitly or explicitly, that the participant will be an "expert" after 3-5 days of training. Second, most would agree that problems in Chemistry and Mechanical Engineering are often open-ended and require professional judgment. Yet many view statistics primarily as containing only the "seven tools" or so; and furthermore, these tools are too often thought to be applied to most problems, as Lynne suggests, in simply a cook-book fashion.

But can we also go too far in the other direction, by making statistics appear too complicated? To continue Lynne's remarks, "Conversely, there are those who would have the world believe that statistics is some kind of magic art and that it is so complex that mere mortals cannot understand it." This belief, unfortunately, can be fostered by statisticians themselves — often without intent to do so — by overuse of technical terms, failure to provide simple interpretations of data, use of needlessly complex tools, and underuse of graphs.

I am of two minds on this issue. I continue to be amazed at the extremely high amount of variation I see in the correct use of statistical tools.

On the one hand, a number of statistical tools do not appear to be very difficult. For example, control charts along with control limits can be calculated without much difficulty. The thinking itself that forms the basis for control charts can be understood by most (although the derivations of the control limits require some sophisticated mathematics). Many quality professionals have an excellent understanding of control charts, and are involved in having them put to appropriate and important uses.

On the other hand, these tools can be, and often are, misused. Some obvious misuses of control charts include using specification limits instead of control limits on the chart; using the charts for impressing a customer or blaming an operator instead of for improving and controlling a process; overuse of charts, such as charting too many characteristics of a CONTINUED ON PAGE 2
Chair's Message continued

part; and either updating control limits too seldom or too often. Less obvious misuses also frequently arise. One misuse I often see is inappropriate subgrouping, such as taking one sample from each spindle of a six-spindle machine, and grouping the six into a subgroup without knowing whether the spindles may be acting as different processes. Another misuse I often encounter is charting a process whose measurement system has not been explored. A follow-up measurement-system study too often finds that the measurement system has so much variation that the control chart is essentially a chart of measurement variation instead of process variation.

What are your thoughts on this issue? Do you see statistical tools being properly applied? Misused? Do you find the proper use of these tools obvious? Obscure?Send Jed Heyes, our new and hard-working Newsletter Editor, your thoughts. Perhaps you could include a key example to buttress your comments. Think about it: with over 16,000 members, say only 1% of you responded to this request: 160 examples! These 160 examples would themselves constitute a worthy contribution to the education of the Division members. (Jed, don't schedule any vacation for the next three months...)  Δ

Send all letters EXCEPT FOR CHANGE OF ADDRESS to: 
Jed Heyes 
Rank Video Services America 
555 Huehl Road 
Northbrook, IL 60062 
(708) 291-4229

EDITOR'S CORNER

Meet the Editor

Born in England, Gerald B. Heyes (Jed's real name) came to the United States in 1968. An ASQC senior member, CQE and CQA, Jed has been active in the St Charles, IL section since 1979, participating on committees and speaking at section meetings. Jed has taught a number of courses and seminars in SQC, Process Troubleshooting and Exploring Data. Having published in periodicals such as Quality, Quality Progress, Quality Engineering, and the Journal of Quality Technology, he is currently writing an SQC textbook. He is past Region 12 Councilor for the Statistics Division and has made technical presentations at Division sponsored conferences including Argonne National Laboratories.

Mr. Heyes was nominated to the International Who's Who in Quality by the St. Charles section of ASQC, and holds a B.A. in Biology from Blackburn College in Carlinville, Illinois.  Δ

ANNOUNCING A NEW "HOW TO" BOOKLET

"How to Detect and Handle Outliers" a new booklet in the Statistics Division "How To" series has just been accepted for publication. The book is co-authored by Boris Iglewicz of Temple University and David C. Hoaglin of Harvard University.

Data sometimes contains apparently deviant observations often regarded as outliers. "How to Detect and Handle Outliers" reviews several specific definitions of "outlier" and the basic ways in which such observations may arise. The analyst faces three main choices: (1) label the outliers for further investigation, (2) use techniques that are resistant and thus accommodate the presence of outliers, and (3) apply a formal test to identify outliers (and set them aside).

This booklet approaches each of these alternatives from a practical point of view, namely for data that could reasonably have come from a normal

CONTINUED ON PAGE 3

VISION

• Our customers' needs will be continuously anticipated and met.
• Our members will be proud to be part of the Division.
• Our Division's operation will be a model for other organizations.
• As a result, we will be a widely influential authority on scientific approaches to quality and productivity improvement.

MISSION

• Promote Statistical Thinking for quality and productivity improvement.
• Serve ASQC, business and industry, academia, and government as a resource for effective use of statistical methods for quality and productivity improvement.
• Provide a focal point within ASQC for problem-driven development and effective use of new statistical methods.
• Support the growth and development of Division members.  Δ
Dear Readers,

I'm sure you've noticed changes in the appearance of our newsletter. This is an attempt on the part of myself and our printer to provide you, our valuable readers, with the most professional, attractive and up-to-date newsletter possible. We are also trying to reduce eye-fatigue by making more effective use of "white space". If you've been receiving this newsletter for some time you already know that the basic appearance has remained the same since our first professional printing back in September 1983 when Ed Mykytka was Editor - almost ten years ago!

So, "How did we do?" Please let us know how you feel about the newsletter's new look by returning the questionnaire on page 11. It will only take a few moments of your time but your responses will affect our newsletter for years.

You may also have detected the "Printed on Recycled Paper" symbol and statement on the back cover of our newsletter. I am happy to report that we have been using recycled paper for a number of years, not just since the last issue. We just thought you'd like to know it. Does anybody know exactly when we started? If so, please write in and let our readers know.

After talking with some of you on the phone, I'm excited about our newsletter. We have commitments from a number of qualified division members for future articles, minipapers, and basic tools columns. My goal is to provide you with these columns in every issue. You know I can't do it alone. We need more of you to volunteer your writing talents. Come on, write in with letters, opinions, news, and of course the minipapers and basic tools columns.

Interested readers will find criteria for these columns, and "How to Submit A Paper" elsewhere is this issue.

Finally, if you have a CHANGE OF ADDRESS please contact ASQC at 800-248-1946, as instructed elsewhere in this issue. It is NO NEED TO NOTIFY ME as our mailing labels are printed automatically by ASQC headquarters in Milwaukee and mailed directly to our printer.

Dear Jed,

I'd like to take this opportunity to thank you for giving the *News & Record* permission to reprint the article "Spilt Milk" by David Rasmussen in our company newsletter. [ASQC Statistics Division Newsletter, Vol 12, No.4, Summer 1992] I came by the article quite by accident when a member of ASQC national and I were exchanging recent "quality" articles. As I mentioned to you on the phone, our company, 600 employees strong, is implementing a Continuous Improvement program based on the Deming methodology. This article "hits the nail on the head", by addressing the concept of everything is a process, and 94% of all problems are process problems, not people problems. This article was written in the simplest of terms so our employees will be able to understand, and relate to it's meaning. Thanks again for the chance to share this enlightening look at continuous improvement with our employees.

Susan Flannagan
*News & Record*
Greensboro, North Carolina

Dear Mr. Heyes,

I enjoyed the Mini Paper entitled "Some Aspects of Sampling for Control Charts" by Andrew C. Palm. [ASQC Statistics Division Newsletter, Vol 12, No.4, Summer 1992] I hope that you will continue publishing good articles such as these. A point of interest is that the term rational sampling was used by Irving W. Burr in his book Statistical Quality Control Methods, (Dekker, 1976), p. 173. He implied that the term rational sampling is equivalent to rational subgrouping.

Alson C. H. Look
San Jose, California

---

**LETTERS TO THE EDITOR**

Dear Jed,

I'd like to take this opportunity to thank you for giving the *News & Record* permission to reprint the article "Spilt Milk" by David Rasmussen in our company newsletter. I came by the article quite by accident when a member of ASQC national and I were exchanging recent "quality" articles. As I mentioned to you on the phone, our company, 600 employees strong, is implementing a Continuous Improvement program based on the Deming methodology. This article "hits the nail on the head", by addressing the concept of everything is a process, and 94% of all problems are process problems, not people problems. This article was written in the simplest of terms so our employees will be able to understand, and relate to it's meaning. Thanks again for the chance to share this enlightening look at continuous improvement with our employees.

Susan Flannagan
*News & Record*
Greensboro, North Carolina

Dear Mr. Heyes,

I enjoyed the Mini Paper entitled "Some Aspects of Sampling for Control Charts" by Andrew C. Palm. I hope that you will continue publishing good articles such as these. A point of interest is that the term rational sampling was used by Irving W. Burr in his book Statistical Quality Control Methods, (Dekker, 1976), p. 173. He implied that the term rational sampling is equivalent to rational subgrouping.

Alson C. H. Look
San Jose, California

---

**CRIERIAS FOR THE BASIC TOOLS COLUMN**

Purpose: To inform/teach the "quality practitioner" about useful techniques that can easily be understood, applied and explained to others.

Criteria:

1. Application oriented/not theory
2. Non-technical in nature
3. Techniques that can be understood and applied by non-statisticians
4. No longer than 3 pages (8 1/2" x 11" single spaced typewritten)
5. Should be presented in "how to use it" fashion
6. Should include applicable examples

Possible Topics:

- New SPC techniques
- Graphical techniques
- Statistical thinking principles
- "Rehash" established methods

Authors should have a conceptual understanding of the topic and should be willing to answer questions related to the article through the newsletter. Authors do not have to be members of the Statistics Division. All articles will be reviewed.

Examples of some Past Basic Tools Articles:

- "How Control Charts Indicate Change"
- "Cause and Effect Diagram"
- "Brainstorming"
- "The Great Stem and Leaf Plot Debate"
- "Statistical Graphics - A Great Tool-Kit"
- "Thinking Like A Statistician"

---

**NEW "HOW TO" CONTINUED**

distribution. Further sections discuss techniques for data from lognormal and exponential distributions, outliers and influential observations in linear regression, and outliers in factorial experiments. Throughout, numerical examples illustrate issues and approaches. Selected references access to related literature.

Be looking for this helpful new booklet which goes to press soon!
HOW TO SUBMIT A PAPER

Readers interested in submitting a paper to this newsletter should do so in one of three columns: Basic Tools, Mini-Paper or a freelance style paper which is not intended to conform to the criteria of the aforementioned columns.

For all papers we require one hard copy, and (if possible) an IBM compatible 5½" floppy diskette double sided, double density. The preferred file type is an ASCII or Wordperfect file with no formatting: bold, fonts, sizes and so on. (Because we have to manually remove all formatting from files for our printer to process.) All equations and graphics should be on separate pages in the back of the manuscript for scanning purposes, and they should be numbered and referred to in the body of the text. If you are unable to conform to the preferred file formats, don’t give up! Call me and we’ll work something out.

To submit anything besides “papers”, ie; news, letters, comments etc., simply type and send in hard copy and we’ll do the rest. All submissions should be sent to the editor’s address on the back page of the newsletter.

RELATED EVENTS

STATFAIR - A UNIQUE, TRIENNIAL EVENT

The Chicago Chapter of the American Statistical Association holds a Statfair every three years. The last event was held on Wednesday September 23rd, 1992. We just missed it! What is Statfair? Statfair is a gathering of statistical software manufacturers and interested users. During Statfair, attendees learned what programs are available, their capabilities, and how to run them. Technical experts are there to answer all kinds of questions. In addition, if users want to purchase a package, they can do so, on the spot, at a large discount. (Imagine buying a $449 package for just $64.) Software packages on display included: ABSTAT, BBN, GENSTAT, MINITAB, RESAMPLING STATS, SAS, SPC DATALYZER, SPSS, SYSTAT and a host of others. Those in attendance also received a listing of over 400 statistical software packages, with addresses, and phones cross referenced by application - on computer diskettes. We will try to get more advanced information on the next Statfair, but don’t hold your breath. It’s not scheduled until 1995.

CUSTOMER SUPPLIER DIVISION

The CSD Fall Technical Conference (not to be confused with our FTC) will be held on November 5 & 6, 1992 at the Hanalei Hotel, in San Diego, CA. The conference is co-sponsored by the San Diego Section of ASQC. For more information contact Chuck Miller at 800-645-6707.

QUALITY AUDIT DIVISION

The QAD is holding it’s Fall Meeting at the Hyatt Regency, Denver Tech Center in Denver, CO on October 9 & 10, 1992. Their 2nd Annual Quality Audit Conference will be held at the Adam’s Mark Hotel in Charlotte, North Carolina on February 25 & 26, 1993. Look forward to a December’92 brochure containing a full program and information. Early registration can be obtained from ASQC’s Conference Department at 800-248-1946. For more information contact Ken Love at 617-421-7375. The Charlotte Section of ASQC is helping out with this conference.
NORMAL PROBABILITY PLOTS

INTRODUCTION
Plotting sample data points on probability paper can be a powerful tool in the hands of quality practitioners. Probability paper is specially ruled graph paper with one "regular" axis, and another axis ruled in units of percent probability for an appropriate distribution. The probability ruling is based on a cumulative distribution function. Keuffel and Esser graph paper provides the probability ruling on the x axis and the "regular" ruling on the y. On the other hand, my computer software uses the y axis for the probability ruling.

It is not hard to perform probability plots by hand, as we shall soon see. Fortunately, many computer software packages now include probability plot options, thus improving accuracy and speed allowing us to focus on interpretation rather than mere mechanics.

Depending on the resultant shape of the graph, data points give clues as to the nature of the underlying distribution from which the data came. The idea behind these plots is that the closer plotted points are to a straight line, the more likely it is that they come from the distribution modeled by the probability paper. Among the most widely used probability papers is normal probability paper. With normal probability paper if the data plot close to a straight line, they can be considered to be reasonably Gaussian (normally distributed).

In addition to indicating general distribution shape, probability plots also uncover gaps, clusters, outliers, and multiple modes. It is also possible to obtain approximate sample statistics such as the mean and standard deviation, and to predict percentages expected beyond certain limits. A detailed treatment of the subject is given by King (1971), and simple illustrations of the use of normal probability plots are found in, Heyes (1983), Ott (1975), and many other sources.

HOW TO CONSTRUCT A NORMAL PROBABILITY PLOT
1. Obtain the data of interest.
2. Order the values from smallest to largest.
3. For each data point obtain a probability axis location using * Cunnanes formula; location = 100 * [i - 3/8] / (n + 1/4) where i is the ith ordered value in a sample of size n.
4. Plot each point at the appropriate x, y location based on its value for the "regular" axis and the probability location above for the other axis.
5. Draw a straight line roughly dividing the points into two equal halves.

* Other formulas for normal probability location exist including: 100% * [(i - 0.5) / n, and i(100% / n + 1), but they yield less accurate estimates of the standard deviation for samples of 20 or less.

EXAMPLE
A number of electronic measurements were taken on 17 prototype video heads. Based on these measurements Engineering wanted to know what kind of performance to expect from future video heads made by the same company.

Table 1 shows the raw data (RF signal level), the ordered data (ascending), and the y-axis position for each ordered data point based upon; location = 100 * [(i - 3/8) / (n + 1/4)], where i are values 1 through 17 and n is fixed at 17.

<table>
<thead>
<tr>
<th>POSITION (i)</th>
<th>Raw Data RF Signal Level mV</th>
<th>Ordered For x-axis</th>
<th>y -axis Probability Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>540</td>
<td>520</td>
<td>3.62</td>
</tr>
<tr>
<td>2</td>
<td>550</td>
<td>525</td>
<td>9.42</td>
</tr>
<tr>
<td>3</td>
<td>530</td>
<td>530</td>
<td>15.22</td>
</tr>
<tr>
<td>4</td>
<td>560</td>
<td>530</td>
<td>21.01</td>
</tr>
<tr>
<td>5</td>
<td>540</td>
<td>540</td>
<td>26.81</td>
</tr>
<tr>
<td>6</td>
<td>560</td>
<td>540</td>
<td>32.61</td>
</tr>
<tr>
<td>7</td>
<td>580</td>
<td>550</td>
<td>38.41</td>
</tr>
<tr>
<td>8</td>
<td>580</td>
<td>550</td>
<td>44.20</td>
</tr>
<tr>
<td>9</td>
<td>560</td>
<td>560</td>
<td>50.00</td>
</tr>
<tr>
<td>10</td>
<td>525</td>
<td>560</td>
<td>55.80</td>
</tr>
<tr>
<td>11</td>
<td>570</td>
<td>560</td>
<td>61.59</td>
</tr>
<tr>
<td>12</td>
<td>600</td>
<td>560</td>
<td>67.39</td>
</tr>
<tr>
<td>13</td>
<td>590</td>
<td>570</td>
<td>73.19</td>
</tr>
<tr>
<td>14</td>
<td>560</td>
<td>580</td>
<td>78.99</td>
</tr>
<tr>
<td>15</td>
<td>530</td>
<td>580</td>
<td>84.78</td>
</tr>
<tr>
<td>16</td>
<td>550</td>
<td>590</td>
<td>90.58</td>
</tr>
<tr>
<td>17</td>
<td>520</td>
<td>600</td>
<td>96.38</td>
</tr>
</tbody>
</table>

From Table 1 the points would plot as:
1st point at x = 520 and y = 3.6 = 100 * [(1 - 3/8) / (17 + 1/4)]
2nd point at x = 525 and y = 9.4 = 100 * [(2 - 3/8) / (17 + 1/4)],
3rd point at x = 530 and y = 15.2 = 100 * [(3 - 3/8) / (17 + 1/4)].

We continue plotting in this manner until the last point is located at x = 600 and y = 96.3. The resulting probability plot is shown in Figure 1 (see next page).

It is evident from Figure 1 that the parameter Signal Level is close to normal because of how well the points follow the line. These data are so well behaved that we can make even more use of the probability plot. First we can estimate the mean by locating the x value corresponding to the 50% probability (y) level. This value of 555 mV is shown in Figure 2 (see next page).

Also shown in Figure 2 are x values of 530 and 580 mV corresponding to y values of 16 and 84% respectively. The difference between these two x values divided by 2 gives an estimate of the standard deviation. Our estimated standard deviation for video signal level is:
(580-530) / 2 = 25.

ASQC STATISTICS DIVISION NEWSLETTER, VOL. 13, NO. 1
Now suppose engineering wanted to predict what percent would be expected to be beyond proposed specification limits of 510 to 600 mV. To estimate these percent-ages follow each x value up until you reach the best-fit line, then read across to the corresponding y value. Figure 3 demonstrates this practice, resulting in expected percents of approximately 4% below 510 mV and 4% above 600 (96% at or below 600). This gives a total of approximately 8% beyond proposed specifications, clearly an unacceptable rate.

References

KNOW YOUR REGIONAL COUNCILOR

Robert A. Dovich is the new Region 12 Councilor for the Statistics Division of ASQC. Bob is an ASQC Fellow, a CQE, and a CRE. He is Quality Manager for Ingersoll Cutting Tool Company in Rockford, Illinois. An energetic teacher and writer, Bob has published two textbooks, Reliability Statistics, Quality Engineering Statistics, and contributed to the Quality Engineer's Handbook. In addition he has published numerous articles in such journals as; Quality Magazine, Machine and Tool Blue Book, and is contributing Editor for Quality in Manufacturing. He is on the Editorial Review Board for Quality, and Quality Progress magazines. Mr. Dovich holds an M.S. in Engineering Management from the University of Massachusetts (with a concentration in Quality and Reliability), a B.A. from Western Illinois University, and an A.A.S. from Rock Valley College in Quality Assurance Technology. As adjunct faculty member of Rock Valley College since 1981, Bob has provided a number of courses and seminars in the Quality and Reliability fields, and presented technical papers at conferences throughout the U.S.
SELECTING QUALITY BOOKS — AN EDITORIAL

When deciding to purchase a book do you judge it by its cover? Do you look inside catalogs or through the pages of Quality Progress to find materials. Are you impressed with a well known name in the field even though you know little about the author or the contents of the book itself?

We are swimming in a sea of new books nowadays, as authors are springing up everywhere (from nowhere), anxious to publish in the fields of quality, statistics, and management. I'm even planning to try my hand at it. Readers should be careful in their selection of books. What do I recommend? I use three simple options.

One, I ask colleagues and friends about recommended books (just as I do with software, restaurants, CD's and service stations). Two, I seek book reviews. My primary source for book reviews is the Journal of Quality Technology although there are a number of other sources such as Quality Progress and Technometrics. Here are a few excerpts from the July 1992 issue of JQT.

"...Books on Statistical techniques written by nonstatisticians seldom give a reviewer anything about which to rejoice. Unfortunately, the book under review is no exception. The author, an analytical chemist, clearly has much practical experience, and he writes well about this. However, some of his explications of basic statistical concepts are quite flawed...."

-Bert Gunter reviewing a book about simplex experimental designs.

"...In summary, I think this book has many valuable features. For those who wish to learn how to apply these techniques to chemical and similar systems I can think of no better source...."


Besides personal recommendations and book reviews there is a third option. Get a trial copy for review prior to purchasing, or purchase on a 30 day trial basis and read, real fast.

Call for Papers continued from 4

Control, and Tutorial/Case Study sessions. Statistics papers should be at or below the level of Technometrics, while Quality Control and Tutorial/Case Study papers should be at or below the level of Journal of Quality Technology. Please send the title, one-page abstract, and one-page outline of your paper by January 22, 1993 to Ralph St. John, Department of Applied Statistics and Operations Research, Bowling Green State University, Bowling Green, OH 43403. For more information contact any of the program chairs: Christopher Nachtsheim of the Section on Physical and Engineering Sciences (612-624-1077), Steven J. Czarniak of the Chemical and Process Industries Division (716-722-1959), or Ralph St. John of the Statistics Division (419-372-8098).

---

### ASQC Statistics Division 1992-1993 Regional Councilors

<table>
<thead>
<tr>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
<th>Region 4</th>
<th>Region 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edward R. Hansen</td>
<td>Donald J. Baird</td>
<td>Thomas Calvin</td>
<td>Ian B. MacNeill</td>
<td>Joseph R. Troxell</td>
</tr>
<tr>
<td>Gillette Company</td>
<td>Xerox Corporation</td>
<td>IBM Corporation</td>
<td>Director, Statistics Laboratory</td>
<td>Lasalle University</td>
</tr>
<tr>
<td>Prudential Tower Building</td>
<td>800 Phillips Road, 0200-18C</td>
<td>13 Malstrome Road</td>
<td>Engrg. and Math. Sciences Bldg.</td>
<td>20th Street and Olney Avenue</td>
</tr>
<tr>
<td>Boston, MA 02199</td>
<td>Webster, NY 14580</td>
<td>Wappingers Falls, NY 12590</td>
<td>University of Western Ontario</td>
<td>Philadelphia, PA 19141</td>
</tr>
<tr>
<td>(617) 421-7770</td>
<td>(716) 422-8085</td>
<td></td>
<td>(519) 361-6618</td>
<td>(215) 951-1031</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region 6</th>
<th>Region 7</th>
<th>Region 8</th>
<th>Region 9</th>
<th>Region 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marilyn Hwan</td>
<td>Thomas J. Vaden</td>
<td>William M. Bleau</td>
<td>Carlos W. Moreno</td>
<td>Gregory F. Gruska</td>
</tr>
<tr>
<td>LSI Logic</td>
<td>Consultant</td>
<td>(Picker International Inc.)</td>
<td>Ultramax Corporation</td>
<td>(The Third Generation, Inc.)</td>
</tr>
<tr>
<td>3115 Alfred St.</td>
<td>5765 Grand Avenue</td>
<td>(216) 473-2385</td>
<td>1251 Kemper Meadow Dr. #900</td>
<td>(4439 Rolling Pine Drive</td>
</tr>
<tr>
<td>Santa Clara, CA 95050</td>
<td>Riverside, CA 92504</td>
<td>(513) 825-7794</td>
<td>Cincinnati, OH 45215</td>
<td>West Bloomfield, MI 48323</td>
</tr>
<tr>
<td>(408) 433-6562</td>
<td>(714) 382-5525</td>
<td>(513) 825-7794</td>
<td>(513) 825-7794</td>
<td>(313) 363-1634</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region 11</th>
<th>Region 12</th>
<th>Region 13</th>
<th>Region 14</th>
<th>Region 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>George W. Marrah</td>
<td>Robert A. Dovich</td>
<td>Richard D. Schleusener</td>
<td>E.B. Godsey</td>
<td>William H. Woodall</td>
</tr>
<tr>
<td>Department of Mathematics</td>
<td>Ingersoll Cutting Tool Co.</td>
<td>Kodak Colorado Division</td>
<td>303 Ridgebriar Drive</td>
<td>Department of Management</td>
</tr>
<tr>
<td>and Computer Science</td>
<td>505 Fulton Avenue</td>
<td>Plate Manufacturing</td>
<td>Richardson, TX 75080</td>
<td>Science and Statistics</td>
</tr>
<tr>
<td>James Madison University</td>
<td>Rockford, IL 61013</td>
<td>C-60, 1st Floor</td>
<td>(214) 690-1794</td>
<td>University of Alabama</td>
</tr>
<tr>
<td>Harrisburg, VA 22807</td>
<td>(815) 987-6542</td>
<td>Windsor, CO 80551-1610</td>
<td>(503) 686-4500</td>
<td>P.O. Box 870226</td>
</tr>
<tr>
<td>(703) 568-6334</td>
<td></td>
<td>(303) 686-4500</td>
<td></td>
<td>Tuscaloosa, AL 35487-0226</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(205) 348-6085</td>
</tr>
</tbody>
</table>
Introduction

Humans tend to complicate matters unnecessarily. We forget that a great deal can be accomplished using only simple tools. When analyzing data it is usually best to use the simple statistical tools first, leaving the complicated methods until last.

The histogram remains one of the most simple yet powerful tools available. It is just a simple little picture of how the data are distributed, but it is so often neglected. Most statistical and spreadsheet software packages have built-in functions to plot histograms, but it is easy enough to do with a pencil and ruler, and the practice helps develop understanding.

A Histogram Tutorial

The basic approach to constructing a histogram for measurements, and other kinds of continuous data, includes the following steps:

1. Choose the number of cells.
2. Choose the lower boundary of the first cell, and the upper boundary of the last cell.
3. Determine the cell width.
4. Compute the cell boundaries.
5. Compute the cell midpoints.
6. Group the data into the cells and plot the frequency for each cell against the cell midpoints.

The trick in constructing a histogram is to choose the right number of cells. A cell is the "bucket" that catches a group of the data points, and the cell boundaries define the highest and lowest values that fall into the bucket. The cells taken together are like a line of buckets catching all of the data. The number of cells should be few enough so that some data collects in almost every cell but not so few that the data are lumped together. Too many (or too few) cells results in too few (or too many) data points falling in each cell, so that the shape of the frequency distribution cannot be seen. Table 1 gives guidelines for choosing the number of cells. These guidelines are based on experience, and should be altered whenever an adjustment seems reasonable. Some textbooks refer to other guidelines such as "Sturges Rule" or the "Power of Two" rule.

TABLE 1 — Suggested Number of Cells for a Histogram

<table>
<thead>
<tr>
<th>Number of Data Points</th>
<th>Number of Cells (k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under 50</td>
<td>5 to 7</td>
</tr>
<tr>
<td>50 to 100</td>
<td>6 to 10</td>
</tr>
<tr>
<td>100 to 250</td>
<td>7 to 12</td>
</tr>
<tr>
<td>Over 250</td>
<td>10 to 20</td>
</tr>
</tbody>
</table>

The lower boundary of the first cell is the starting point for the histogram. It should be a convenient number that is just less than the smallest value in the data set. The upper boundary of the last cell is likewise the end point of the histogram, and it should be just greater than the largest value in the data set. Some authors recommend choosing cell boundaries that are mid-way between observed values. For instance if the first cell was 20 to 29 and the second 30 to 39, the lower boundary for the first cell might be 19.5 and the upper boundary 29.5 (also the lower boundary for the next cell). Subtracting the lower boundary of the first cell (L) from the upper boundary of the last cell (U), and dividing by the number of cells (k), gives the cell width, as follows:

\[ cw = \frac{U - L}{k} \]

Whenever possible, the cell width should be rounded to a convenient number.

The cell boundaries can be worked out very easily, because each upper boundary is just one cell width greater than the lower boundary of the cell, and each lower boundary is the same as the upper boundary of the previous cell. The cells must all be the same width. It is useful to list the cell boundaries on paper, leaving spaces to add the midpoint of each cell and a summary of the frequencies of the data.

The midpoint is the value of the outcome that characterizes all of the data that falls within the cell. As the name suggests, it is halfway between the lower and upper boundaries of each cell. The midpoint of each cell is also one cell width greater than the midpoint of the previous cell.

Using the cell boundaries, the frequencies can be tabulated by going through the data set one value at a time, and making a mark alongside the cell where the value would fall. Sometimes a rule is needed for handling values that fall on a cell boundary. It does not really matter if such values are put in the cell below the boundary, or in the cell above, as long as the rule is consistently observed.

The histogram is drawn as a bar graph. The cell midpoints are given on the horizontal axis and the frequency counts are plotted on the vertical axis. A cell with no data in it must be shown as an empty cell. If the histogram has too many empty cells to see the shape of the distribution, or if all the data fall in one or two cells, then the histogram should be redrawn with a different number of cells.

A typical histogram is shown in Figure 1. The data used for this figure are contamination measurements collected daily from a process over the course of a year.
The real power of the histogram is realized when you are trying to infer the kind of distribution that best "fits" the data. It takes some practice, and there are some useful tricks that help in the effort.

The best trick is to plot the probability density function of the candidate distribution right on the histogram. But first, this requires a few adjustments. Assuming that a total number of N data points was used to construct a histogram with k cells, the relative frequency occurring in each cell is

\[ f_i = \frac{n_i}{N} \]

where \( n_i \) is the frequency in cell i, and \( i = 1, 2, ..., k \).

This change does not alter the shape of the histogram in any way. It merely changes the scale of the vertical axis of the histogram. The histogram shown in Figure 2 presents the same data that was used to construct Figure 1, but the vertical axis of Figure 2 represents relative frequency instead of frequency. This type of scale change is useful when comparing histograms of data sets of different sizes.

There is another change in the vertical scale that is less intuitive, but at least as useful. We can approximate the probability density function in cell i as

\[ f_i \frac{X(x)}{cw} \]

(Details of how we derived equation 3 are given in the appendix.)

This histogram has relative frequency divided by cell width plotted on the vertical axis, but again the shape of the histogram does not change. The change in vertical scale permits us to plot the probability density function on the same axes with the histogram, using the values of the cell midpoints. When the histogram and the hypothesized distribution are a close fit, the probability density function will pass very near the top-center of every bar of the histogram.

The histogram shown in Figure 3 was constructed using the same data that were used for Figures 1 and 2.

Maximum likelihood parameter estimates for a lognormal distribution were also computed using these data. With the relative frequency divided by cell width as the ordinate, the probability density function was plotted over the histogram, and as one can see, the fit is quite good. We had a high degree of confidence in the lognormal hypothesis even before the goodness-of-fit test was carried out.

Conclusion

Simple tools are really more flexible and more powerful than most people believe. Readers are encouraged to use a variety of exploratory tools including histograms, stem-and-leaf plots, box plots and line graphs before performing more complex statistical tasks. Using the procedures described above is a good beginning to any data analysis task, and the lowly histogram may be the most powerful tool of all.

APPENDIX

Suppose that a process generates observations according to some known distribution. We know that the probability of an observation will fall into an incremental interval is

\[ P(x \leq X \leq x+dx) = f(x) \, dx \]

where \( f(x) \) is the probability density function. With a little algebra we can represent the probability density function as

\[ f(x) = \frac{P(x \leq X \leq x+dx)}{dx} \]

Since the probability in the interval is the relative frequency of occurrence in that interval, we can approximate the probability density function in cell i as

\[ f_i \frac{X(x)}{cw} \]
<table>
<thead>
<tr>
<th>Title</th>
<th>Name</th>
<th>Organization</th>
<th>Address</th>
<th>Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chair</td>
<td>Joseph G. Voelkel</td>
<td>Rochester Inst. of Technology</td>
<td>1 Lomb Memorial Dr., Bldg. 14, P.O. Box 9887, Rochester, NY 14623-0887</td>
<td>(716) 475-2231</td>
</tr>
<tr>
<td>Chair-Elect</td>
<td>Galen C. Gritz</td>
<td>3M Center</td>
<td>220-9W-08, St. Paul, MN 55144-1000</td>
<td>(612) 736-6499</td>
</tr>
<tr>
<td>Treasurer</td>
<td>Richard A. Lewis</td>
<td>Monsanto Co.</td>
<td>800 N. Lindbergh Boulevard, St. Louis, MO 63167</td>
<td>(314) 694-7735</td>
</tr>
<tr>
<td>Secretary</td>
<td>Nancy E. Baxter-Belunis</td>
<td>Thomas J. Lipton, Co.</td>
<td>800 Sylvan Avenue, Englewood Cliffs, NJ 07632</td>
<td>(201) 894-4011</td>
</tr>
<tr>
<td>Past Chair</td>
<td>Conrad A. Fung</td>
<td>Center for Quality and Productivity Improvement, University of Wisconsin</td>
<td>WARF/575, 610 Walnut Street, Madison, WI 53705</td>
<td>(608) 263-2652</td>
</tr>
<tr>
<td>Membership</td>
<td>Robert H. Mitchell</td>
<td>3M Consumer Products Plant</td>
<td>Highway #22 South, Hutchinson, MN 55350</td>
<td>(612) 234-4000, Ext. 1864</td>
</tr>
<tr>
<td>Newsletter Editor</td>
<td>Jed Heyes</td>
<td>Rank Video Services America</td>
<td>555 Huehl Road, Northbrook, IL 60062</td>
<td>(708) 291-4229</td>
</tr>
<tr>
<td>&quot;How To&quot; Booklet Editor</td>
<td>Edward F. Mykytka</td>
<td>Wright-Patterson Air Force Base</td>
<td>Dayton, Ohio 45433-6585</td>
<td>513-235-3362</td>
</tr>
<tr>
<td>&quot;Glossary&quot; Editor</td>
<td>James L. Bossert</td>
<td>Eastman Kodak Company</td>
<td>Building 6, 7th Fl, Kodak Pk. MC23401, Rochester, NY 14652</td>
<td>(716) 477-3837</td>
</tr>
<tr>
<td>Standards Representative</td>
<td>Oswald Willner</td>
<td>Westinghouse Electric Corp.</td>
<td>Betts Atomic Power Lab, P.O. Box 79, West Mifflin, PA 15122</td>
<td>(412) 476-5954</td>
</tr>
<tr>
<td>Examining</td>
<td>Robert L. Perry</td>
<td>Pillsbury Company</td>
<td>330 University Ave. S.E., Minneapolis, MN 55414</td>
<td>(612) 688-0309</td>
</tr>
<tr>
<td>William G. Hunter Award</td>
<td>Peter J. Jacobs</td>
<td>Statistical Consulting</td>
<td>3M, 224-45-19, St. Paul, MN 55144</td>
<td>(612) 733-4257</td>
</tr>
<tr>
<td>Education</td>
<td>Annabeth L. Propst*</td>
<td>Quality Transformation Services</td>
<td>P.O. Box 519, LaGrange, IL 60525-0519</td>
<td>702-482-0339</td>
</tr>
<tr>
<td>1992 Fall Technical Conference - Short Courses</td>
<td>Richard D. Schlesusener</td>
<td>Kodak Colorado Division Plate Manufacturing, C-60, 1st Floor Windsor, CO 80551-1610</td>
<td>(303) 686-4530</td>
<td></td>
</tr>
<tr>
<td>1992 Conference on Applied Statistics - Program</td>
<td>Frank B. Alt</td>
<td>University of Maryland College Park, MD 20742</td>
<td>(301) 454-2231</td>
<td></td>
</tr>
<tr>
<td>1992 Annual Quality Congress - Program</td>
<td>Michael J. Mazu (Alcoa - Warrick Operations)</td>
<td>5311 Moodridge Drive, Newburgh, IN 47630</td>
<td>(812) 853-4893</td>
<td></td>
</tr>
<tr>
<td>ASQC Divisions Liaison Coordinator</td>
<td>Joseph J. Pignatiello, Jr.</td>
<td>Texas A&amp;M University Dept. of Industrial Engrg. Zachary Engineering Center College Station, TX 77843-3131</td>
<td>(409) 862-2081</td>
<td></td>
</tr>
</tbody>
</table>

*Also 1993 Annual Quality Congress - Short Course  **Also Liaison with ASA Section on Quality & Productivity

ASQC STATISTICS DIVISION NEWSLETTER, VOL. 13, NO. 1
READERS SURVEY OF NEWSLETTER CHANGES

MAIL THIS PAGE TO:  
JED HEYES  
Editor, ASQC Statistics Division Newsletter  
c/o Rank Video Services America  
555 Huehl Road  
Northbrook, IL 60062

Please circle one:

1. I'm glad you changed it!    Yes    No
2. It looks better than before. Yes    No
3. Please rate the overall content of this issue:

1=Poor - 5=Excellent

Overall Appearance & Layout 1 2 3 4 5
Color Schemes 1 2 3 4 5
Color of Paper 1 2 3 4 5
Type of Paper 1 2 3 4 5
Readability 1 2 3 4 5
Graphics & Art Work 1 2 3 4 5
Technical Content 1 2 3 4 5
Amount of Technical Information 1 2 3 4 5
Usefulness 1 2 3 4 5
Newsworthiness 1 2 3 4 5

What did you like most?
________________________________________________________________________

What did you like least?
________________________________________________________________________

What further improvements should we make?
________________________________________________________________________

I have been a member of the Stat. Division ________ years

Name:(Optional) ___________________________________________________________
The ASQC Statistics Division Newsletter is published quarterly by the Statistics Division of the American Society for Quality Control.

All communications regarding this publication, EXCLUDING CHANGE OF ADDRESS, should be addressed to:

Jed Heyes, Editor
ASQC Statistics Division Newsletter
Rank Video Services America
555 Huehl Road
Northbrook, IL 60062

Other communications relating to the Statistics Division of ASQC should be addressed to:

Joseph G. Voelkel
Rochester Inst. of Technology Center for Quality and Applied Statistics
1 Lomb Memorial Dr., Bldg. 14
P.O. Box 9887
Rochester, NY 14623-0887
(716) 475-2231

Communications regarding change of address should be sent to ASQC at:
American Society for Quality Control
P.O. Box 3005
Milwaukee, WI 53201-3005

This will change the address for all publications you receive from ASQC including the newsletter. You can also handle this by phone (414) 272-8575 or (800) 248-1946.

UPCOMING NEWSLETTER DEADLINES

<table>
<thead>
<tr>
<th>Issue</th>
<th>Vol. No.</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter '93</td>
<td>13</td>
<td>Nov 20</td>
</tr>
<tr>
<td>Spring '93</td>
<td>13</td>
<td>Feb 19</td>
</tr>
<tr>
<td>Summer '93</td>
<td>13</td>
<td>May 21</td>
</tr>
<tr>
<td>Fall '93</td>
<td>14</td>
<td>Aug 15</td>
</tr>
</tbody>
</table>

Statistics Division planning sessions at A.Q.C. ‘92. From left Roger Hoerl and Beth Propst. Second photo: Ed Hansen and Joe Troxell.