Letter for the ASQ Statistics Digest—February 2019:

Happy New Year from the new 2019 ASQ Statistics Division leadership team!

Some of you may know me, but I thought I would start with a brief introduction: I have been a practicing engineering statistician in the aerospace industry going on (ahem) 20 years now. I have been involved in the Statistics Division since 2008, when my friend, colleague, and new past-chair Daksha Chokshi asked me to help her out by becoming the Secretary for the Division. I followed that with 3 years as Newsletter Editor, then 2 years as Treasurer, including 3 years on the Technical Program Committee for the ASQ/ASA Annual Fall Technical Conference (FTC)—following by this last year as general co-Chair of the 62nd edition, also with Daksha, and held in West Palm Beach, Florida. More about that later.

I would first like to thank Steve Schuelka, outgoing Chair, for his long-standing involvement in the Division. He has been involved since before I was involved and has been a constant and steady source of support over the years, pushing various activities forward, particularly in the areas of Outreach and Voice of the Customer. I greatly appreciate his society-wide perspective and detailed knowledge of ASQ policies and practices stemming from his extensive and active parallel involvement in Section affairs over the years. Steve has been a tenacious advocate for the Division this year as ASQ goes through its internal reorganization/transformation. More on that later as well.

Moving on to leadership changes, I would like to thank Alex Gutman for his service this year as Treasurer, a critical role in making sure the Division runs smoothly. He will be replaced by Joyce Crum, a new volunteer this year. Jennifer Williams is moving on from Secretary, transitioning to a new Vice Chair role for Community Involvement. (This is a new title for this role, intending to capture what had previously been Member Development and a few other roles under a broader umbrella term.) Shoshana Bokelman is a new volunteer last year stepping in to a larger role as our new official Secretary. She has been taking care of announcements as part of our Content committee. Thanks to everyone for their service this past year and to our new contributors who are taking on new roles and responsibilities.

Continued on page 3
The Statistics Division was formed in 1979 and today it consists of both statisticians and others who practice statistics as part of their profession. The division has a rich history, with many thought leaders in the field contributing their time to develop materials, serve as members of the leadership council, or both. Would you like to be a part of the Statistics Divisions’ continuing history? Feel free to contact chair@asqstatdiv.org for information or to see what opportunities are available. No statistical knowledge is required, but a passion for statistics is expected.

**Vision**

The ASQ Statistics Division promotes innovation and excellence in the application and evolution of statistics to improve quality and performance.

**Mission**

The ASQ Statistics Division supports members in fulfilling their professional needs and aspirations in the application of statistics and development of techniques to improve quality and performance.

**Strategies**

1. **Address core educational needs of members**
   - Assess member needs
   - Develop a “base-level knowledge of statistics” curriculum

2. **Build community and increase awareness by using diverse and effective communications**
   - Webinars
   - Newsletters
   - Body of Knowledge
   - Web site
   - Blog
   - Social Media (LinkedIn)

3. **Foster leadership opportunities throughout our membership and recognize leaders**
   - Advertise leadership opportunities/positions
   - Invitations to participate in upcoming activities
   - Student grants and scholarships

4. **Establish and Leverage Alliances**
   - ASQ Sections and other Divisions
   - Non-ASQ (e.g. ASA)
   - CQE Certification

**Updated October 19, 2013**
Message From the Chair  Continued

I am also happy to welcome Amy Ruiz as our incoming Chair-Elect. Amy started out as Webinar Coordinator, also helped out with Social Media, and moved up to Vice Chair of Content. She has been quietly making things happen in the background for several years now. Some of you may have met her at various outreach events, as she has often volunteered to help out at the Statistics Division table or booth, for example at the World Conference for Quality & Improvement (WCQI). Thank you for joining me, Amy, I’m sure this will be an adventure (!).

Other key leads include Peter Parker as Vice Chair of Awards, continuing on. Gary Gehring, another volunteer who has taken on numerous roles over the years, will become Vice Chair of Content, replacing Amy. Brian Sersion is another long-time volunteer continuing on with several roles. I would also like to welcome several other new faces to our group: Harry Rowe, Geoff Farmer, and Michael Kirchner, picking up several committee support roles. Thanks for your involvement and interest in helping out this year!

One additional major change I would be remiss not to mention (although I am certain it will be mentioned elsewhere here) is the upcoming retirement/transition of Matthew Barsalou, our intrepid Statistics Digest editor, who is beginning to transfer his editorial responsibilities to incoming editor Didier Greenleaf, Jr. I have to thank Matt for his work the past four (!) years with the Statistics Digest. Our leadership team at the time had the initial vision to convert the original Statistics Division Newsletter to focus more on the technical content that was more interesting to our members. Matt implemented the entire thing, identifying both columnists and potential sources for feature articles, as well as handling all the publication details with the support of Alex Gutman and Kurtis Shuler as technical reviewers. The Division received a Silver PAR Award for Innovation for this in 2016 and we extend great thanks to Matt for his dedication and diligence turning the Statistics Digest into a reality. And welcome Didier! We are happy to have you on board.

The Statistics Division as one of the four sponsoring organizations is heavily involved in the planning for this event, so the FTC is always the capstone of our year. We use it to meet members, share technical content through our invited session, and present a slew of annual awards. The theme of the conference was “Statistics and Quality: Riding the Big Data Wave,” and there were a plethora of sessions showing a variety of tools and techniques for various case studies and application areas. I would like to thank everyone for their involvement in pulling off this successful event. I won’t rehash everything here—if you are interested in the gory details there was a summary published in the December 2018 issue of the AmStat News (pgs. 34–35) http://magazine.amstat.org/wp-content/uploads/2018/11/AMSTATNEWS_DECEMBER2018.pdf; however, I will mention the Youden Memorial Address “Leveraging Industrial Statistics for the Data Revolution”, given by Dr. L. Allison Jones-Farmer of the Miami University of Ohio, which explored how modern data technologies are transforming our roles and responsibilities as statisticians. A summary of this address is included in this issue of the Digest, with a more complete version to be published in the journal Quality Engineering.

The 2019 FTC will be held at the National Institute for Standards and Technology (NIST) in Gaithersburg, Maryland, just outside Washington D.C., and chaired by Statistics Division Past-Chair Adam Pintar, a long-time NIST denizen. If you are interested in being a part of this always worthwhile event, the deadline for contributed abstracts is rapidly approaching (Feb 28th!). The Call for Papers is available here: http://www.falltechnicalconference.org/wp-content/uploads/2019/2019_Call_for_Papers.pdf. Online abstract submission (piloted last year) is available on the conference website http://www.falltechnicalconference.org/. Additional proposals can be made to members of the Technical Program and Short Course Committees. Travel and other information will be posted when available.
Message From the Chair  Continued

Please do make FTC plans early this year! The NIST facility is a government facility with limited access and all participants must be pre-registered. To accommodate government restrictions, registration will close earlier than usual and there will be NO on-site registration available to last-minute participants.

Some of you may be aware of the changes going on with the ASQ Transformation. This will present significant challenges to us, but also many opportunities. Some operational challenges we have struggled with over the past 10 years (or more) should be alleviated. For example, you should see a big change in the ASQ Statistics Division website by mid-year as ASQ upgrades their platform and moves to a more flexible system. We have been well aware that the current website has needed upgrades but have been waiting on developments at the ASQ level to begin major renovations. This project has been imminent for quite some time but the system is now in place and we are scheduled for the second wave of implementation. The plan is to begin migrating content behind the scenes early this year, do an initial rollout and then continue to populate the remainder of the content in batches during the course of the year. We are really looking forward to this change and in seeing what we can do with the new system!

The new myASQ platform has also just been rolled out, which you are invited to join. This provides an opportunity to connect with people in various communities and networks across ASQ. You are also invited to join the Statistics Division’s group on LinkedIn if you are interested in networking with members there. The Statistics Division sends out a monthly E-Zine by email which contains the latest information on upcoming events and deadlines, e.g., for conferences and scholarships. The Statistics Digest is published 3 times a year (February, June, October). You do need to be signed up with ASQ to accept email from the Statistics Division to receive any communications from us, so if you are reading this from the website and wondering why you don’t get anything, it is because of your settings. ASQ does try to limit the amount of communications they send to avoid drowning everyone in email, but you can tailor your settings to the level of interaction that is right for you.

Other behind-the-scenes changes include improvements to the financial management system, which as a former Treasurer and current Chair make me very happy, but won’t impact anyone else unless you are considering becoming involved as an ASQ member leader. This should make the jobs of member leaders easier in the future, so certainly a plus. Other changes have to do with the restructuring of the internal support organization and development is still ongoing. Again, these changes don’t impact the general population, but you may start to encounter some terminology differences: you will see the term Geographic Community replacing Section in some areas, and the term Technical Community is becoming more widespread as an umbrella term for Divisions, Forums, and Interest Groups. The Statistics Division is staying the Statistics Division though—there was some thought given to changing names, but that was quashed last year. The Statistics Division has a long heritage as the Statistics Division and there is no compelling reason to change it now.

In fact, this year the Statistics Division will be celebrating the 40th anniversary of its founding in 1979. We will be recognizing this milestone as we go through the year and will be recognizing the event at this year’s FTC in Gaithersburg, which is where the very first meeting of the Statistics Division took place in 1979. A little history for those interested in such things: the Statistics Division was formed from the Chemical and Process Industries Division (CPID), which is where most people in ASQ (then ASQC) with a statistics interest had congregated prior to that point. The first meeting was chaired by William G. “Bill” Hunter, the namesake of the Hunter Award, one of several awards given out by the Statistics Division at the FTC each year. A list of the topics discussed at this meeting is on our website under Division history, with a note that the first...
Message From the Chair  Continued

The newsletter contained an article from Phil Crosby, ASQC president at the time. In his article he commented on the critical role of statistics in supporting the entire quality profession.

While the nature of the statistics profession has certainly evolved in the 40 years since he wrote that, statistics still plays a foundational role in quality methods, applications, and management. It continues to evolve and grow with changes in technology that allow closer integration between statistics and quality, or and other fields. This presents challenges and requires the development of new skills and tools, but also provides new opportunities to both statisticians and those they support.

Every 4–6 years the Statistics Division revisits its Vision, Mission, and Strategies (Ref. page 2 of the Digest) at a Long-Term Strategy planning meeting. The most recent of these was held in 2013, so we are certainly due for another one, especially given the rapid pace of technology change and impacts in our arena. So, this is likely on the agenda as well for early 2020, pending input from ASQ, which is itself is in the process of reassessing their own goals and strategies.

As a statistician I am happy to see such greater access to an abundance of data, as well as wider recognition of its criticality (data-driven decision-making!), but it is important to remember that big data isn’t always better data. Context is critical. I could take a million measurements of my height every day but they wouldn’t serve any purpose. It is important to remember why we are collecting data and clearly understand the objectives.

Another recently published example of the importance of context was a study recently published in the British Medical Journal (BMJ) in their December 2018 issue “Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial” (full article available at https://www.bmj.com/content/363/bmj.k5094, summarized here by NPR “Researchers Show Parachutes Don’t Work, But There’s A Catch” https://www.npr.org/sections/health-shots/2018/12/22/679083038/researchers-show-parachutes-dont-work-but-there-s-a-catch). This tongue-in-cheek study on a classic thought example showed no difference in benefit for the intervention (the parachute) vs. the control (the backpack); however the test was conducted only using people jumping off the airplane two feet above the ground—due to the inability to enroll people in more realistic scenarios. The authors do warn against extrapolation. The provides a humorous example of what we refer to in aerospace as “Test Like You Fly” (TLYF) as a way to make sure that we do everything possible to ensure that we test in realistic or at least reasonably representative conditions.

I always enjoy seeing statistics in the news.

A few other updates and announcements:

- The Data Science and Analytics Technical Community, organized by Past Chair Herb McGrath, is getting off the ground. For information contact Rnmcgra@bgsu.edu.
- Webinars will be announced via E-Zine. If you are interested in giving a webinar on a topic of interest to our member community, please submit to Harry Rowe, our Webinar Coordinator, at webinars@asqstatdiv.org.
- The Ellis R. Ott Scholarship fund awards two $7500 scholarships every year to graduate students in the area of applied statistics and/or quality management. Information about this prestigious award is available at the Stat Div website. The deadline for applications is April 1st.
Message From the Chair  Continued

- **WCQI**—The Statistics Division will have a booth at the expo at the World Conference on Quality & Improvement (WCQI) to be held May 20–22 in Fort Worth, TX [https://asq.org/conferences/wcqi]. If you are there, stop by and see us!

- **FTC Grants**—Several travel and registration grants are available for students and early career individuals to attend the FTC. Application required. Announcements with updated deadlines will be posted as soon as possible due to the accelerated FTC schedule and registration deadline this year.

- If you are interested in becoming an **ASQ Fellow**, the Statistics Division sponsors a limited number of nominations every year. Minimum eligibility requirements are 5 years as an ASQ Senior member and 15 full years of active experience in quality-related professions, and demonstration of expertise across a broad range of categories. The application process is based on a points system and requires detailed documentation and supporting evidence that the criteria are met. For more information contact our Examining Chair, Daksha Chokshi, at examining@asqstatdiv.org. Specific requirements are available at ASQ.org.

I would also like to put in a plug for the Statistics Division’s most recent publication, **Statistical Roundtables: Insights and Best Practices**, published 2016, and for which we also received a Silver PAR Award for Innovation. Edited by frequent contributor and Past-Chair Christine Anderson-Cook and Lu Lu, this book includes a comprehensive collection representing 10 years of popular “Statistics Roundtable” articles, a regular feature in Quality Progress, now called “Statistics Spotlight”. A portion of the proceeds goes to the Ott Scholarship fund. This book is available at ASQ Press: [https://asq.org/quality-press/display-item?item=H1508](https://asq.org/quality-press/display-item?item=H1508).

On a final note, if you are interested in becoming involved with the Statistics Division, please feel free to contact me at chair@asqstatdiv.org. We have a lot of ongoing activities and are looking for people available to help out. Also we are always interested in new ideas and suggestions for ways that the Statistics Division can better provide new resources to its members.

Best wishes for 2019!
Message from the Past Chair

Steven Schuelka

It has been an honor to serve as the Chair of the Statistics Division this past year. As someone with both a BS and MS in Statistics, these past 12 years being on the leadership team have been fulfilling, both personally and professionally. I joined the division soon after its inception in 1979 and have had the pleasure of interacting with some of the giants and thought leaders in our profession since then. It has been a pleasure meeting George Box, Stuart Hunter, Ken Stephens, and others whose books I studied in college. Now there is a new generation of experts who I’ve met at WCQI and at FTC or read their contributions in the Stat Digest.

The division has seen many changes over the years and 2018 was no exception. The transformation journey for ASQ began in earnest last year and our division began to adapt to these changes. As we enter 2019, one thing that hasn’t changed is us providing statistical knowledge to our members. We will still be offering webinars, peer-reviewed articles, speaking opportunities, etc. into the future. As the Statistics Division enters its 40th year of providing exceptional content, training and conference speakers, I invite you to continue being part of one of the most awarded divisions in the Society.

Steven Schuelka

Editor’s Corner

Matt Barsalou

Welcome to the February 2019 issue of Statistics Digest. There are many changes being announced in this issue. First, our 2018 Division Chair, Steven Schuelka, is now the Past-Chair and Mindy Hotchkiss will be taking over as the 2019 Chair. Long-time readers may recall Mindy the one who trained me for the position of Newsletter editor.

This issue concludes a summary of Dr. K. Allison Jones-Farmer’s Youden Address Leveraging Industrial Statistics in the Data Revolution. A more detailed version will be published in Quality Engineering, volume 31, issue number 2. This issue’s Mini-paper is Transforming Individuals Control Chart Data by Forrest Breyfoglle III and the feature is A Brief Discussion on Conjugate Beta Priors by Harish Jose.

This issue’s Standards InSide-Out column will be Mark E Johnson’s final column and I would like to sincerely thank Mark for his contributions. We also have Donald J. Wheeler explain SPC and Jim Frost discussing hypothesis testing.

This will be my final issue as editor and the past four years have truly been a privilege for me. I am pleased to introduce Didier Greenleaf Jr, who will be taking over as the editor of Statistics Digest starting with the June 2019 issue. Didier graduated with honors from the Pennsylvania State University and holds a Bachelor of Science in Chemical engineering. He has over 5 years of experience in pharmaceutical manufacturing and has held various roles in quality systems and engineering. He currently works in quality validation where he applies statistical techniques on a daily basis.
Youden Address: Leveraging Industrial Statistics in the Data Revolution - Summary

L. Allison Jones-Farmer

Department of Information Systems and Analytics, Miami University, Oxford, Ohio

It was one of the great honors of my career to be invited to give this lecture. I want to thank the Youden Address nominating committee for inviting me and the FTC organizers for hosting me. This is a brief summary of my W.J. Youden Memorial Lecture presented at the 2018 Fall Technical Conference in West Palm Beach, Florida. A more comprehensive paper based on this lecture is being prepared for publication in a future issue of Quality Engineering.

We are in the midst of a Data Revolution that is transforming our economy. I believe that this revolution is as large and profound as other major economic shifts including the Industrial Revolution. Many call our time the Information Age, the Computer Age, the Digital Age, or the age of “Big Data”. The Information Age started during the latter quarter of the 20th century, following the Machine Age, which began in the 1800’s. The Machine Age represented a major turning point in our history, as every facet of daily life was affected. It was during this timeframe that we saw the rise of our field, Industrial Statistics.

Industrial Statistics, the field, emerged to address problems and questions that arose from the new types of data that came about as production moved off the farm and into the factory. The early works of Shewhart, Box, Hunter, Duncan, and Jack Youden were at the forefront of this movement. Youden was a Ph.D. trained chemist who lived from 1900–1971 and was self-motivated and self-taught in statistics. He was particularly skilled at adapting statistical methodologies to solve a particular problem. Jack Youden brought tremendous value to the field of industrial statistics and embodied five important characteristics that are evident through his work. Youden was (1) a self-taught, life-long learner; (2) an innovative problem solver; (3) a clear communicator to audiences at all levels; (4) an open sharer of knowledge; who (5) emphasized good study design. These characteristics embody the ideal traits for an Industrial Statistician.

The field of Industrial Statistics emerged to address the new problems that arose during the Machine age, and the field of Data Science is emerging to solve the new problems that are arising during the Information Age. The Information Age began when the manipulation and storage of data by computers and networks became easily accessible. Before our eyes, we are watching as the businesses and industries that most efficiently transform data into meaningful information are emerging as the leaders. Like the success of the Machine Age rested on the ability to leverage machinery and skilled workers in meaningful ways, the success of this time will be defined by how we leverage data.

The ability to build new technology is a key differentiator for companies in the Information Age. A second key differentiator is owning proprietary data, and third key differentiator is being able to leverage this data for effective decision
making. These three differentiators, the ability to build technology, the owning of proprietary data, and the ability to leverage this data for decision-making distinguish the companies that are thriving in the Information Age from those that are stagnating. As companies try to gain the ability to differentiate themselves, they seek employees who have the skills to move them forward in this process. This has created a new profession: the Data Scientist.

There are numerous definitions for the role of a Data Scientist. SAS defines Data Scientists as “a new breed of analytical data experts who have the technical skills to solve complex problems—such as the curiosity to explore what problems to be solved. They’re part mathematician, part computer scientist and part trend-spotter.” (SAS, 2018) Davenport and Patil (2012) states that a “Data Scientists’ most basic, universal skill is the ability to write code.” He goes on to say “. . . the dominant trait among Data Scientists is an intense curiosity—a desire to go beneath the surface of a problem, find the questions at its heart, and distill them into a very clear set of hypotheses that can be tested.”

Cegielski and Jones-Farmer (2016) conducted a mixed-method study of industry professionals to determine the skills necessary for entry-level Analytics/Data Science positions. They found that six of the top ten skills included software and coding skills and the remainder of skills related to analytical skills such as problem framing as well as professional skills such as being an independent learner. Interestingly, these skills are similar to primary job domains tested on the Certified Analytics Professional (CAP) Exam (Informs 2018) and are consistent with the definition given by Davenport and Patil (2012).

As industrial statisticians, we embrace the skills that are identified in the CAP job domains and Data Science/Analytics job-task surveys. We frame problems both within a domain and methodologically. We acquire and process data. We develop models, communicate and deploy results, and we manage models throughout the lifecycle to make sure they are still valid. However, there are critical skills that we, as Industrial Statisticians, possess that are necessary in Data Science, but are lacking in the job task analyses for Data Science and Analytics professionals.

Study design and the understanding inferential validity are critical, yet we rarely see these skills discussed in the role of a Data Scientist or Analytics professional. In addition to up front study design, inferential validity, or the understanding of the different conclusions that can be drawn from confirmatory or explanatory or predictive models, and the near impossibility of inferring causation from secondary data have been largely ignored with the emergence of Data Science as a field.

As statisticians, we can use our skills to transform the practice of Data Science. We can embrace the ideal characteristics of an Industrial Statistician that we learned from Jack Youden’s work and be (1) self-taught, life-long learners; (2) innovative problem solvers; (3) good communicators; and (4) open sharers of knowledge who (5) emphasize good study design.

One of the key differentiators of companies in the Information Age is that they build their own technology. Industrial statisticians in industry can contribute to teams at the algorithm development stage to ensure that the technology and algorithms are developed using scientific methods and sound statistical practice. In addition, industrial statisticians in both industry and academia can develop statistically valid methodologies that are scalable for large problems, fast, and deployable. A second key differentiator of companies in the Information Age is that they own proprietary data. As a profession, we can work to make sure that the Data Scientists understand concepts and importance of data quality, scientific sampling, missing data, and systematic bias. All of us need to educate others on the importance of the four dimensions of data quality including completeness, consistency, accuracy, and timeliness to the validity of their analysis and conclusions. The third key differentiator of technology giants is their ability to leverage data for decision-making.
There is no substitute for the advanced education and experience that we, as a profession, possess in study design and inferential validity. It is time for us, as a profession, to leverage these skills and to help transform the Data Science field.

References


Operational Excellence (OE), according to Wikipedia, is to create a system for sustainable improvement of key performance metrics. To achieve this Wikipedia OE directive, measurements need to be tracked from a process-output point of view. These metrics also need to be structurally linked to the processes that created them. An **Operational Excellence System for achieving these OE objectives** is Integrated Enterprise Excellence.

**Operational Excellence System with 30,000-foot-level Measurement Report-outs**

The IEE 30,000-foot-level reporting format provides a structured methodology for examining a process output response from a high-level point of view. The creation of these metrics involves a two-step process:

1. Determine process stability (from a 30,000-foot-level perspective)
2. Create a capability/performance statement that describes how the process is not only currently performing but expected to perform in the future.

If the stable-process-prediction-statement is unsatisfactory, this metric improvement needs pulls for a process enhancement effort; e.g., Lean Six Sigma process improvement project.

A 30,000-foot-level individuals chart is used to assess process stability; however, **this chart is not robust to data non-normality.** Because of this chart characteristic, **data may need to be transformed so that false special-cause signals are not created.** A bounded-physical-situation example is when the time to execute a process cannot be less than zero. When using a log or other transformation, the data alteration needs to make physical sense. For this situation where negative numbers are not possible, a log transformation can be an appropriate physical choice.

The following describes the creation of an individuals control chart as the first step to the formulation of a 30,000-foot-level metric report-out. For this flatness of a part illustration, there is a lower bound of zero.

**Individuals Control Chart for Non-normal Data**

Transforming individuals control chart data should be an important consideration when providing control charting of individuals data, since an individuals control chart is not robust to non-normality. A data transformation, which makes physical sense, may be necessary for an individuals control chart (i.e., XmR or ImR chart) to adequately assess process stability. The implication of this is that when processes are assessed at a 30,000-foot-level an erroneous decision could be made relative to one of following three considerations (see below), if an appropriate transformation is not made.
Transforming Individuals Control Chart Data

Individuals Control Chart Statistical Tracking and Reporting

Three considerations for statistical tracking and reporting of transactional and manufacturing process outputs are:

1. Is the process unstable or did something out of the ordinary occur, which requires action or no action?
2. Is the process stable and meeting internal and external customer needs? If so, no action is required.
3. Is the process stable but does not meet internal and external customer needs? If so, process improvement efforts are needed.

To enhance the process of selecting the most appropriate action or non-action from the three listed reasons, a 30,000-foot-level approach will be used which will include a method to describe process capability/performance reporting in terms that are easy to understand and visualize.

Transforming Individuals Control Chart Data Application Example

Let's consider a hypothetical application. A panel's flatness, which historically had a 0.100 inch upper specification limit, was reduced by the customer to 0.035 inches. Consider, for purpose of illustration, that the customer considered a manufacturing nonconformance rate above 1 percent of the specification limit (i.e., 0.100 or 0.035 inches) to be unsatisfactory.

Physical limitations are that flatness measurements cannot go below zero, and experience has shown that common-cause variability for this type of situation often follows a log-normal distribution; i.e., transforming individuals control chart data may later be necessary.

The person who was analyzing the data wanted to examine the process at a 30,000-foot-level view to determine how well the shipped parts met customers' needs. She thought that there might be differences between production machines, shifts of the day, material lot-to-lot thickness, and several other input variables. Because she wanted typical variability of these inputs as a source of common-cause variability relative to the overall dimensional requirement, she chose to use an individuals control chart that had a daily subgrouping interval. She chose to track the flatness of one randomly-selected, daily-shipped product during the last several years that the product had been produced.

She understood that a log-normal distribution might not be a perfect fit for a 30,000-foot-level assessment, since a multimodal distribution could be present if there were a significant difference between machines, etc. However, these issues could be checked out later since the log-normal distribution might be close enough for this customer-product-receipt point of view.

To model this situation, consider that 1,000 points were randomly generated from a log-normal distribution with a location parameter of two, a scale parameter of one, and a threshold of zero (i.e., log normal 2.0, 1.0, 0). The distribution from which these samples were drawn is shown in Figure 1. A normal probability plot of the 1,000 sample data points is shown in figure 2.

From Figure 2, we statistically reject the null hypothesis of normality technically, because of the low p-value, and physically, since the normal probability plotted data do not follow a straight line. This is also logically consistent with the problem setting, where we do not expect a normal distribution for the output of
Transforming Individuals Control Chart Data

From Figure 3, we fail to statistically reject the null hypothesis of the data being from a log-normal distribution, since the p-value is not below our criteria of 0.05, and physically, since the log-normal probability plotted data tend to follow a straight line. Hence, it is reasonable to model the distribution of this variable as log normal.

If the individuals control chart is robust to data non-normality, an individuals control chart of the randomly generated log-normal data should be in statistical control. In the most basic sense, using the simplest run rule (a point is “out of control” when it is beyond the control limits), we would expect such
Transforming Individuals Control Chart Data

Data to give a false alarm on the average around three times out of 1,000 points. Further, we would expect false alarms below the lower control limit to be equally likely to occur, as would false alarms above the upper control limit.

Figure 4 shows an individuals control chart of the randomly-generated data.

The individuals control chart in Figure 4 shows many out-of-control points beyond the upper control limit. In addition, the individuals control chart shows a physical lower boundary of zero for the data, which is well within the lower control limit of -22.9. If no transformation is needed when plotting non-normal data in a control chart, then we would expect to see a random scatter pattern within the control limits, which is not prevalent in the individuals control chart.
Transforming Individuals Control Chart Data

Figure 5 shows a control chart using a Box-Cox transformation with a lambda value of zero, the appropriate transformation for log-normally distributed data. This control chart is much better behaved than the control chart in Figure 4. Almost all 1,000 points in this individuals control chart are in statistical control. The number of false alarms is consistent with the design and definition of the individuals control chart control limits.

![I Chart of Random1000 Using Box-Cox Transformation With Lambda = 0.00](image)

**Figure 5: Individuals Control Chart with a Box-Cox Transformation Lambda Value of Zero**

Determining Actions to Take

Previously three decision-making action options were described, where the first option was:

**Consideration 1**: Is the process unstable or did something out of the ordinary occur, which requires action or no action?

For organizations that did not consider transforming data to address this question, as illustrated in Figure 4, many investigations would need to be made where common-cause variability was being reacted to as though it were special cause. This can lead to much organizational firefighting and frustration, especially when considered on a plant-wide or corporate basis with other control chart metrics.

If data are not from a normal distribution, an individuals control chart can generate false signals, leading to unnecessary tampering with the process. For organizations that did consider transforming data to address this question, as illustrated in Figure 5, there is no overreaction to common-cause variability as though it were special cause.

For the transformed data analysis, let’s next address the other questions:

**Consideration 2**: Is the process stable and meeting internal and external customer needs? If so, no action is required.

**Consideration 3**: Is the process stable but does not meet internal and external customer needs? If so, process improvement efforts are needed.

When a process has a recent region of stability, we can make a statement not only about how the process has performed in the stable region but also about the future, assuming nothing will change...
Transforming Individuals Control Chart Data

in the future either positively or negatively relative to the process inputs or the process itself. However, to do this, we need to have a distribution that adequately fits the data from which this estimate is to be made.

For the previous specification limit of 0.100 in., Figure 6 shows a good distribution fit and best-estimate process capability/performance or nonconformance estimate of 0.5 percent (100.0–99.5). For this situation, we would respond positively to item number two since the percent nonconformance is below 1 percent; i.e., we determined that the process is stable and meeting internal and external customer needs of a less than 1-percent nonconformance rate; hence, no action is required.

However, from Figure 6 we also note that we expect the nonconformance rate to increase to about 6.3 percent (100–93.7) with the new specification limit of 0.35 in. Because of this, we would now respond positively to item number three, since the nonconformance percentage is above the 1-percent criterion. That is, we determined that the process is stable but does not meet internal and external customer needs; hence, process improvement efforts are needed. This metric improvement need would be “pulling” for the creation of an improvement project.

It is important to present the results from this analysis in a format that is easy to understand, such as the 30,000-foot-level metric approach described in Figure 7. With this approach, we demonstrate process predictability using a control chart in the left corner of the report-out and then use, when appropriate, a probability plot to describe graphically the variability of the continuous-response process with its demonstrated predictability statement. With 30,000-foot-level reporting, a statement at the bottom of the plots nets out how the process is performing relative to the new specification requirement of 0.035; i.e., a predictable process with an approximate nonconformance rate of 6.3 percent.

A Lean Six Sigma improvement project could be executed to determine what should be done differently in the process so that the new customer requirements are met. Within this project it might be determined in the analyze phase that there is a statistically significant difference in production machines that now needs to be addressed because of the tightened 0.035 tolerance. This statistical difference between machines was probably also prevalent before the new specification requirement; however, this difference was not of practical importance since the customer requirement of 0.100 was being met at the specified customer frequency level of a less than 1-percent nonconformance rate.
Upon satisfactory completion of an improvement project, the 30,000-foot-level control chart would need to shift to a new level of stability that had a process capability/performance metric that is satisfactory relative to a customer 1 percent maximum nonconformance criterion.

Generalized Statistical Assessment

The specific distribution used in the prior example, log normal (2.0, 1.0, 0), has an average run length (ARL) for false rule-one errors of 28 points. The single sample used showed 33 out-of-control points, close to the estimated value of 28. If we consider a less-skewed log-normal distribution, log normal (4, 0.25, 0), the ARL for false rule-one errors drops to 101. Note that a normal distribution will have a higher false rule-one error ARL.

The log-normal (4, 0.25, 0) distribution passes a normality test over half the time with samples of 50 points. In one simulation, a majority, 75 percent, of the false rule-one errors occurred on the samples that tested as non-normal. This result reinforces the conclusion that normality or a near-normal distribution is required for a reasonable use of an individuals chart or a significantly higher false rule-one error rate will occur.

Summary

The output of a process is a function of its steps and input variables. Doesn’t it seem logical to expect some level of natural variability from input variables and the execution of process steps? If we agree to this assumption, shouldn’t we expect a large percentage of process output variability to have a natural state of fluctuation; that is, to be stable?

To me this statement is true for many transactional and manufacturing processes, with the exception of things like naturally auto-correlated data situations such as the stock market. However, with traditional control charting methods, it is often concluded that the process is not stable even when logic tells us that we should expect stability.

Why is there this disconnection between our belief and what traditional control charts tell us? The reason is that often underlying control-chart-creation assumptions are not valid in the real world. Figures 4 and 5 illustrate one of these points where an appropriate data transformation is not made.
Transforming Individuals Control Chart Data

The questions that should be addressed when tracking a process can be expressed as determining which actions or non-actions are most appropriate.

1. Is the process unstable or did something out of the ordinary occur, which requires action or no action?
2. Is the process stable and meeting internal and external customer needs? If so, no action is required.
3. Is the process stable but does not meet internal and external customer needs? If so, process improvement efforts are needed.

This article described why appropriate transformations from a physical point of view need to be a part of this decision-making process.

The statement at the bottom of Figure 7 describes the state of the examined process in terms that everyone can understand; i.e., the process is predictable with an estimate 6.3-percent nonconformance rate.

Benefiting from Application of 30,000-foot-level Predictive Performance Metric throughout Organizations

An organization gains much when this form of scorecard-value-chain reporting is used throughout its enterprise and is part of its decision-making process and improvement project selection. A system for accomplishing this objective is Integrated Enterprise Excellence (IEE).

Eight application examples of converting traditional dashboards and scorecards to 30,000-foot-level reporting is available in the article Predictive Performance Reporting, where more insight is gained from this reporting format.

About the Author

Forrest Breyfogle is an ASQ Fellow and the CEO of Smarter Solutions, Inc. He has authored or co-authored over a dozen books. His five-book set, Integrated Enterprise Excellence provides radical management advancements in the utilization and integration of scorecards, strategic planning, and process improvement.

Mr. Breyfogle was named Quality Professional of the Year for 2011 by Quality Magazine and in 2012 was awarded alumni of the year by Missouri University of Science and Technology. He also received the prestigious Crosby Medal from ASQ in 2004 for an earlier book, Implementing Six Sigma and was presented the Leadership Award at the 2013 Lean & Six Sigma World Conference. He can be reached at Forrest@SmarterSolutions.com or 512-918-0280.
All charts for count-based data are charts for individual values. Regardless of whether we are working with a count or a rate, we obtain one value per time period and we will want to plot a point every time we get a value. This is why four specialty charts for count-based data had been developed before a general approach for charting individual values was discovered. These four charts are the p-chart, the np-chart, the c-chart, and the u-chart. The question addressed in this column is when to use these and other specialty charts with your count-based data.

The first of these specialty charts, the p-chart, was created by Walter Shewhart in 1924. At that time the idea of using the two-point moving range to measure the dispersion of a set of individual values had not yet occurred. (John von Neumann et. al. would not publish their paper on the use of successive differences until 1941.) So the problem Shewhart faced was how to create a process behavior chart for individual values based on counts. While he could plot the data in a running record, and while he could use an average value as the central line for this running record, the obstacle was how to measure the dispersion so as to filter out the routine variation. With individual values he did not see how to use the within-subgroup variation, and he knew better than to try and use the global standard deviation statistic which would be inflated by any exceptional variation present. So he decided to use theoretical limits based on a probability model.

The classic probability models for simple count data are the Binomial and the Poisson, and Shewhart knew that both of these models have a dispersion parameter that is a function of their location parameter. This meant that the estimate of location obtained from the data could also be used to estimate the dispersion. Thus, with one location statistic he could estimate both the central line and the three-sigma distance.

<table>
<thead>
<tr>
<th>Data Characterized by Binomial Model</th>
<th>Data Characterized by Poisson Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area of Opportunity = n</td>
<td>Area of Opportunity = a</td>
</tr>
<tr>
<td>Constant n</td>
<td>Constant a</td>
</tr>
<tr>
<td>Variable n</td>
<td>Variable a</td>
</tr>
<tr>
<td>np-Chart for Counts</td>
<td>p-Chart for Proportions</td>
</tr>
<tr>
<td>p-Chart for Proportions</td>
<td>c-Chart for Counts</td>
</tr>
<tr>
<td>u-Chart for Rates</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Specialty Charts for Count-Based Data

This dual use of an average to characterize both location and dispersion means that p-charts, np-charts, c-charts, and u-charts all have limits that are based upon a theoretical relationship between the mean and the dispersion. Hence these specialty charts can all be said to use theoretical limits. If the counts can be reasonably modeled by either a Binomial distribution or a Poisson distribution, then one of these specialty charts will provide appropriate limits for the data.

The Binomial model will be appropriate for count-based data when the counts are counts of items where the probability that an item possesses the attribute being counted remains the same within each area of opportunity (each time period).
The Poisson model will be appropriate for count-based data when the counts are counts of events where there is a uniform likelihood of an event happening throughout the finite area of opportunity examined for each time period.

Over the years many textbooks and standards have forgotten that the assumption of a Binomial model or a Poisson model is a prerequisite for the use of these specialty charts. This is a problem because there are many types of count-based data that cannot be characterized by either a Binomial or a Poisson distribution. When such data are placed on a p-chart, np-chart, c-chart or u-chart the theoretical limits obtained will be wrong.

So what are we to do? The problem with the theoretical limits lies in the assumption that we know the exact relationship between the central line and the three-sigma distance. The solution is to obtain a separate estimate of dispersion, which is what the XmR Chart does: While the average will characterize the location and serve as the central line for the X Chart, the average moving range will characterize dispersion and serve as the basis for computing the three-sigma distance for the X Chart.

Thus, the major difference between the specialty charts and the XmR Chart is the way in which the three-sigma distance is computed. The p-chart, np-chart, c-chart, and u-chart will have the same running record, and essentially the same central lines, as the X Chart. But when it comes to computing the three-sigma limits the specialty charts use an assumed theoretical relationship to compute theoretical values while the XmR Chart actually measures the variation present in the data and constructs empirical limits.

To compare the specialty charts with the XmR Chart we shall use three examples. The first of these will use the data of Figure 2. These values come from an accounting department which keeps track of how many of their monthly closings of departmental accounts are finished “on-time.” The counts are the number of closings, out of 35 closings each month, that are completed on time.

Here both the np-chart and the X Chart computations give essentially the same lower limit. (The upper limits are not shown since they exceed the maximum value of 35 on-time closings.) Here the two approaches are essentially identical because these counts seem to be appropriately modeled by a Binomial distribution. If you are sophisticated enough to determine when this happens, then you will know when the np-chart will work and can use it successfully. On the other hand, if you are not sophisticated enough to know when a Binomial model is appropriate, then you can still use an XmR Chart. As may be seen here, when the np-chart would have worked, the empirical limits of the X Chart will mimic the theoretical limits of the np-chart, and you will not have lost anything by using the XmR Chart instead of the np-chart.
Our next example will use the on-time shipments for a plant. The data are shown in Figure 3 along with both the $X$ Chart and the $p$-chart for these data.

| Month       | Year | Total No. | On-Time % | | Month       | Year | Total No. | On-Time % |
|-------------|------|-----------|-----------| |-------------|------|-----------|-----------|
| January     | 01   | 191       | 92.1      | | January     | 02   | 170       | 91.2      |
| February    | 01   | 203       | 91.6      | | February    | 02   | 270       | 91.1      |
| March       | 01   | 220       | 91.8      | | March       | 02   | 167       | 90.4      |
| April       | 01   | 200       | 91.5      | | April       | 02   | 216       | 90.7      |
| May         | 01   | 236       | 91.1      | | May         | 02   | 227       | 90.7      |
| June        | 01   | 213       | 91.1      | | June        | 02   | 149       | 91.3      |
| July        | 01   | 212       | 90.1      | | July        | 02   | 182       | 91.8      |
| August      | 01   | 241       | 89.2      | | August      | 02   | 224       | 90.6      |
| September   | 01   | 159       | 89.9      | | September   | 02   | 246       | 91.5      |
| October     | 01   | 217       | 90.8      | | October     | 02   | 185       | 91.9      |
| November    | 01   | 181       | 91.2      | | November    | 02   | 261       | 91.6      |
| December    | 01   | 113       | 91.2      | | December    | 02   | 140       | 91.4      |

The $X$ Chart shows a process with three points at or below the lower limit. The variable-width $p$-chart limits are five times wider than the limits found using the moving ranges. No points fall outside these limits. This discrepancy between the two sets of limits is an indication that the data of Figure 3 do not satisfy the Binomial conditions. Specifically, the probability of a shipment being on time is not the same for all of the shipments in any given month. Because the Binomial model is inappropriate the theoretical $p$-chart limits are incorrect. However, the empirical limits of the $XmR$ Chart, which do not depend upon the appropriateness of a particular probability model, are correct.

Our final comparison will use the data of Figure 4. There we have the percentage of incoming shipments for one electronics assembly plant that were shipped using air freight. Two points fall outside the variable width $p$-chart limits while no points fall outside the $X$ Chart limits.
Figure 4 is typical of what happens when the area of opportunity for a count of items gets excessively large. The Binomial model requires that all of the items in any given time period will have the same chance of possessing the attribute being counted. Here this requirement is not satisfied. With thousands of shipments each month, the probability of a shipment being shipped by air is not the same for all of the shipments. Thus, the Binomial model is inappropriate, and the theoretical $p$-chart limits which depend upon the Binomial model are incorrect. The $X$ Chart limits, which here are twice as wide as the $p$-chart limits, properly characterize both the location and dispersion of these data and are the correct limits to use.

Thus, the difficulty with using a $p$-chart, $np$-chart, $c$-chart, or $u$-chart is the difficulty of determining whether the Binomial or Poisson models are appropriate for the data. As seen in Figures 3 and 4, if you overlook the prerequisites for a specialty chart you will risk making a serious mistake in practice. This is why you should avoid using the specialty charts if you do not know how to evaluate the appropriateness of these probability models.

In contrast to this use of theoretical models which may or may not be correct, the $XmR$ Chart provides us with empirical limits that are actually based upon the variation present in the data. This means that you can use an $XmR$ Chart with count based data anytime you wish. Since the $p$-chart, the $np$-chart, the $c$-chart, and the $u$-chart are all special cases of the chart for individual values, the $XmR$ chart will mimic these specialty charts when they are appropriate and will differ from them when they are wrong.

<table>
<thead>
<tr>
<th>All Count-Based Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant Area of Opportunity</td>
</tr>
<tr>
<td>$XmR$ Chart for Counts</td>
</tr>
<tr>
<td>Variable Area of Opportunity</td>
</tr>
<tr>
<td>$XmR$ Chart for Rates</td>
</tr>
</tbody>
</table>

Figure 5: An Assumption Free Approach for Count-Based Data
Thus, if you do not have advanced degrees in statistics, or if you simply have a hard time determining if your counts can be characterized by a Binomial or a Poisson distribution, you can still verify your choice of specialty chart for your count-based data by comparing the theoretical limits with the empirical limits of an XmR chart. If the empirical limits are approximately the same as the theoretical limits, then the probability model works. If the empirical limits do not approximate the theoretical limits, then the probability model and the theoretical limits are probably wrong.

Of course, you can guarantee that you have the right limits for your count-based data by simply using the XmR chart to begin with. The empirical approach will always be right.

Confidence Intervals to Assess Differences between Groups

To determine whether the difference between two means is statistically significant, analysts often compare the confidence intervals for those groups. If those intervals overlap, they conclude that the difference between groups is not statistically significant. If there is no overlap, the difference is significant.

However, visually assessing the overlap in this manner is an overly conservative approach. It’s true that when the confidence intervals don’t overlap, the difference between groups is statistically significant. However, when there is some overlap, the results might still be statistically significant. In other words, this visual comparison method fails to reject the null hypothesis more frequently than the corresponding hypothesis test. This method decreases the statistical power of your assessment (higher type II error rate), which means you might miss important findings.

This visual method of assessing the overlap is easy to perform, but it comes with the cost of reducing your ability to detect a difference. Fortunately, there is a simple solution to this problem that allows you to perform a visual assessment yet not diminish the power of your analysis.

In this column, I’ll start by showing you the problem in action and explain why it happens. Then, we’ll proceed to an easy alternative method.

Comparing Groups Using Confidence Intervals of each Group Estimate

For all hypothesis tests and confidence intervals, you are assessing the properties of population parameters. These parameters can be properties like population means, standard deviations, proportions, and rates. For these examples, I’ll use means, but the same principles apply to the other types of parameters.

Imagine that you’re comparing the means of two groups. You graph the 95% confidence intervals for the group means, as shown below.
Upon seeing how these intervals overlap, you conclude that the difference between the groups is not statistically significant. After all, if they’re overlapping, they’re not different, right? This conclusion sounds logical, but it’s not necessarily true. In fact, for these data, the 2-sample t-test results are statistically significant with a p-value of 0.044. Despite the overlap, the difference between these two means is statistically significant.

This apparent discrepancy between confidence intervals and hypothesis results might surprise some analysts. There is an expectation that confidence intervals with a confidence level of \((100 - X)\) will always agree with a hypothesis test that uses a significance level of \(X\) percent. For example, analysts often pair 95% confidence intervals with tests that use a 5% significance level. It’s true. Confidence intervals and hypothesis test should always agree. So, what is happening in the example above?

The problem occurs because we are not comparing the correct confidence intervals to the hypothesis test result. The test results apply to differences between means while the CIs apply to the estimate of each group mean—not the difference between the means. We’re comparing apples to oranges, so it’s not surprising that the results differ.

To obtain consistent results, we must use confidence intervals for differences between group means—we’ll get to those CIs shortly.

However, if you’re determined to use CIs of each group to make this determination, there are several possible methods.

Goldstein and Healy (1995) find that for barely non-overlapping intervals to represent a 95% significant difference between two means, use an 83% confidence interval of the mean for each group. The graph below uses this confidence level for the same dataset as above, and they don’t overlap.
Cumming & Finch (2005) find that the degree of overlap for two 95% confidence intervals for independent means allows you to estimate the p-value for a 2-sample t-test when sample sizes are greater than 10. When the confidence limit of each CI reaches approximately the midpoint between the point estimate and the limit of the other CI, the p-value is near 0.05. The first graph in this column, with the 95% CIs, approximates this condition, and the p-value is near 0.05. Lower amounts of overlap correspond to lower p-values. For example, 95% CIs where the end of one CI just reaches the end of the other CI corresponds to a p-value of about 0.01.

To me, these approaches seem kludgy. Using a confidence interval of the difference is an easier solution that even provides additional useful information.

**Assessing Confidence Intervals of the Differences between Groups**

Previously, we saw how the apparent difference between the group CIs and the 2-sample test results occurs because we used the wrong confidence intervals. Instead, we need a CI for the difference between group means. This type of CI will always agree with the 2-sample t-test—just be sure to use the equivalent combination of confidence level and significance level (e.g., 95% and 5%). We’re now comparing apples to apples!

Using the same dataset as above, the confidence interval below presents a range of values that likely contains the difference between the means for the entire population. The interpretation continues to be a simple visual assessment. Zero represents no difference between the means. Does the interval contain zero? If it does not include zero, the difference is statistically significant because the range excludes no difference. At a glance, we can tell that the difference is statistically significant.

The pooled standard deviation is used to calculate the intervals.
This graph corresponds with the 2-sample t-test results below. Both test the difference between the two means. This output also provides a numerical representation of the CI of the difference [0.06, 4.23].

**Two-Sample T-Test and CI: Strength B, Strength A**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strength B</td>
<td>20</td>
<td>22.84</td>
<td>3.08</td>
<td>0.69</td>
</tr>
<tr>
<td>Strength A</td>
<td>20</td>
<td>20.69</td>
<td>3.41</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Difference = μ (Strength B) − μ (Strength A)
Estimate for difference: 2.15
95% CI for difference: (0.06, 4.23)
T-Test of difference = 0 (vs ≠): T-Value = 2.09 **F-Value = 0.044** DF = 37

In addition to providing a simple visual assessment, the confidence interval of the difference presents crucial information that neither the group CIs nor the p-value provides. It answers the question, based on our sample, how large is the difference likely to be in the population? Like any estimate, there is a margin of error around the point estimate. It’s important to factor in this margin of error before acting on findings.

For our example, the point estimate of the mean difference is 2.15, and we can be 95% confident that the population difference falls within the range of 0.06 to 4.23.

As with all CIs, the width of the interval for the difference reveals the precision of the estimate. Narrower intervals suggest a more precise estimate. And, you can assess whether the full range of
Hypothesis Testing

values is practically significant. Remember, statistical significance doesn’t necessarily indicate that the results are meaningful in the real world.

When the interval is too wide (imprecise) to be helpful and/or the range includes differences that are not practically significant, you have reason to hesitate before making decisions based on the results. These types of CI results can indicate that you won’t obtain meaningful benefits even when a difference is statistically significant.

There’s no statistical method for answering questions about how precise an estimate must be or how large an effect must be to be practically useful. To use the CI of the difference to answer these questions, you’ll need to apply your subject-area knowledge. For this example, it’s important to note that the low end of the CI is very close to zero. It will not be surprising if the actual population difference falls close to zero, which might not be practically significant despite the statistically significant result.

When you’re comparing groups, assess confidence intervals of those differences rather than comparing confidence intervals for each group. This method is simple, and it even provides you with additional valuable information.

References


Standards InSide-Out

Mark Johnson, PhD, University of Central Florida, Standards Representative for the Statistics Division

Final Column

It has been a great honor and pleasure to contribute these columns (Standards InSide-Out) in the Statistics Division newsletter and more recently with the splendid Statistics Digest. The contemporary column name was a (possibly) obvious play on the International Standards Organization (ISO). This wordplay is the variety of “humor” that encouraged my early retirement from university teaching to students the age of my grandchildren.

I understand that Didier Greenleaf, Jr. has agreed to take over the standards reporting in the Digest and I wish him the best in this endeavor. Personally, I have an increased respect for newspaper columnists and op-ed types who pump out columns on a regular basis. It is a bit embarrassing to admit how much time it was taking me to generate each column—I would have never cut it in a newsroom. Perhaps it is indicative of the effects of aging.

The primary purpose of the column was to report on activities of the ISO TC69 committee on statistical standards, with which I have participated almost thirty years. The first meeting I attended was in Berlin in November 1989, just as the Wall was coming down. During the week of my initial meeting I had the chance to visit East Berlin through Checkpoint Charlie. In June of this year the meeting returned to Berlin and I was once again able to visit Pergamon Museum, Fernsehturm tower and other great sites. International standards work has its privileges! Another purpose of the column was to encourage additional participation of experts in standards work. Again, I implore interested experts to contact Jennifer Admussen at Standards@asq.org.
I appreciate the Statistics Division support of travel over the past several years. Previously, the Division had supported Ed Schilling for many years.

Finally, I would also like to recognize the long term and outstanding contributions to international and US standards by Dr. Michèle Boulanger, Chair of ISO TC69. She completes her term as Chair by the next meeting in Tokyo. We collaborated on a number of journal publications (International Statistical Review, Quality Engineering and most recently Quality Progress). More generally, without her support and encouragement I would not have been able to work in standards these past many years.

Mark E. Johnson
Professor Emeritus, UCF
Senior Member, ASQ
Fellow, ASA

FEATURE

A Brief Discussion on Conjugate Beta Priors

Harish Jose

Will your favorite team win their next game? Will the next coin toss result in Heads? Will the next lot of product pass inspection? All of these are binomial inference problems. There are only two results that can happen, generally notated as Pass or Fail. In Bayesian Statistics, we can easily represent this with the help of the beta distribution. This article looks at the Conjugate Beta Priors in simple terms.

Simply put, Bayesian inference relies on your prior beliefs regarding the parameter you are interested in and updates your belief in the presence of new data. The use of Beta distributions helps us with easily calculating the prior distribution and the posterior distribution (the probability distribution for the parameter given the new data). The prior distribution describes our beliefs, what we think is the parameter, without the new data. If we are looking at a phenomenon where we have no insight, like how many hours a new product will last, we would want our prior beliefs to have minimal impact on the calculation of the posterior data. These types of priors are conventionally called “noninformative priors.” Even though
they are called noninformative priors, they do contain some information. The other type of priors is called "informative priors," where you have some idea about the parameter you are interested in. Thus, you can make educated guesses. This could be based on expert opinions or empirical data that is readily available. The term "conjugate" simply means that the prior and posterior distributions belong to the same family of distribution.

The prior distribution is expressed as Beta(a, b), where a is the number of successes and b is the number of failures. Thus, the sample size is (a + b). The mean, median and standard deviation calculations are easy to calculate.

\[
\text{Mean} = \frac{a}{a + b} \quad (1)
\]

\[
\text{Median} \approx \frac{(a - 1/3)}{(a + b - 2/3)} \quad (2)
\]

\[
\text{Mode} = \frac{(a - 1)}{(a + b - 2)} \quad (3)
\]

\[
\text{Standard Deviation} = \sqrt{\frac{a \times b}{(a + b)^2 \times (a + b + 1)}} \quad (4)
\]

From prior distribution, we can easily get to the posterior distribution by adding successes to "a" and adding failures to "b". Thus, for example the mean for the posterior distribution is:

\[
\text{Posterior Mean} = \frac{(a + x)}{(a + x + b + n - x)}
\]

The other distribution descriptors can be easily calculated for the posterior distribution.

Noninformative Beta Priors
We will look at four noninformative priors in this article. They are uniform prior, Haldane’s prior, Jeffreys prior, and neutral prior. The common characteristics of these priors are that both a and b are the same to achieve symmetry, and they are both \(\leq 1\). If we do not have a good prior belief, then we should use as small a prior distribution sample size as possible. This also indicates that the distribution spread is wider, which makes sense because we cannot concentrate around a particular central point without good belief.

Uniform Prior
The main idea behind the noninformative beta priors is the principle of indifference. If you do not have a good belief on your prior, you should treat all values as equally likely. This is given as Beta(1, 1). Intuitively, this makes sense from the standpoint that the posterior mode is the same as \(x/n\), the maximum likelihood value (sample mean). The other aspect for using uniform prior is the idea of maximum entropy which indicates higher levels of ignorance about the parameter. This was also the prior that Thomas Bayes and later on Pierre-Simon Laplace used. This prior is useful when looking at rare events or when \(x = n\). [1] The uniform prior distribution is depicted below.
Haldane’s Prior
Haldane’s prior refers to Beta(0, 0). This is an “improper prior” in the sense that the prior distribution does not integrate to 1. However, when $a$ and $b = 0$, the posterior mean equals $x/n$, the maximum likelihood value (sample mean). Since both $a$ and $b$ are equal to zero, this is the lowest sample size possible to have the greatest ignorance possible. In other words, there is no information available as a prior. These characteristics makes this an interesting prior. This prior is an extreme example because values of 0 and 1 are highly represented. This is useful for instances where the probabilities hover around 0 and 1 with nothing in between.

Jeffreys’s Prior
Jeffreys’s prior for our purpose is Beta(0.5, 0.5). It has the valuable property where we get the same results no matter which scale we use for the parameter we are interested in. Intuitively, this prior may be perceived as a compromise between Haldane’s prior and Uniform Prior.
Neutral Prior

Neutral prior was introduced by Jouni Kerman, and has the values Beta(1/3, 1/3). Kerman explained that the conventional noninformative conjugate priors tend to shrink the posterior quantities toward the boundary or toward the middle of the parameter space, making them appear excessively informative. Kerman’s solution to this was the neutral prior. Similar to uniform prior and Haldane’s prior, the neutral prior has the interesting property of the posterior median being equal to x/n, the maximum likelihood value (sample mean). This prior is useful when looking at rare events or when x = n.

Informative Priors

As the name suggests, the informative priors bring information to the table. You have a good idea of what the distribution would look like. The general advice is to use the Beta prior \( \text{Beta}(a, b) \), where a and b are the number of successes and failures you would expect based on your experience, prior knowledge and/or expert opinion.

Another approach is to estimate the mean and standard deviation based on elicitation from experts. The value for a and b can be algebraically solved using equations (1) and (4). We should be careful to use lower values for a and b since higher values indicate higher confidence which might add unwanted bias in the model.

Heuristics to Use

Some valuable heuristics for determining priors are below.

a) Gelman, Simpson and Betancourt advise that “The prior can generally only be understood in the context of the likelihood.” Logically, the prior distribution should come before the data model, but in practice, priors are often chosen with reference to a likelihood function. In their paper, the authors propose that we think generatively by considering the potential measurements consistent with a given prior and predictively by validating those potential measurements against data that we collect.

b) The effect of sample size is included in how the posterior distribution comes out. A sharp focused distribution with a distinct peak indicates an adequate sample size. Use more samples if this is not achieved. A larger \((a + b)\) value results in lower variance.

c) If the prior distribution is truly uninformative, the posterior is very much impacted by the data. This is evident in the use of uniform prior.

d) If the prior distribution carries information, the posterior is a combination of prior and data. Thus, if the prior belief is highly biased, you will need a lot more data to “normalize” the bias. In other words, the \((a + b)\) should be relatively small compared to the new data sample size.
A Brief Discussion on Conjugate Beta Priors

e) Today’s posterior is tomorrow’s prior. You can use a_post and b_post values from the posterior distribution as the prior values for the next dataset. A small run could be pursued beforehand to determine the a (number of successes) and b (number of failures), which can then be used for the prior distribution.

f) Always be transparent and directly state the rationale for choosing the priors.

g) In the event of zero failures, start with uniform or neutral priors. Gather additional samples, if possible.

h) If in doubt and if feasible, use more samples.

Hypothetical Example

Your company is launching a new light bulb “A”. You would like to understand the quality of the product. You were given 20 bulbs to perform testing. At the end of test, you had 17 bulbs that still worked after 100 hours and 3 that failed.

The use of conjugate Beta priors makes it easy to do the calculations in Microsoft Excel. You can try using the uniform prior (Beta (1, 1)) to calculate the posterior mean. Using equations (6) and (7), we can calculate a_post and b_post as 18 and 4 respectively.

Using equation (5), you were able to calculate the posterior mean as 0.818. Compare this to the frequentist proportion 17/20 or 0.85. In Microsoft Excel, you can also easily calculate the different quantile values. For example, the 95% credible interval can be calculated using the betainv function in Excel as follows:

\[ \text{BetaInv}(0.025, 18, 4) = 0.6366 \]
\[ \text{BetaInv}(0.975, 18, 4) = 0.9455 \]

Based on our sample of 20, using a 95% credible interval we expect between 64% and 95% of “A” type light bulbs will last 100 hours.

References


[2] Neutral Noninformative and Informative Conjugate Beta and Gamma Prior Distributions. Author: Jouni Kerman. 2011

[3] The prior can generally only be understood in the context of the likelihood. Authors: Andrew Gelman, Daniel Simpson, Michael Betancourt. 2017


About the Author

Harish Jose has over 10 years experience as a Quality Engineer in the medical devices field. He is a graduate of the University of Missouri-Rolla (USA) where he obtained a master’s degree in Manufacturing Engineering and published two articles. He is an ASQ member with multiple ASQ certifications including Reliability Engineer, Six Sigma Black Belt and Quality Engineer. He has subject matter expertise in lean, data science, database programming and industrial experiments. He can be reached at harishjose@gmail.com. His LinkedIn profile is available at https://www.linkedin.com/in/harishjose.
2018 William G. Hunter Award: Dr. James M. Lucas

The Statistics Division of the American Society for Quality (ASQ) is pleased to announce that Dr. James M. Lucas is the recipient of the 2018 William G. Hunter Award. The William G. Hunter Award was established by the Statistics Division in 1987 to recognize the many contributions of its founding chair at promoting the use of applied statistics and statistical thinking. The attributes that characterize Bill Hunter's career - consultant, educator for practitioners, communicator, and integrator of statistical thinking into other disciplines - are used to help decide the recipient.

Dr. Lucas is the principal at J. M. Lucas and Associates, a consulting firm in Statistics and Quality Management. This firm implements business systems with statistical aspects. Before starting his own consulting firm Dr. Lucas was a Senior Consultant at DuPont’s Quality Management and Technology Center for over twenty years where he conducted his early seminal work on applied statistics. Dr. Lucas’s research focuses on practical solutions to real-world problems, emphasizing the underlying science for the problem. He has successfully integrated statistical thinking with other disciplines throughout his career. For example, he was a major contributor to the development of statistical systems used throughout DuPont, including experimental design systems and statistical process control initiatives.

He has been an Adjunct Professor at the University of Delaware and at Drexel University and he has directed six PhD dissertations. He is a Fellow of the American Statistical Association (ASA) and of the American Society for Quality (ASQ), an Associate Editor of the Journal of Quality Technology, and a past Associate Editor of Chemometrics and Intelligent Laboratory Systems and of Technometrics. He has over 70 publications and many are cited frequently. He authored the most cited paper in two volumes of Technometrics and in two volumes of the Journal of Quality Technology. He has won many awards including the Shewhart Medal, the Brumbaugh Award, the H. O. Hartley Award, the Ellis R. Ott Foundation Award, the Don Owen Award, the Shewell Award, and the Youden Prize. He has a PhD in Statistics from Texas A&M University, a MS in Statistics from Yale University, and a BS in Engineering from The Pennsylvania State University.
Lean and Six Sigma Conference
4–5 March 2019 | Phoenix, AZ | https://asq.org/

Do you have technical proficiencies and leadership responsibilities within your organization? Are you actively involved in process improvement, organizational change, and development dynamics related to a successful lean and Six Sigma culture? This conference is for you!

World Conference on Quality and Improvement
20–22 May 2019 | Fort Worth, TX | https://asq.org/

The World Conference on Quality and Improvement is ASQ’s flagship conference. Join a broad attendee base representing a wide array of industries and over 45 countries from around the world to gain knowledge, improvement methodologies, quality tools, best practices, and networking contacts. With over 2,500 attendees, the World Conference is the ideal forum to meet quality professionals with a wide range of backgrounds and experiences.

Joint Statistical Meetings
27 July–1 August 2019 | Denver, Colorado | http://ww2.amstat.org/

JSM (the Joint Statistical Meetings) is the largest gathering of statisticians and data scientists held in North America. It is also one of the broadest, with topics ranging from statistical applications to methodology and theory to the expanding boundaries of statistics, such as analytics and data science.

JSM also offers a unique opportunity for statisticians in academia, industry, and government to exchange ideas and explore opportunities for collaboration. Beginning statisticians (including current students) are able to learn from and interact with senior members of the profession.

ENBIS-19
2–4 September 2019 | Budapest, Hungary | https://enbis.org/

The 19th annual conference of the European Network for Business and Industrial Statistics will be hosted by the Department of Probability Theory and Statistics at the Eötvös Loránd University, Budapest and will take place from the 2nd to 4th of September, 2019.
2019

ELECTED POSITIONS
CHAIR-ELECT/AUDITING
Amy Ruiz
chair@asqstatdiv.org

CHAIR
Mindy Hotchkiss
chair@asqstatdiv.org

PAST CHAIR/NOMINATING
Steven Schuelka
pastchair@asqstatdiv.org

SECRETARY
Bokelman
secretary@asqstatdiv.org

TREASURER
Joyce Crum
treasurer@asqstatdiv.org

APPOINTED POSITIONS

Vice Chair Content
Gary Gehring
content@asqstatdiv.org

STATISTICS DIGEST EDITOR
Didier Greenleaf
didiergreenleafjr@gmail.com

NEWSLETTER CONTENT REVIEW
Kurtis Shuler
kurtis.shuler@gmail.com

WEBINAR / YOUTUBE
Harry Rowe
webinars@asqstatdiv.org

SOCIAL MEDIA MANAGER
Brian Sersion
bsersion@gmail.com

WEBSITE CONTENT MANAGER
Geoff Farmer
gfarmer118@yahoo.com

Vice Chair Community Involvement
Jennifer Williams
community@asqstatdiv.org

FTC PROGRAM REP
Richard McGrath
rnmgra@bgsu.edu

FTC STEERING COMMITTEE
Bill Myers
myers.wr@pg.com

FTC SHORT COURSE CHAIR
Brian Sersion
bsersion@gmail.com

WCQI COORDINATOR
Steven Schuelka
pastchair@asqstatdiv.org

DATA SCIENCE/ANALYTICS INTEREST GROUP CHAIR
Herb McGrath
rnmgra@bgsu.edu

MEMBERSHIP/OUTREACH CHAIR
Jennifer Williams
membership@asqstatdiv.org

EZINE
Shoshana Bokelman
secretary@asqstatdiv.org

VOC CHAIR
Michael Kirchner
michael.r.kirchner@gmail.com

Vice Chair Awards
Peter Parker
awards@asqstatdiv.org

FTC STUDENT/EARLY CAREER GRANTS
Jennifer Williams
ftcgrants@outlook.com

NELSON AWARD
TBD

HUNTER AWARD
Joel Smith
joelmcquarriesmith@gmail.com

YOU DEN ADDRESS
Steven Schuelka
pastchair@asqstatdiv.org

BISGAARD AWARD
TBD

EXAMINING CHAIR
Daksha Chokshi
examining@asqstatdiv.org

FTC SHORT COURSE CHAIR
Brian Sersion
bsersion@gmail.com

WCQI COORDINATOR
Steven Schuelka
pastchair@asqstatdiv.org

DATA SCIENCE/ANALYTICS INTEREST GROUP CHAIR
Herb McGrath
rnmgra@bgsu.edu

MEMBERSHIP/OUTREACH CHAIR
Jennifer Williams
membership@asqstatdiv.org

VOC CHAIR
Michael Kirchner
michael.r.kirchner@gmail.com

Vice Chair Awards
Peter Parker
awards@asqstatdiv.org

FTC STUDENT/EARLY CAREER GRANTS
Jennifer Williams
ftcgrants@outlook.com

NELSON AWARD
TBD

HUNTER AWARD
Joel Smith
joelmcquarriesmith@gmail.com

YOU DEN ADDRESS
Steven Schuelka
pastchair@asqstatdiv.org

BISGAARD AWARD
TBD

EXAMINING CHAIR
Daksha Chokshi
examining@asqstatdiv.org
The ASQ Statistics Division Newsletter is published three times a year by the Statistics Division of the American Society for Quality.

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

Didier Greenleaf Jr.
Editor
email: newsletter@asqstatdiv.org

Other communications relating to the ASQ Statistics Division should be addressed to:

Mindy Hotchkiss
Division Chair
email: chair@asqstatdiv.org

Communications regarding change of address should be sent to ASQ at:

ASQ
P.O. Box 3005
Milwaukee, WI 53201-3005

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

Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.

John Tukey