

## Gini's Coefficient of Dispersion - The Range and The Standard Deviation

Long before Shewhart started working with Xbar and R charts, Gini, in 1912, started wondering about how to measure dispersion by looking at the distances between each possible pair of X's in the sample. That is, he considered all possible distances between the X's rather than just the maximum distance (the Range).

$$\text{Range} = X_{\max} - X_{\min}$$

$$d(i) = X(i) - X(j)$$

where  $i$  &  $j$  take values from 1 to  $n$  (sample size)

This definition caused a problem since one would naturally average these distances to get a measure of dispersion. The average of these simple measures of distance would, of course, be zero since:

$$X(i) - X(j) = -(X(j) - X(i)) \text{ for all } i \text{ and } j.$$

So, Gini decided to use, as a measure of dispersion,

$D(\text{Gini}) = \text{Dispersion} = \sum |d(i)|/n^2$  - since there are  $n^2$  possible differences to consider.

This measure is referred to as Gini's Coefficient of Dispersion. It is the average of the ranges of every possible subsample size 2 that can be formed from the sample of size  $n$ .

So, Gini's Coefficient would provide us with a variety of ranges that can be used to calculate an Rbar for each subgroup that could be used in conjunction with the Xbar for each subgroup. This makes a nice feeling of equal attention being given to the Center and the Dispersion of the process as represented by the subgroup data.

But, Gini's Coefficient was probably kept in the bottle for the lack of high speed computer power. There would be  $n^2$  ranges to calculate for each subgroup.

Let's try the more usual statistical approach to getting rid of the +'s cancelling out the -'s; we'll square those distances rather than taking the absolute value as Gini did.

$$D(\text{Squares}) = \text{Dispersion} = \sum (d(i)^2)/n^2 \\ = \sum (X(i) - X(j))^2/n^2$$

**Example:** Suppose the X's in the subgroup with  $n = 5$  are 1, 2, 3, 4, 5.

The single range would be:  $R = 5 - 1 = 4$ . And the estimate of Std Dev would be:

$$\text{Range}/d2 = 4/2.33 = 1.717$$

The direct estimate of Std Dev would be:

$$s_x/c4 = 1.581/0.7979 = 1.981$$

NEXT, let's look at the average of the squares of the "useful"  $d(i)$ . All possible ranges are shown in the table below. Of those 25 ranges, only 20 (the off-diagonal ones) provide useful information about the dispersion. It is clear that using the distance of one variable to itself provides no real information about how disperse the values are - those distances are all automatically zero.

	1	2	3	4	5
1	0	1	2	3	4
2	1	0	1	2	3
3	2	1	0	1	2
4	3	2	1	0	1
5	4	3	2	1	0

$D(\text{squares})$  - for only the off-diagonal distances  
 $= \sum d^2/(n(n-1)) = 100/20 = 5$

To get the estimate of Std Dev, we must take the  $\sqrt{5/2}$  (the "2" is all in the math) and then adjust by dividing by  $c4$

The result is: 1.581/0.7979

**which is the SAME as the std. dev. calculated in the usual fashion.**

Amazing? Not really. The two are algebraically equivalent. The variance is the average of the squared differences between X's divided by 2.

### Conclusions:

1. Although the Std Dev is often said to be a dispersion measure based on distances that X's are from the center (Xbar), the Std Dev can also be said to be a dispersion measure which measures the distances between pairs of values and hence doesn't require the average value at all.

### In this Issue:

SPC Training  
 CQC Datalink  
 Dispersion Article  
 QMLite  
 Stochos New Service  
 Stochos Users Group

2. The issues of  $n$  vs,  $(n-1)$  and "degrees of freedom" are simplified from this approach; the distances between  $X(i)$  and  $X(i)$  is zero, of course. Thus the distances on "main diagonal" furnish no information about the dispersion of the set. The issue is the number of samples containing real information about distances. The concept that the number of samples with real information is about distances is  $(n^2 - n)$  (or  $n*(n-1)$ ) would seem to be more easily explained than "degrees of freedom".

### CQC Datalink Provides Seamless Data Extraction to Stochos Custom/QC™ Software

Stochos' new product, "CQC Datalink" provides easy selection of data from an ODBC database for seamless formatting and integration into Stochos' off-line SPC/SQC analysis software, Custom/QC for Windows™.

CQC Datalink has an easy-to-use interface for extracting and querying data from an ODBC database table. Utilization of CQC Datalink eliminates timely column setup procedures, allowing users to quickly perform off-line analysis with a few simple keystrokes. Users can preview data, sort fields, and save templates for automatic launch of data into Custom/QC's mathematical spreadsheet format.

Flexible template functionality allows template creation during any phase of data extraction making repetitive queries fast and easy. Distribution of template files to key personnel ensures consistent data selection throughout the corporation.

Stochos' current QMDS™ and SPC Direct™ customers will be pleased with the elimination of the export/import burden of transferring data, while Custom/QC™ users can add the CQC Datalink tool to extract data from other ODBC-based systems.

For more information on CQC Datalink and Stochos products, please visit our web page at

[www.stochos.com](http://www.stochos.com) or call 1-800-426-4014.

**Stochos' 2nd Annual User's Meeting will be held Oct. 7 & 8, 2002**  
 at the "Inn and Meeting Place" Rensselaerville, NY

<http://www.innandmeetingplace.com>

Located in the beautiful & unique Stonecrop building, room selection is on a first come first serve basis. The cost is \$500 per person and includes meals, sleeping accommodations, training sessions and transportation to and from the Albany airport. Additional information and registration is available on our web page at: [www.stochos.com](http://www.stochos.com) or by contacting us at (800) 426-4014.

Driven by current manufacturing needs, Stochos has developed a new lighter side Quality Management and SPC software solution - "QMLite". Based on Stochos' current solutions QMDS and SPC Direct, the new QMLite provides a SPC data collection interface that includes detailed information, live SPC charts, on-screen alarms, test scheduling and reporting.

### All the functionality for a fraction of the cost

- Data Collection
- Live SPC Charts
- Detailed Information Tracking
- Easy Query and Reporting
- Test Scheduling
- On-Screen Alarms

The value of a great quality database system lies in the information that can be tracked and utilized to make formidable conclusions. These analyses can then be further carried to include Corrective & Preventive action, ideally resulting in better quality and lower yields. Thus, QMLite tracks all pertinent process information such as: Machine, Shift, Job, Product, Unit ID, Tests, Run, Operation Sequence, Corrective Action and more.

QMLite also incorporates comprehensive Pareto Analysis and valuable production and job reports. Flexible configuration options allow each installation of QMLite to appear like a customized application providing user-friendly, consistent field selection.

Diversified functionality such as scheduling and test frequency is configured per product presenting operators with specific inspection requirements. Assignable test types include variable, mathematical, characteristic and defect. Real-time, on-screen alarms allow operators to immediately react to out-of-control conditions whether performing a corrective action or simply adding a comment.

Stochos also ensures database integrity and security via user login options and integrated ADO technology for connection.

Available for both SQL and MSDE, QMLite can easily be expanded to integrate other quality management tools from Stochos or upgrade to the entire Quality Management and Shop-Floor Data Collection (QMDS) system.

## What's New:

### Stochos Offers Leasing

Manufacturers, both large and small, must continue to improve their production methods and the quality of the products they produce to remain competitive in today's marketplace. Achieving these objectives demands a major company-wide commitment to quality, requiring extensive internal training and plant-floor technological improvements.

These aggressive needs are often met with limited budgets. Manufacturers are forced to put their quality and process improvement projects on hold or settle for limited capability software. Since 1968, Stochos has provided a wide range of quality solutions to virtually all types and sizes of manufacturers. In an effort to continue to provide better opportunities, Stochos employed this value-added service.

### Why lease...?

Leasing is a *cost-effective* method to acquire software to enhance your business at times when there are not sufficient cash resources. Rapid advancement of process improvement methods and tools continually change the market. Leasing allows manufacturers to utilize new and improved technologies in a timely fashion with a *low up-front investment* risk.

- Increased purchasing power
- Cutting-Edge technology
- Preserve capital for other needs
- Tax Deductions
- Lower initial investment

### Benefits to leasing...

Leasing provides flexible options, such as length of time and inclusion of implementation, training and maintenance costs. Projected return on investment coupled with low monthly lease payments allow companies to forge ahead without compromising their current needs.

- Inclusive of all costs
- Continued support
- Flexible payment options
- Satisfy all requirements
- Sharpen your competitive edge
- Accommodate a phased-in installation

For questions regarding leasing options or more information on Stochos products and services, contact us at ...**(800) 426-4014**

Email: [Sales@Stochos.com](mailto:Sales@Stochos.com)

## Statistical Process Control Training

### Basic and Advanced Topics - 3 to 5 Day Course

June 24 - 26 Basic ~ June 26 - 28 Advanced

Course Director: Donald S. Holmes

**Course Description:** This course has two components: a two and 1/2 day basic overview and a two and 1/2 day advanced topics. The basic course deals with such standard tools as histograms, Xbar and R charts, process capability studies and sampling plans. The advanced course deals with statistical process control in automated industries: Key variable identification, regression, correlation, principal components, autocorrelation, discriminant analysis and evolutionary operation (EVOP). The course will cover the basic concepts of statistical analysis and their application to practical problems in process control.

- \* Review of basic SPC tools and advanced tools important to your type of industry
- \* Introduction to Control Charts
- \* Correlation and Regression, Acceptance Sampling
- \* Multiple Variables and Key Variable Identification
- \* Autocorrelation and Time Series Modeling
- \* Case studies used throughout discussions

**Directed by: Donald S. Holmes**  
**President of Stochos, Inc.**

Donald S. Holmes, who founded *STOCHOS* in 1968, is well versed in the teaching and application of quantitative methods to management systems. He is a Fellow of the American Society for Quality (ASQ) and a certified quality engineer.

### Training Location:

Stochos, Inc.  
14 N. College Street,  
Schenectady, NY 12305

1968  2002

For more information and registration contact:

Konnie Steele (800) 426-4014 or  
Visit: [www.stochos.com](http://www.stochos.com)