2004-04-30

Using Applied Statistics to Study a Pharmaceutical Manufacturing Process

John P. Tiani
Worcester Polytechnic Institute

Using Applied Statistics to Study a Pharmaceutical Manufacturing Process

by

John Tiani

A Project Report

Submitted to the Faculty

of

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Master of Science

in

Applied Statistics

May 2004

ABSTRACT

The pharmaceutical manufacturing process of interest produces a suspension for inhalation. Currently, the product is manufactured on two lines. A third and fourth line are in the process of being commissioned and plans are currently in place to construct three additional lines. The manufacturing lines operate independently of one another. Each manufacturing line consists of two actives compounding tanks so their utilization can be rotated to improve manufacturing capacity.

The objective of this project was to study the content uniformity assay values for the 0.25 mg/mL (0.5 mg) manufacturing process through the application of statistical techniques. The study focused on three separate topics:

1. Monitoring process behavior for content uniformity assay values
2. Ascertaining the equivalence of batches manufactured on Line 1 vs Line 2.
3. Monitoring the signal to noise ratio of the content uniformity assay values

In order to accomplish the three tasks above, the following statistical techniques were applied:

1. Control chart techniques were applied to the data, including standard control chart techniques ($\bar{x}$ and S), individuals control chart techniques, and modified limits.
2. An equivalence test for the means of the two processes was conducted.
3. A new control chart, the SNR chart, was developed and implemented.

The results/conclusions of the application of statistical techniques were:

1. The content uniformity assay values were in statistical process control with respect to modified limit control chart techniques.
2. The Line 1 and 2 data were statistically equivalent.
3. The quantity ($\sqrt{n}\ \bar{x}\ /$ s) was in statistical process control. The SNR control chart displayed superior performance to the Individuals control chart.

ACKNOWLEDGMENTS

I would like to take this opportunity to thank my project advisor, Dr. Joseph Petruccelli for all his technical expertise and patience while putting this project together. It has been a long and challenging journey and I would not have made it through without his support. I also appreciate the flexibility he granted me during the course of the preparation of this paper due to personal life changes that came along during the exercise.

I also would like to thank my family, especially my mom (and dad), for all their assistance and guidance through the years.

Most of all, I would like to thank my wife Carrie for all her patience and support while writing this thesis. Her understanding and good nature made it possible for me to devote the time necessary to finish. I also want to thank Nolan and Nina for putting up with dad's continual computer use when he should have been spending more time with them. I love you all! Thanks!

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

1. Introduction

The quality of health care is very important to any society. Patients and their families need and expect high-quality health care. One does not like to think that a physician's diagnosis may be wrong, that a hospital chart is inaccurate, or that a treatment mix-up may have occurred. The drug obtained from the pharmacy is expected to cure the infection or relieve the pain and have its intended effect. People expect that the bottle of medicine has the specified number of tablets and that each tablet contains the specified quantity of the correct drug.

Law requires quality control in the pharmaceutical industry. The organization of the quality control unit, its responsibilities, and the way it performs its duties are covered by federal regulations. The regulations tend to outline a quality control function that emphasizes inspection and defect detection, and pharmaceutical quality control technology. In today's pharmaceutical companies, people in all parts of the organization are being empowered with increased responsibility and authority for the quality of the products and services delivered to internal and external customers. It is not left solely to the quality control department to act as "policemen" to catch the defects. In today's environment, quality of products and services are of concern in all parts of the organization. Quality is introduced in product design, process development, distribution, sales, and marketing, rather than in the production component alone.

Chemical and Biological Considerations in Drug Product

The features of a drug product that are the easiest to understand, measure, and control tend to fall under the classification of chemical or physical attributes. For example, the amount of active ingredient, the type and amount of excipient, the amount of impurity, the time it takes for a tablet or capsule to dissolve, the rate and duration of drug release from an extended release formulation, can be measured. In addition, the stability characteristics of a drug product throughout the manufacturing process and up until patient use can be monitored.

The quality characteristics of most interest to the patient are: "Does the drug work?" and “Is it safe to use?” These issues are addressed by performing large scale experiments called controlled clinical trials, where the efficacy and safety of the drug are carefully studied and documented in a large number of patients. Concurrently, the chemical and physical properties of the drug are determined and specified. Therefore, if the manufacturer produces a uniform drug product that is equivalent to the one proven to work in clinical trials, the patient's expectation that the drug will work is reasonable.

Although it is not possible for the manufacturer to guarantee the patient that a specific tablet, capsule, or injection will cure the ill, the manufacturer can assure the patient that the dosage unit was produced by a controlled process that yields units satisfying the same chemical and physical specifications as dosage units shown experimentally to be bioavailable and therapeutically effective (C. Ralph Buncher and Jia-Yeong Tsay 1994).

The Role of the United States Pharmacopoeia (USP) in Assuring Quality

Marketed pharmaceutical products are subject to recognized standards of identity, quality, strength, and purity documented in various compendia around the world. Some examples include the United States Pharmacopoeia (USP) and National Formulary, the British Pharmacopoeia, the European Pharmacopoeia, the Japanese Pharmacopoeia, and the International Pharmacopoeia.

The USP contains standards for drugs and pharmaceutical substances, which are called Official Monographs. In addition, the USP also contains sections called General Tests and Assays and General Information. These sections cover analytical methods, the design and analysis of biological assays, as well as general information on how to clean glassware, the laws and regulations governing drug manufacturing and distribution, stability considerations, validation of compendial methods, and sterility.

The standards published in the USP are recognized as official and required by law in the United States under a set of regulations called the Food, Drug, and Cosmetic Act. These standards apply at any time in the life of a drug from production to consumption. Manufacturers are expected to develop and utilize release tests and specifications that assure that a drug will comply with compendial standards until its expiration date when stored as directed (C. Ralph Buncher and Jia-Yeong Tsay 1994).

The Role of the Food and Drug Administration (FDA) in Assuring Quality

The FDA is the federal agency which oversees the quality of drugs discovered, developed, and produced for use in the United States. Other countries around the world have similar agencies. It is through the FDA and similar agencies worldwide that the patient and physician have input in the quality control of pharmaceutical products. The regulations which guide the FDA and the pharmaceutical industry in the manufacturing process are the Current Good Manufacturing Practices (cGMPs), published in the Federal Register and in the USP. The FDA, through its field offices around the country, enforces the cGMP regulations through plant inspections, audits, and sample analyses.

In order to meet the increasing demands of the FDA, Quality Assurance Departments have been expanded. In most pharmaceutical companies, Validation Departments have been created to directly deal with increasing FDA demands for documented proof of process quality. With resources at a minimum in most companies, new ways of showing and presenting data had to be developed and implemented. The need to present documented proof of quality processes over time has lead to an increased reliance on statistical techniques to show that a particular process is in control.

1.1. Process Description

Due to the competitive nature of the pharmaceutical industry and the intellectual property at stake, the details of the production process have been omitted for confidentiality reasons.

The product of interest is a New Drug Application (NDA) approved product which is currently aseptically manufactured and filled and then packaged at the subject pharmaceutical company. The drug is manufactured at two strengths (0.125 mg/mL and 0.25 mg/mL), however, the 0.25 mg/mL (0.5 mg) manufacturing process was the focus of this study. This product is the first suspension product to be approved for manufacturing at the site and is the only one of its kind approved for use in its intended market.

The Manufacturing Area consists of two filling lines, one Buffer Preparation Room (consisting of 2 buffer preparation tanks which support both Actives Compounding Rooms), two Actives Compounding Rooms (#1 and #2), and two Isolator Rooms (#1 and #2). Each Actives Compounding Room (consists of 2 compounding tanks) and each Isolator Room (equipped with a processing isolator) supports one filling line. Either of the buffer tanks in the Buffer Preparation Room can be utilized during the manufacturing of BIS. In addition, either of the compounding tanks in the Actives Compounding Room can be utilized during the manufacturing of the product.

The suspension consists of the active ingredient and a number of inactive ingredients. The pH of the suspension is buffered and the final product is protected from light and moisture loss by an aluminum foil envelope to maximize stability. The suspension is filled into single dose units made of low density polyethylene (LDPE). Each single dose unit contains 2 mL of suspension. A strip of five single dose units is packed into an aluminum foil envelope, for protection from light and moisture loss.

Primary Packaging

The primary packaging for all units consists of a LDPE primary container formed from a plastic resin. The microbiological quality of the suspension is maintained by the LDPE used in the Blow-Fill-Seal manufacturing process. The container is translucent and allows visual inspection of the suspension which is important for assurance that the product is properly resuspended prior to dispensing. Since the foil package protects the units from moisture loss and light it is considered part of the primary packaging system. The package insert recommends that the patient retain the foil package so that unused product can be properly stored.

1.2. Problem Statement

The pharmaceutical manufacturing company received approval to manufacture two strengths (0.125 mg/mL and 0.25 mg/mL) of drug product. This product is the first suspension product to be approved for manufacturing at the site. Currently, the product is manufactured on two lines. A third and fourth line are in the process of being commissioned and plans are currently in place to construct three additional lines. The manufacturing lines operate independently of one another. Each manufacturing line consists of two actives compounding tanks so their utilization can be rotated to improve manufacturing capacity.

Production on Line 1 started in the beginning of 2000 and production on Line 2 started in the beginning of 2001. At least 100 batches of each concentration have been manufactured on both lines. The company would like to use statistical techniques to study the manufacturing process for Lines 1 and 2 during the year 2001 by analyzing the content uniformity assay values for the 0.25 mg/mL (or 0.5 mg) manufacturing process. The analysis began in 2001 to allow the process to reach steady state. The study focused on three separate topics:

1.2.1. Developing control charts to assess whether or not the process for content uniformity assay values was in statistical control
1.2.2. Developing a statistical equivalence test to compare the means of assay values of batches manufactured on Line 1 vs Line 2.
1.2.3. Developing a control chart to monitor the signal to noise ratio of the content uniformity assay values.

1.3. Significance of Problem

The above study is very important to the pharmaceutical company. The product is the only one approved for use in its intended market in the U.S. and therefore, it is critical for the company to be able to continue to supply high quality product. It is critical that the company understand its sampling processes and their ability to detect when any of those processes is out of control.

2. Control Charts For Content Uniformity Assay Values

2.1. Objective

The objective of this analysis was to assess the statistical control of the process for the content uniformity assay values. Based on process experience and knowledge, no distinction was made in the active ingredient assay data between the two actives compounding tanks for each line.

2.2. Data

Active ingredient assay data (0.5 mg) for individual units (used to determine content uniformity) produced on Lines 1 and 2 were collected for the year 2001. At least 45 batches were produced on each line during this time period. Although production on Line 1 started in the beginning of 2000 and production on Line 2 started in the beginning of 2001, the data analysis focused on the year 2001 to allow the process to reach steady state.

Content uniformity testing is reported in the batch record as 10 individual active ingredient content values. Samples are taken throughout the batch (4 from the beginning, 3 from the middle, and 3 from the end of fill).

2.3. Procedure

After examining the data for spurious or missing values, we conducted the Box-Pierce test for autocorrelation on the average of the 10 values for each batch. This test failed to reject the null hypothesis of no autocorrelation at the 5% level for any of the data sets considered. We concluded that the use of control charts for uncorrelated data was justified.

Prior to constructing the control charts, an analysis of the begin, middle, and end samples was conducted to see if there was a statistically significant difference between the locations where the samples were taken. In order to check the location factor, a PROC MIXED procedure in SAS was conducted. The following model was utilized for the analysis:

$$Y = X\beta + Z\gamma + \varepsilon,$$

where Y represents a vector of observed data, β is an unknown vector of fixed-effects parameters with known design matrix X, γ is an unknown vector of random-effects parameters with known design matrix Z, and ε is an unknown random error vector.

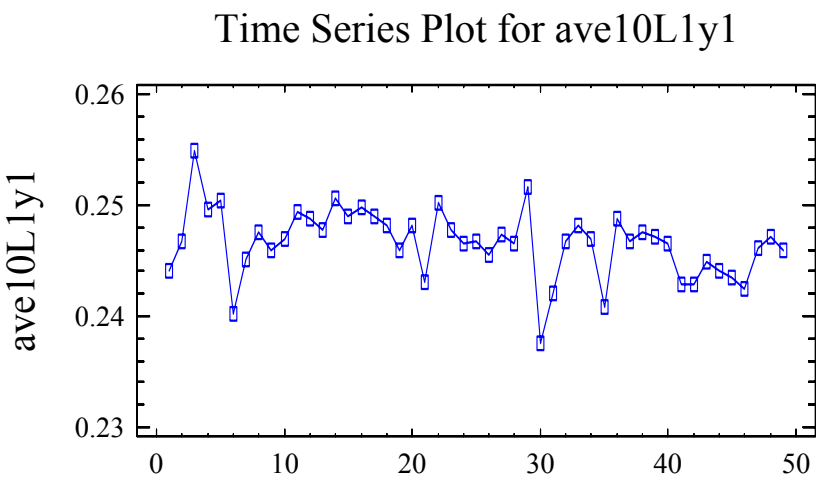
For this analysis, the fixed effect was the location of the sample (begin, middle, and end) and the random effect was the batch from which the sample was taken. The analysis indicated that there was not a statistically significant difference between the locations (begin, middle, and end) at the $\alpha = 0.05$ level.

2.4. Analysis

2.4.1. Correlation Analysis

An autocorrelation analysis was performed for the Line 1 Content Uniformity (CU) data from 2001. The time series plot below represents averages of 10 assay values (4 begin, 3 middle, and 3 end). Based on the Box-Pierce Test for autocorrelation, we cannot reject the hypothesis that the autocorrelation is 0 ($p = 0.86$). In addition, the time series plot below appears to be stationary.

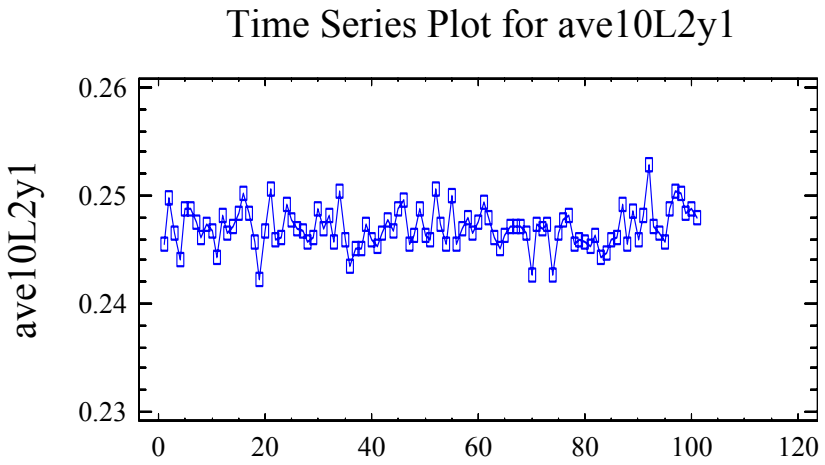
Figure 2.4.1-1: Time Series Plot of Line 1 CU Data from 2001




```
Box-Pierce Test
---------------
     Test based on first 16 autocorrelations
     Large sample test statistic = 10.1872
     P-value = 0.856669
```

An autocorrelation analysis was performed for the Line 1 Content Uniformity (CU) data from 2001. The time series plot below represents averages of 10 assay values (4 begin, 3 middle, and 3 end). Based on the Box-Pierce Test for autocorrelation, we cannot reject the hypothesis that the autocorrelation is 0 ($p = 0.73$). In addition, the time series plot below appears to be stationary.

Figure 2.4.1-2: Time Series Plot of Line 2 CU Data from 2001




```
Box-Pierce Test
---------------
     Test based on first 24 autocorrelations
     Large sample test statistic = 19.4028
     P-value = 0.730183
```

2.4.2. Location Effect Analysis

For this analysis, there were three levels of the fixed effect "location" and 149 levels of the random effect "batch". For the location effect, 1 = beginning, 2 = middle, and 3 = end.

The location effect was not statistically significant at the $\alpha = 0.05$ level (prob > F = 0.1464).

Since there was not a statistically significant difference between the beginning, middle, and end of a batch, the data analysis that was conducted in Sections 3 and 4 did not distinguish between different portions of the batch.

The results of the analysis are detailed below.

Figure 2.4.2-1: SAS PROC MIXED Results for the Line 1 and 2 CU Data from 2001

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
LOC	2	294	1.93	0.1464

Least Squares Means

Effect	LOC	LSMEAN	Std Error	DF	t	Pr > |t|	Alpha	Lower	Upper
LOC	1	0.24806318	0.00059193	294	419.08	0.0001	0.05	0.2469	0.2492
LOC	2	0.24694932	0.00067008	294	368.54	0.0001	0.05	0.2456	0.2483
LOC	3	0.24641532	0.00067008	294	367.74	0.0001	0.05	0.2451	0.2477

Differences of Least Squares Means

Effect	LOC	_LOC	Difference	Std Error	DF	t	Pr > |t|	Adjustment	Adj P
LOC	1	2	0.00111385	0.00086752	294	1.28	0.2002	Tukey-Kramer	0.4054
LOC	1	3	0.00164786	0.00086752	294	1.90	0.0585	Tukey-Kramer	0.1406
LOC	2	3	0.00053401	0.00092262	294	0.58	0.5632	Tukey-Kramer	0.8316

Differences of Least Squares Means

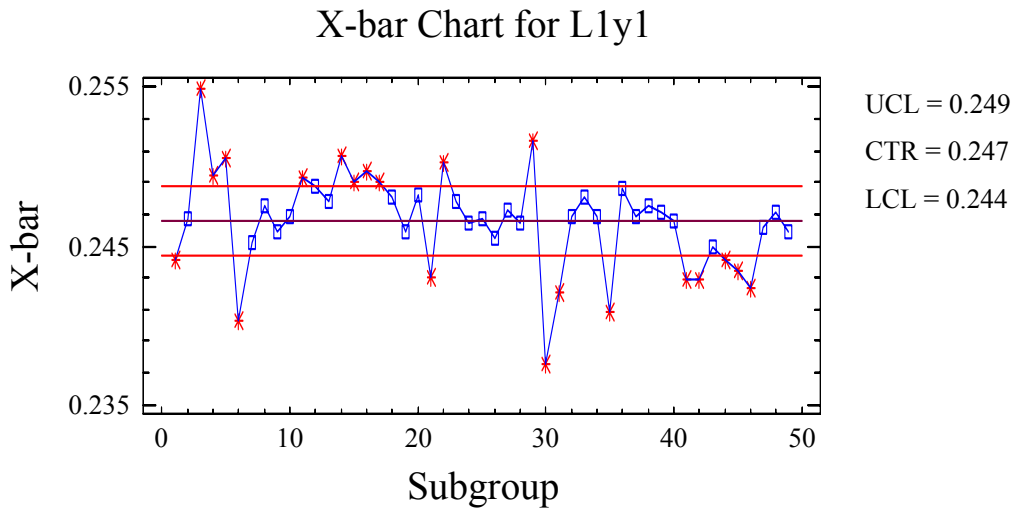
Alpha	Lower	Upper	Adj Low	Adj Upp
0.05	-0.0006	0.0028	-0.0009	0.0032
0.05	-0.0001	0.0034	-0.0004	0.0037
0.05	-0.0013	0.0023	-0.0016	0.0027

2.4.3. $\bar{x}$ and S Control Charts

$\bar{x}$ and S control charts are designed to assess whether subgrouped data come from a process which is in a state of statistical control. These control charts are constructed under the assumption that the data come from a normal distribution.

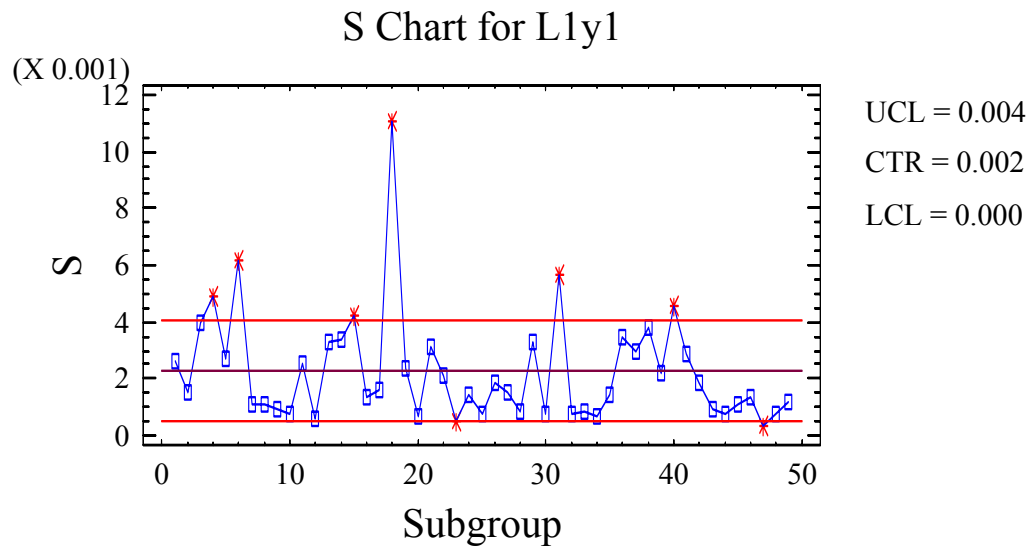
Below are the traditional $\bar{x}$ and S control charts for the uncorrelated Line 1 year 2001 content uniformity (CU) data, using batches as groups each containing ten observations. The charts were based on $\pm 3.29\sigma$ limits (0.0005 two-sided or 0.001 one-sided).

Figure 2.4.3-1: $\bar{x}$ Chart of the Line 1 Begin CU Data from 2001



In the $\bar{x}$ chart in Figure 2.4.3-1, 20 of the 49 points are beyond the control limits. Specifically, the process means go well outside the limits on four occasions, twice above and twice below. Many other means are close to the limits, both outside and inside. A possible explanation for the out of control signals is the result of mixing (addressed in more detail at the end of the section). The out of control points appear to be random with no systematic pattern.

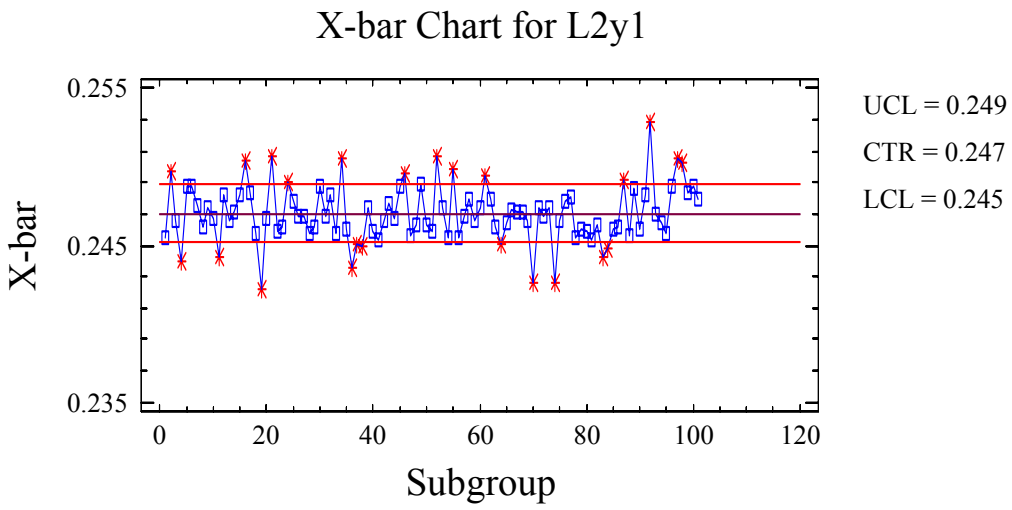
Figure 2.4.3-2: S Chart of the Line 1 Begin CU Data from 2001



In the S chart in Figure 2.4.3-2, 8 of the 49 points are beyond the control limits. Specifically, the process standard deviations go well outside the limits on three occasions

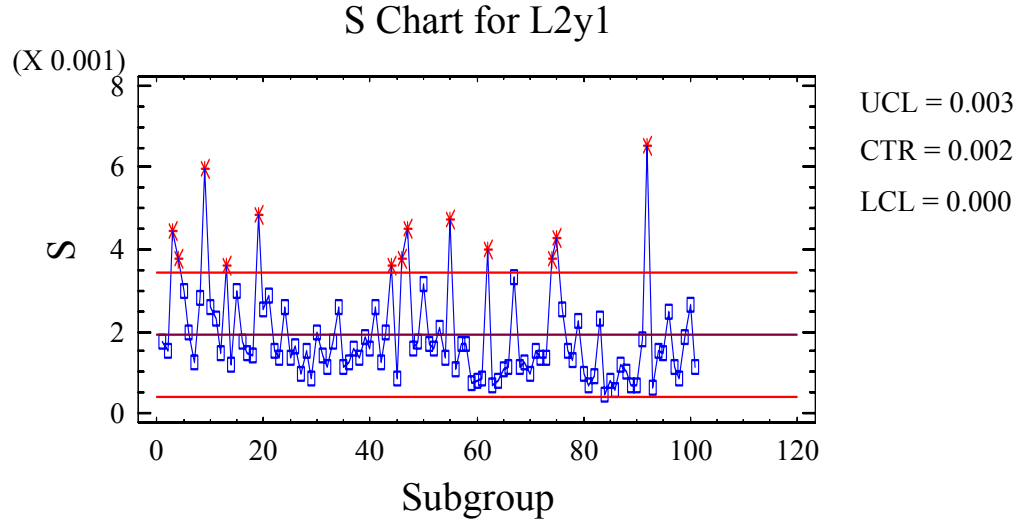
(one value is much larger than the others). Most other standard deviations are very low. The out of control points appear to be random with no systematic pattern.

Figure 2.4.3-3: $\bar{x}$ Chart of the Line 2 CU Data from 2001



In the $\bar{x}$ chart in Figure 2.4.3-3, 20 of the 101 points are beyond the control limits. Specifically, the process means go well outside the limits on one occasion (above). Many other means are close to the limits, both outside and inside. A possible explanation for the out of control signals is the result of mixing (addressed in more detail at the end of the section). The out of control points appear to be random with no systematic pattern.

Figure 2.4.3-4: S Chart of the Line 2 CU Data from 2001



In the S chart in Figure 2.4.3-4, 13 of the 101 points are beyond the control limits. Specifically, the process standard deviations go well outside the limits on seven occasions (two values are much larger than the others). Most other standard deviations are very low. The out of control points appear to be random with no systematic pattern.

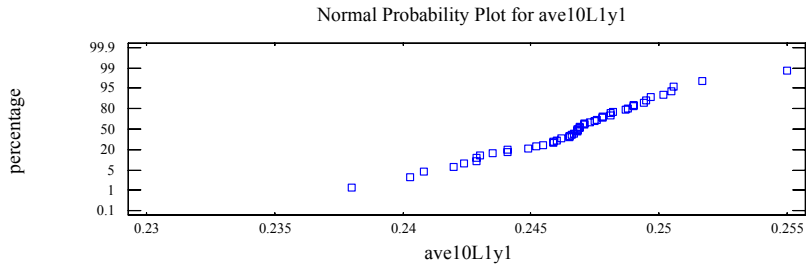
The $\bar{x}$ and S control charts indicate that the Line 1 and 2 processes in 2001 were not in statistical control. The patterns that are shown on the $\bar{x}$ charts may be symptomatic of mixing. Since there are 10 independent filling needles on the Blow-Fill-Seal Machine and samples of 4 (begin), 3 (middle), and 3 (end) are randomly collected for each batch, the performance of a specific filling needle is not being consistently represented. This phenomenon could be a subject for future study. Unfortunately, this issue could not be addressed appropriately in this document since the author did not have ownership of the process.

2.4.4. Individual Measurement Control Charts

The $\bar{x}$ and S control charts indicated that the Line 1 and 2 processes in 2001 were not in statistical control. The fact that between-batch variation was large relative to within-batch variation, and that the units were filled on multiple nozzles suggested that mixing was at least partly to blame for the poor patterns observed on the $\bar{x}$ charts. Since there was no record of which nozzles filled which of the tested units, a more appropriate method was sought for charting the process level. By eliminating within-batch variation as the measure of the size of between-batch variation, Individual Measurement Control Charts provide an alternative, and perhaps, fairer measure of process control. The X individual control charts are constructed under the assumption that the data are uncorrelated (previously verified in Section 2.4.1) and normally distributed (verified in this section). The charts were created by treating each batch as a single observation (e.g. the content uniformity values from begin, middle, and end portions of each batch were averaged to obtain one data point). The control procedure uses the moving range of two successive observations to estimate the process variability. The moving range (MR) is defined as $MR_i = \left|x_i - x_{i-1}\right|$ (Douglas C. Montgomery 1996).

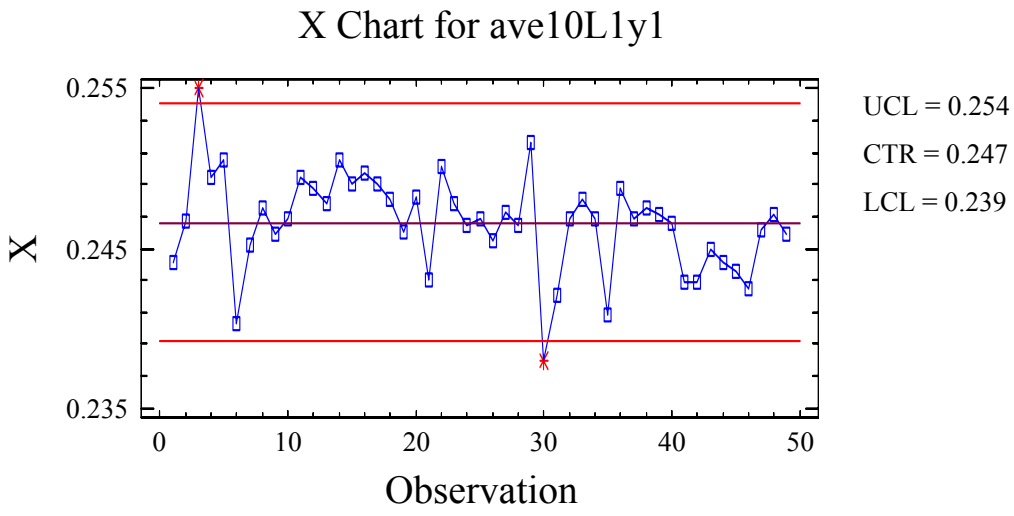
Below are the X individual control charts and MR(2) control charts for the uncorrelated Line 1 year 2001 content uniformity (CU) data. The charts were based on ± 3.29σ limits (0.0005 two-sided or 0.001 one-sided).

Figure 2.4.4-1: Normal Probability Plot for the Line 1 CU Data from 2001



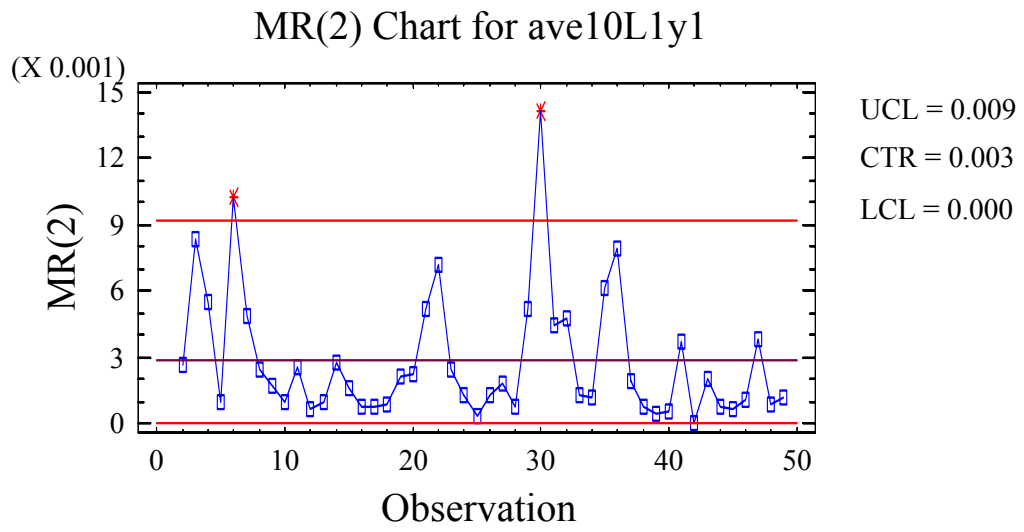
The normal probability plot indicates that normality is a reasonable assumption for these data. Based on the Shapiro-Wilks Test for normality, we cannot reject the hypothesis that the data comes from a normal distribution (p = 0.64).

Figure 2.4.4-2: X Individual Chart for the Line 1 CU Data from 2001



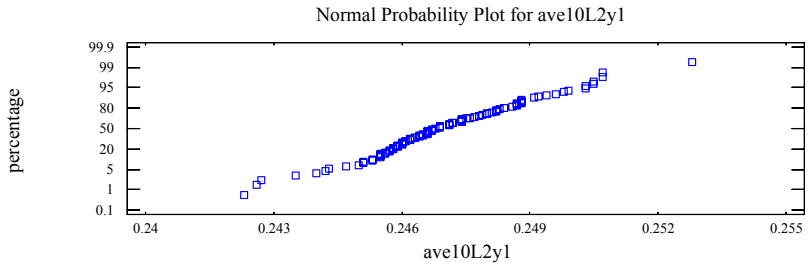
In the X Individuals Chart in Figure 2.4.4-2, 2 of the 49 points are slightly beyond the control limits. Specifically, the process means are mostly situated along the center line. Very few points are close to the control limits.

Figure 2.4.4-3: MR(2) Individual Chart for the Line 1 CU Data from 2001



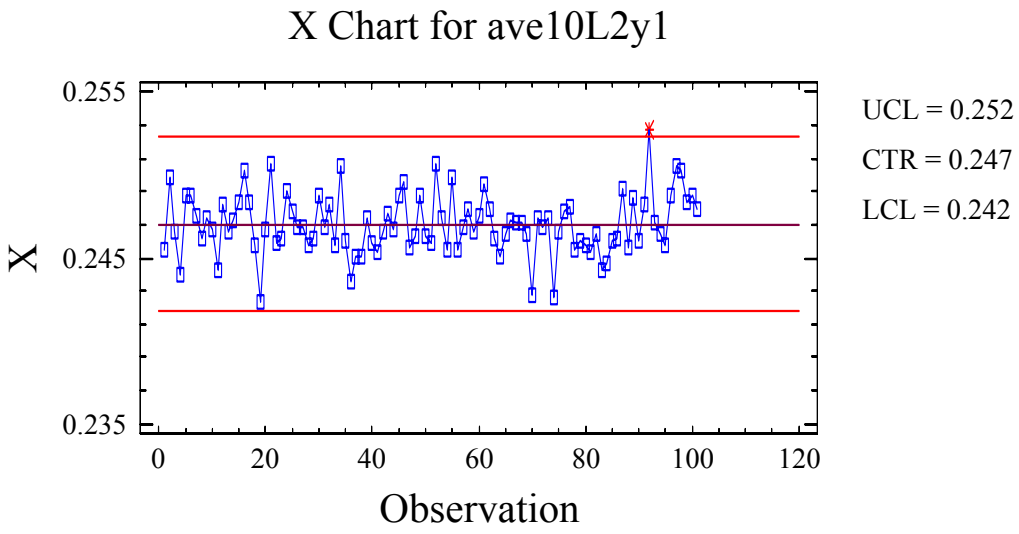
In the MR(2) chart in Figure 2.4.4-3, 2 of the 49 points are beyond the control limits. One point is slightly beyond and one point is far beyond the upper control limit. Specifically, the moving ranges are mostly situated along or below the center line. Very few points are close to the control limits.

Figure 2.4.4-4: Normal Probability Plot for the Line 2 CU Data from 2001



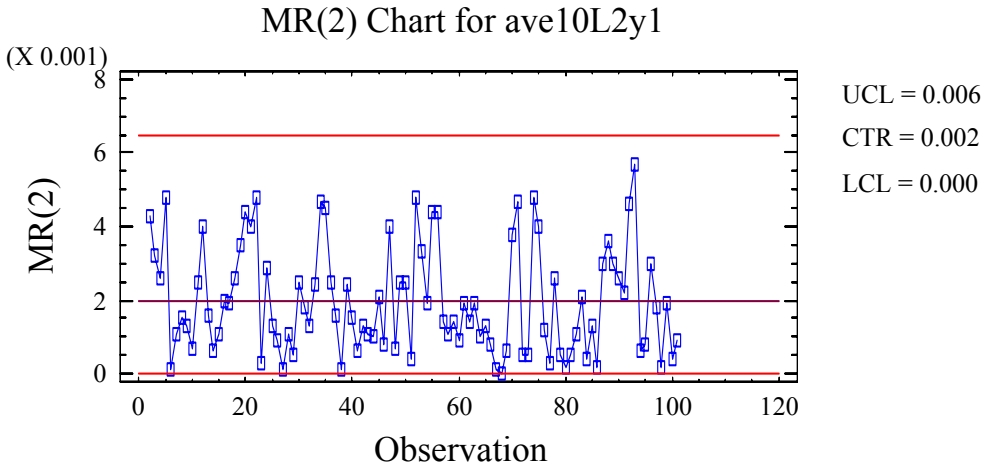
The normal probability plot indicates that normality is a reasonable assumption for these data. Based on the Shapiro-Wilks Test for normality, we cannot reject the hypothesis that the data comes from a normal distribution (p = 0.59).

Figure 2.4.4-5: X Individual Chart for the Line 2 CU Data from 2001



In the X Individuals Chart in Figure 2.4.4-5, 1 of the 101 points is slightly beyond the control limits. Specifically, the process means are mostly situated along the center line. Very few points are close to the control limits.

Figure 2.4.4-6: MR(2) Individual Chart for the Line 2 CU Data from 2001



In the MR(2) chart in Figure 2.4.4-6, 0 of the 49 points are beyond the control limits. Specifically, the moving ranges are mostly situated along or below the center line. One point is fairly close to the upper control limit.

Although the X and MR(2) individual control charts have wider limits and only a few points were outside the limits on the various charts, these control charts reaffirm that the Line 1 and 2 processes in 2001 were not in statistical control.

2.4.5. Modified Control Limits

Despite the fact that traditional $\bar{x}$ and S control charts and Individual Measurement control charts indicated that the Line 1 and 2 processes in 2001 were not in statistical control, the natural variability of the processes is much smaller than the range of the specification limits. 0.237 – 0.262 mg/mL. In such situations, Modified Control Charts are appropriate for process monitoring if some drift in the process mean is acceptable. The modified control charts are constructed under the assumption that the data are uncorrelated and normally distributed. The modified $\bar{x}$ control chart is concerned only with detecting whether the true process mean μ is located such that the process is producing a fraction non-conforming in excess of some specified value δ. In practice, μ is allowed to vary over an interval $\mu_L \leq \mu \leq \mu_U$, whose endpoints are chosen as the smallest and largest permissible values of μ, consistent with producing a fraction non-conforming of at most δ (Douglas C. Montgomery 1996).

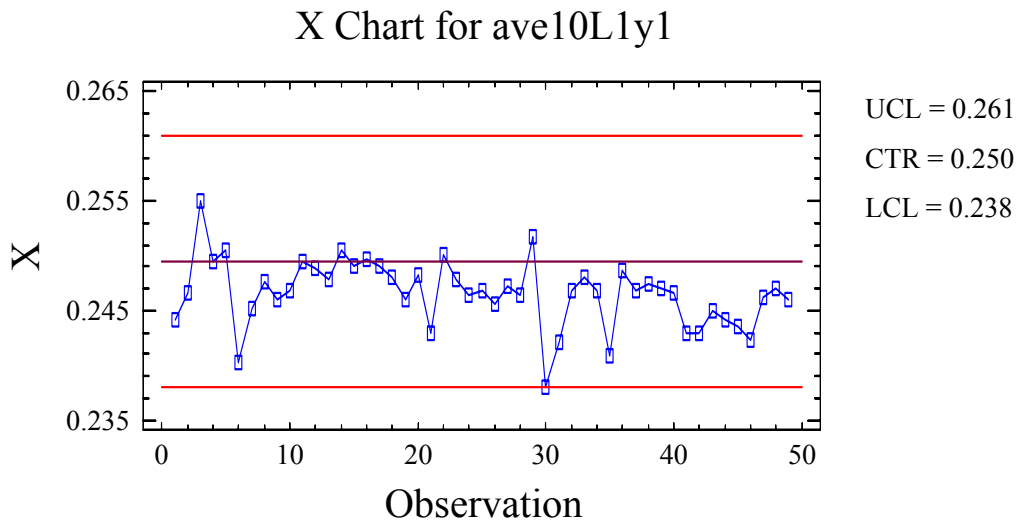
For consistency with the previously displayed control charts ($\bar{x}$ and S charts and X Individuals charts), the modified control limits were based on ± 3.29σ limits (0.0005 two-sided or 0.001 one-sided). The control limits were derived with the following formula:

$$UCL = USL - (z_\delta - 3.29/n\text{^}1/2)\sigma \text{ and } LCL = LSL + (z_\delta - 3.29/n\text{^}1/2)\sigma,$$

Where $\delta = 0.01$ (process fraction non-conforming),
$z = 2.33$,
n = subgroup size,
σ = process standard deviation

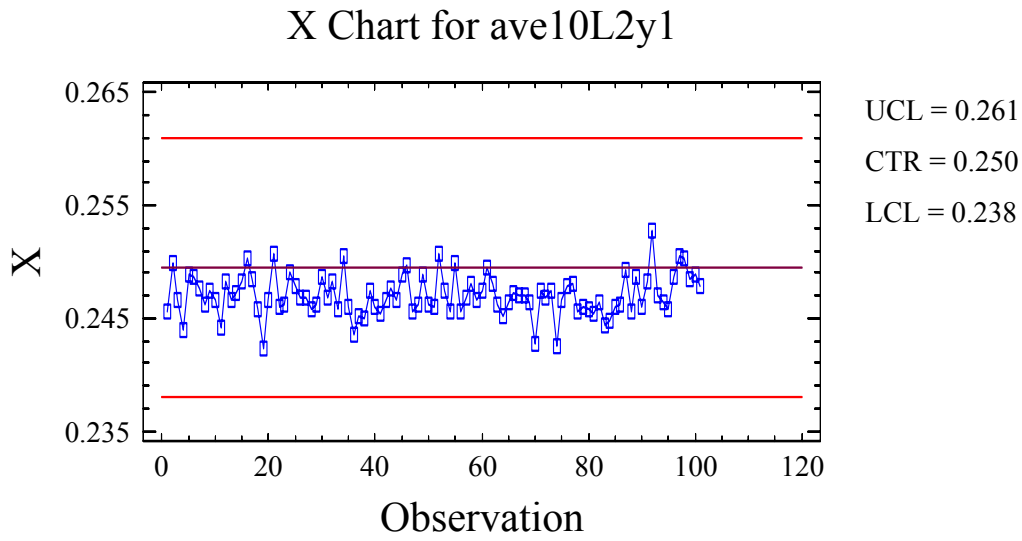
Below are the results obtained by applying modified control limits to the content uniformity Line 1 and 2 data for 2001.

Figure 2.4.5-1: Modified $\bar{x}$ Chart for the Line 1 CU Data from 2001



In the Modified $\bar{x}$ chart in Figure 2.4.5-1, none of the 49 points are beyond the control limits. Specifically, the process means are mostly situated along or below the center line. Very few points are close to the control limits. There is one point which is very close to the lower control limit.

Figure 2.4.5-2: Modified $\bar{x}$ Chart for the Line 2 CU Data from 2001



In the Modified $\bar{x}$ chart in Figure 2.4.5-2, 0 of the 101 points are beyond the control limits. Specifically, the process means are mostly situated along or below the center line. No points are close to the control limits.

The Line 1 and 2 processes in 2001 were in statistical control with respect to the modified $\bar{x}$ control charts.

3. Equivalence Testing

3.1. Objective

The objective of this analysis was to show statistical equivalence between the means of active ingredient assay values of batches manufactured on Line 1 and Line 2. No distinction was made in the active ingredient assay data between the two actives compounding tanks for each line. Based on the results of the PROC MIXED SAS procedure in Section 2.4.2, the location was not factored into the analysis.

3.2. Data

Active ingredient assay data for individual units (used to determine content uniformity) from the year 2001 on Lines 1 and 2 were collected.

Content uniformity testing is reported in the batch record as 10 individual active ingredient content values.

3.3. Procedure (Bolton 1990)

In order to compare the active ingredient assay data from Lines 1 and 2, an equivalence study was conducted using the following steps:

1) Calculate summary statistics of the batches manufactured in 2001 for Line 1 (L_1) and Line 2 (L_2). Use these as estimates for the mean ($\bar{x}$ **)** and standard deviation (s).
2) Display data graphically to get an idea of the distribution. Verify the assumption of normally distributed data.
3) Select α, the probability of committing a type I error (rejecting a true null hypothesis, claim the lines have equivalent active ingredient assay values when they do not) for the study.
4) Select β, the probability of committing a type II error (failing to reject a false null hypothesis, claim the lines do not have equivalent active ingredient assay values when they do) for the study.
5) Select δ, the equivalence margin, for the study.
6) Calculate the sample size required (per group) to achieve the desired level of power for the study. For an equivalence study (based on a two-sided 100 x (1-2α)% confidence interval), the sample size required per group ($n_1 = n_2$) is:

$$n_2 = 2[(Z_{1-\alpha} + Z_{1-\beta/2})(s/\delta)]^2$$

$$\text{Power: } 1 - \beta = 2\,\Phi[\delta(2s^2/n_2)^{-1/2} - Z_{1-\alpha}] - 1,$$

$$\text{when } \mu_1 = \mu_2.$$

7) Calculate the two-sided 100 x (1-2α)% confidence interval for $\mu_1 - \mu_2$ (since the population variances are unknown and sample sizes are large):

$$(\bar{x}_1 - \bar{x}_2) \pm z_{1-\alpha}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

8) Conclude equivalence if the two-sided 100 x (1-2α)% confidence interval for $\mu_1 - \mu_2$, $[C\mu_1, C\mu_2]$ is in $[-\delta, \delta]$ or $[-\delta < C\mu_1 < C\mu_2 < \delta]$, where $C\mu_1$ and $C\mu_2$ are the lower and upper limits of the confidence interval, respectively.

3.4. Analysis

Content uniformity data from 0.5 mg strength production batches manufactured during 2001 were used for the analysis since the process was in statistical control during this time period. These values were most representative of individual units as the values were not averaged. Content uniformity testing is reported in the batch record as individual active ingredient content values to 4 decimal places.

3.4.1. Equivalence Study: Lines 1 and 2

Table 3.4.1-1: Summary Statistics for Line 1 and 2 CU Data from 2001

Summary Statistics

	L1y1	L2y1
Count	490	1010
Average	0.246572	0.247025
Median	0.24665	0.2469
Variance	0.0000173501	0.0000080143
Standard deviation	0.00416534	0.00283095
Standard error	0.000188171	0.0000890784
Minimum	0.2258	0.2286
Maximum	0.2603	0.264
Range	0.0345	0.0354

An α value = 0.05, which is the probability of committing a type I error, was selected for the study. A δ value = 0.0015, which is the equivalence margin, was selected for the study.

Based on the summary statistics, the two-sided 100 x (1-2α)% confidence interval for $\mu_1 = \mu_2$ (assuming population variance is known) where,

μ_1 = mean of Line 1
μ_2 = mean of Line 2

was calculated using the following formula:

$$(\bar{x}_1 - \bar{x}_2) \pm z_{1-\alpha}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The resulting confidence interval for the difference in the content uniformity means is (-0.0014, 0.0006). Since (-0.0014, 0.0006) is in the interval (-0.0015, 0.0015), equivalence of the means is concluded.

4. Control Charts For SNR Values of the Content Uniformity Assay Data

4.1. Objective

The objective of this analysis was to establish control charts for the signal-noise ratio (SNR) values of the content uniformity assay data. Distinction was made in the active ingredient assay data between the two actives compounding tanks for each line. Normal practice at company is to calculate the relative standard deviation, $s/\bar{x}$ (specification is NMT 6.0%) for the content uniformity data since this statistical acceptance criterion is referenced in the USP. Rather than looking at the relative standard deviation, for this project, its reciprocal, the SNR ($\bar{x}$ / s) was analyzed since this quantity has a more desirable distribution (non-central t). Although SNR values may be charted by Individuals Measurement charts, a new control chart, which we call the SNR Chart, was developed specifically to monitor SNR data. No description of the SNR Chart has been located in related literature.

4.2. Data

Content uniformity testing is currently reported in the batch record as 10 individual active ingredient content values (4 begin, 3 middle, and 3 end samples).

The analysis detailed below focused on batch production data for Line 2 during the beginning of 2001 where 4 begin, 3 middle, and 3 end content uniformity values were reported in the batch records. For each group of 10 content uniformity values, the relative standard deviation ($s / \bar{x}$) is calculated. Since the location effect analysis in Section 2.4.2 determined that there was not statistical difference between the begin, middle, and end content uniformity data, the data was combined for the analysis.

4.3. Procedure

The SNR $\sqrt{n}\,(\bar{x}$ / s) was assumed to be non-centrally t distributed based on the following:

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$$

and

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

therefore,

$$\sqrt{n}\,\frac{\bar{x}}{s} \sim \sqrt{n}\,\frac{N(\mu,\frac{\sigma^2}{n})}{\sigma\sqrt{\frac{\chi^2_{n-1}}{n-1}}} \sim \frac{N(\delta,1)}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} \sim t_{n-1,\delta} \quad \text{where}$$

t is the non-central t distribution with n-1 degrees of freedom,

and δ is the non-centrality parameter, $\delta = \frac{\sqrt{n}\mu}{\sigma}$

An interesting property of the SNR is that the process may change in such a way that the value of δ remains unchanged. Therefore, while changes in δ imply changes in the process, changes in level and spread in the original measurements do not necessarily translate into changes in the distribution of the SNR.

The non-central t distribution has two parameters (Evans et al., 1993): the degrees of freedom, f (a positive integer), and the non-centrality parameter, δ (a real number). It is denoted by t: f, δ. Its probability density function (p.d.f.) (Lehmann, 1959) can be expressed as:

$$p(x; f, \delta) = \frac{1}{2^{(f+1)/2}\Gamma(f/2)\sqrt{\pi f}} \int_0^\infty y^{(f-1)/2} \exp\left[-\frac{1}{2}y - \frac{1}{2}(x\sqrt{\frac{y}{f}} - \delta)^2\right] dy$$

Its rth moment about the origin is:

$$\mu_r^{'} = (\frac{f}{2})^{r/2} \frac{\Gamma((f-r)/2)}{\Gamma(f/2)} \sum_{j=0}^{r/2} \binom{r}{2j} x \frac{(2j)!}{2^j\, j!} \delta^{r-2j} \qquad (f > r)$$

To simplify the notation, let $g(f) = \frac{\Gamma((f-1)/2)}{\Gamma(f/2)}$

Using these formulae, we can write the mean $\mu(f,\delta)$, variance $\sigma^2(f,\delta)$, and coefficient of skewness $\eta_3(f,\delta)$ of the t: f, δ distribution as:

$$\mu(f,\delta) = \sqrt{\frac{f}{2}}\, g(f)\delta \quad (f > 1), \qquad (1)$$

$$\sigma^2(f,\delta) = \frac{f}{f-2}\left[1+\delta^2\left(1-\frac{f-2}{2}g^2(f)\right)\right] \quad (f > 2), \qquad (2)$$

$$\eta_3(f,\delta) = \sqrt{\frac{f-2}{2}} x \frac{g(f)\delta\left[3+(f-2)\delta^2((f-3)g^2(f)-(2f-7)/(f-2))\right]}{(f-3)\left[1+\delta^2(1-(f-2)/2g^2(f))\right]^{3/2}} \quad (f > 3)$$

Three different methods for estimating the non-centrality parameter δ were considered First, δ was estimated using equation (1) and the mean of the signal to noise ratio values ($\bar{x}_{SNR}$).

$$\hat{\delta} = \frac{\bar{x}_{SNR}\sqrt{n}}{g(f)\sqrt{\frac{f}{2}}} \qquad (3)$$

Second, δ was estimated using equation (2) and the variance of the signal to noise ratio values (s^2_{SNR}).

$$\hat{\delta} = \sqrt{\frac{ns^2_{SNR}-(\frac{f}{f-2})}{(\frac{f}{f-2})(1-(\frac{f}{f-2})g^2(f))}} \qquad (4)$$

Third, the non-centrality parameter (δ) was estimated using the following equation:

$$\hat{\delta} = \frac{\bar{\bar{x}}\sqrt{n}}{\bar{s}}, \qquad (5)$$

where

$\bar{\bar{x}}$ is the mean of the batch means and
$\bar{s}$ is the mean of the batch standard deviations.

4.4. Determining the Best Method for Estimating the Non-centrality Parameter

In order to ascertain the best of these three estimation methods, a small simulation study was conducted. The SAS macro *simulation* detailed in Appendix 1 was used to simulate subsets of data from a normal distribution.

Since data analysis will involve the first 25 batches manufactured during the year 2001 for Line #2 (Tanks V-211 and V-212) and 10 data points from each batch will be analyzed, the SAS macro *simulation* was run to create 1000 data sets, each consisting of 25 samples of size of 10. The mean was taken to be 0.248 and standard deviation to be 0.003, the values for the actual data from Line #2.

For each data set, each of the 3 different methods given in (3)-(5) was used to estimate the non-centrality parameter δ, which in turn was used to calculate the level 0.001 critical value of the non-central t distribution. The ($\sqrt{n}\ \bar{x}\ /$ s) values were then compared to the level 0.001 critical values to obtain false-alarm rates. Tables 4.4-1 – 4.4-3 show the results of three runs of the simulation. Delta1, Delta2, and Delta3 refer to non-centrality parameter estimates calculated from equations (3), (4), and (5), respectively.

Table 4.4-1: Results from Simulation #1

	0.001 Critical Value	**# of Minimum Values Below 0.001 Critical Value**	$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
Delta 1	148.4	24	(0.015, 0.033)
Delta 2	144.3	13	(0.006, 0.020)
Delta 3	148.8	25	(0.015, 0.035)

Table 4.4-2: Results from Simulation #2

	0.001 Critical Value	**# of Minimum Values Below 0.001 Critical Value**	$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
Delta 1	148.7	20	(0.011, 0.029)
Delta 2	144.9	11	(0.005, 0.017)
Delta 3	149.1	20	(0.011, 0.029)

Table 4.4-3: Results from Simulation #3

	0.001 Critical Value	**# of Minimum Values Below 0.001 Critical Value**	$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
Delta 1	148.4	20	(0.011, 0.029)
Delta 2	144.5	11	(0.005, 0.017)
Delta 3	148.9	25	(0.015, 0.035)

The expected number of values below the critical value would be 25 since 1000 sets of 25 were compared to the 0.001 critical value of the non-central t distribution.

After comparing the results from the 3 simulations, Delta2 appears clearly inferior to either Delta1 or Delta3. Of the remaining two, Delta3 performed marginally better. Therefore, Delta3 was used to estimate δ in all subsequent calculations.

4.5. Analysis of SNR Data

4.5.1. Tank V-211

Begin, middle, and end data from the first 25 batches manufactured during the year 2001 for Line #2 (Tank V-211) were analyzed. For each group of 10 content uniformity values from a single batch, the quantity SNR= ($\sqrt{n}\ \bar{x}\ /$ s) was calculated. The summary statistics for the SNR values are detailed below:

Table 4.5.1-1: Summary Statistics – Tank V-211 SNR Values in 2001

```
Summary Statistics for v211bmey01snrn

Count = 25
Average = 590.277
Median = 575.256
Mode =
Geometric mean = 572.288
Variance = 22320.5
Standard deviation = 149.4
Standard error = 29.8801
Minimum = 319.399
Maximum = 928.657
Range = 609.258
Lower quartile = 499.389
Upper quartile = 635.998
Interquartile range = 136.61
Skewness = 0.549081
Stnd. skewness = 1.12081
Kurtosis = 0.444296
Stnd. kurtosis = 0.453458
Coeff. of variation = 25.3102%
Sum = 14756.9


Percentiles for v211bmey01snrn

1.0% = 319.399
5.0% = 337.91
10.0% = 440.081
25.0% = 499.389
50.0% = 575.256
75.0% = 635.998
90.0% = 832.105
95.0% = 889.214
99.0% = 928.657
```

Figure 4.5.1-1: Box and Whisker Plot for Tank V-211 SNR Values in 2001

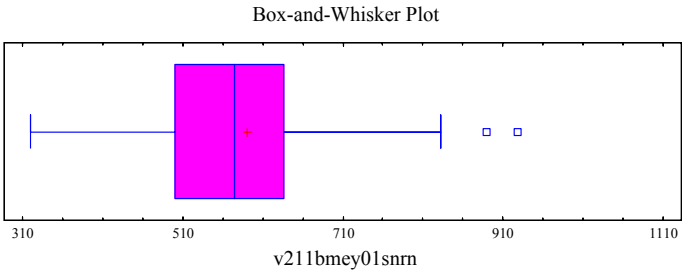


Figure 4.5.1-2: Histogram for Tank V-211 SNR Values in 2001

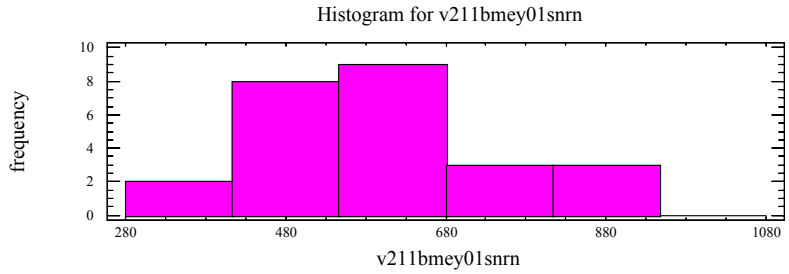
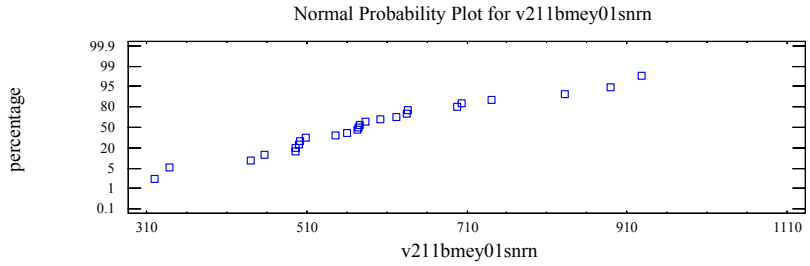


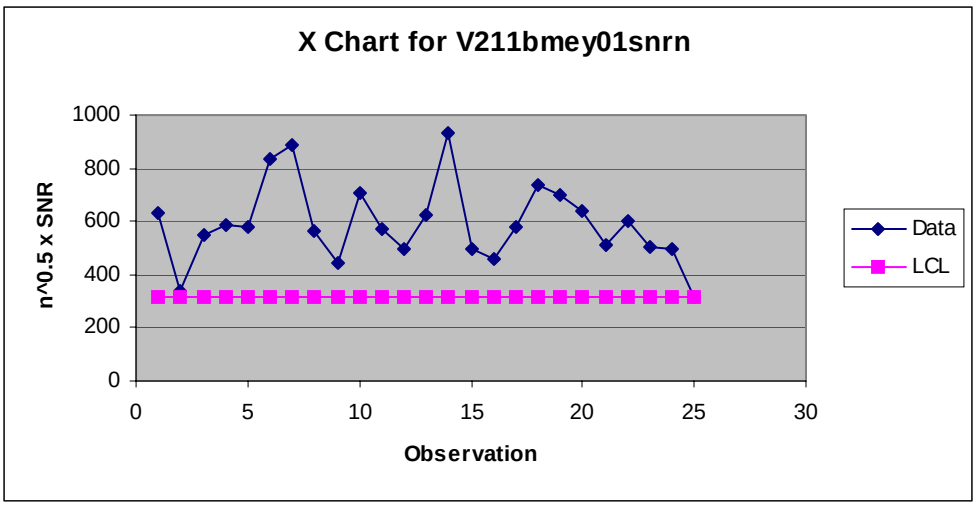
Figure 4.5.1-3: Normal Probability Plot for Tank V-211 SNR Values in 2001



The normal probability plot of the data below indicates non-normality (as expected).

Delta3 was used to calculate the 0.001 critical value (δ = 554.2, 9 degrees of freedom) of 314.9 for the SNR chart displayed in Figure 4.5.1-4.

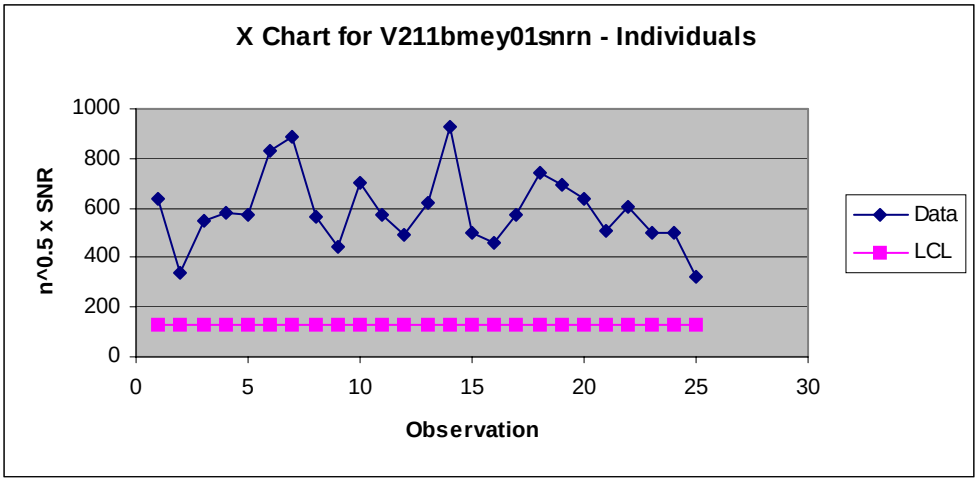
Figure 4.5.1-4: SNR Chart for Tank V-211 SNR Values in 2001



All SNR values were above the one-sided lower control limit (level 0.001 critical value = 314.9) indicating that the process is in control. Values ranged from 319 to 929.

An Individuals Control Chart for the same SNR values charted in Figure 4.5.1-4 is shown in Figure 4.5.1-5. The lower 0.001 control limit was calculated to be 127.1 based on the summary statistics of the SNR values ($\hat{\mu}$ = 590.28 and $\hat{\sigma}$ = 149.4). The lower 0.001 control limit was calculated by assuming normality (e.g. limit = $\hat{\mu} - 3.1\hat{\sigma}$).

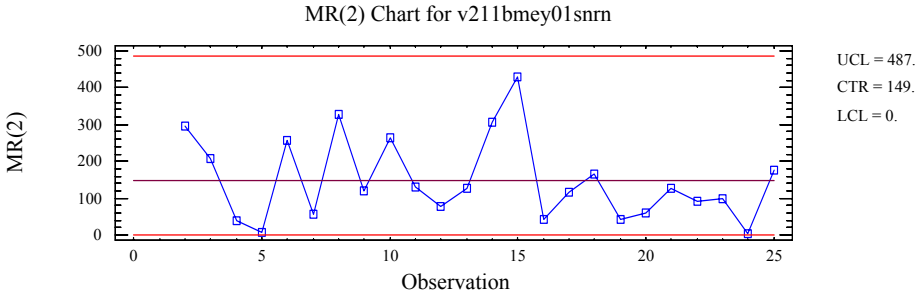
Figure 4.5.1-5: Individuals Chart for Tank V-211 SNR Values in 2001



All SNR values were above the one-sided lower control limit indicating that the process is in control. Values ranged from 319 to 929. The limit in this chart is smaller than the limit in the SNR chart, indicating less sensitivity to out of control signals.

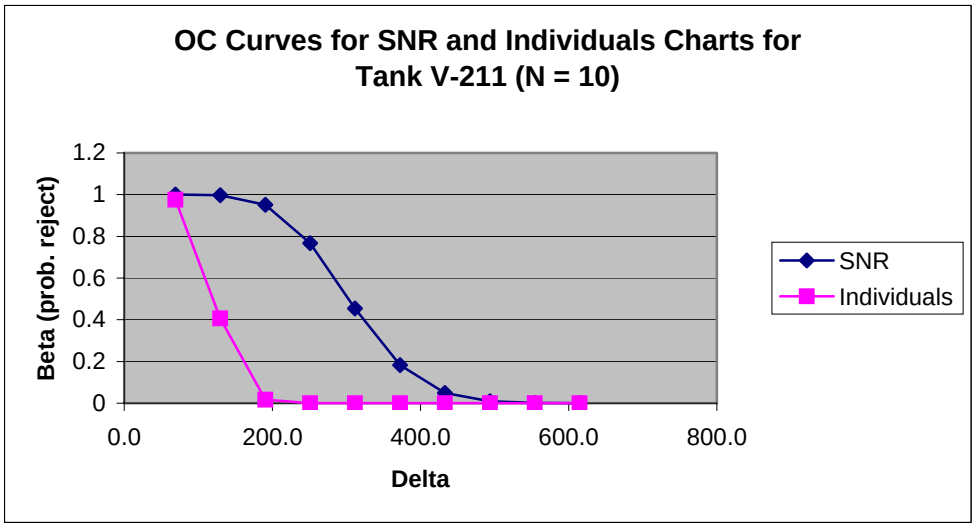
The chart below (Figure 4.5.1-6) is an individuals MR(2) chart for the SNR values with control limits based on +/- 3 σ limits of the range between two consecutive values. All MR(2) values plotted within the control limits indicating that the process is in control.

Figure 4.5.1-6: MR(2) Chart for Tank V-211 SNR Values in 2001



In order to compare the performance of the SNR and individuals charts, the plot below (Figure 4.5.1-7) shows Operating Characteristic (OC) curves for the SNR and individuals control charts shown in Figures 4.5.1-4 and 4.5.1-5. The curves estimate the probability of the SNR values plotting outside the control limits for a given value of δ.

Figure 4.5.1-7: OC Curves for SNR and Individuals Charts for Tank V-211 (N=10)



The data resulted in a non-centrality parameter of 554.2 and a 0.001 critical value of 314.9 (9 degrees of freedom).

Based on a review of the chart above, the SNR Chart OC Curve displays superior performance to the Individuals Chart OC Curve. For smaller values of δ, the SNR chart has a much higher probability of detecting out of control data points. The Individuals Chart has an extremely low probability (near 0) of detecting an out of control signal for δ values as low as 200. In contrast, the SNR chart has an extremely high probability (near 1) of detecting an out of control signal for δ values as low as 200.

4.5.2. Tank V-212

Begin, middle, and end data from the first 25 batches manufactured during the year 2001 for Line #2 (Tank V-212) were analyzed. For each group of 10 content uniformity values from a single batch, the quantity ($\sqrt{n}\ \bar{x}\ /$ s) was calculated. The summary statistics for the quantity ($\sqrt{n}\ \bar{x}\ /$ s) are detailed below:

Table 4.5.2-1: Summary Statistics – Tank V-212 SNR Values in 2001

```
Summary Statistics for v212bmey01snrn

Count = 25
Average = 442.019
Median = 446.877
Mode =
Geometric mean = 428.717
Variance = 12364.0
Standard deviation = 111.194
Standard error = 22.2387
Minimum = 275.338
Maximum = 701.081
Range = 425.743
Lower quartile = 375.703
Upper quartile = 510.089
Interquartile range = 134.386
Skewness = 0.455695
Stnd. skewness = 0.930184
Kurtosis = 0.103541
Stnd. kurtosis = 0.105676
Coeff. of variation = 25.1559%
Sum = 11050.5



Percentiles for v212bmey01snrn

1.0% = 275.338
5.0% = 297.601
10.0% = 298.364
25.0% = 375.703
50.0% = 446.877
75.0% = 510.089
90.0% = 556.276
95.0% = 666.406
99.0% = 701.081
```

Figure 4.5.2-1: Box and Whisker Plot for Tank V-212 SNR Values in 2001

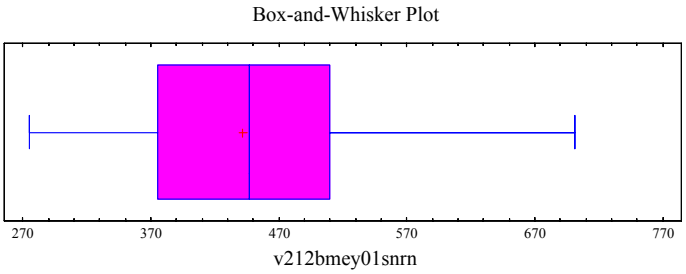


Figure 4.5.2-2: Histogram for Tank V-212 SNR Values in 2001

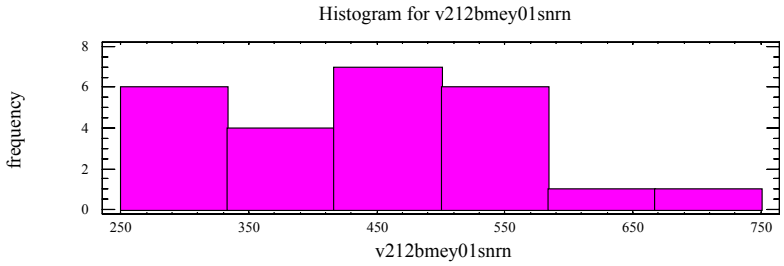
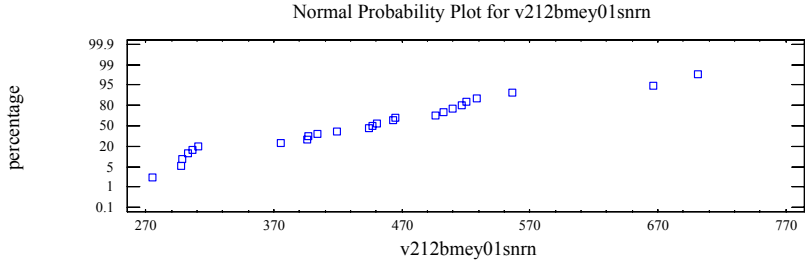


Figure 4.5.2-3: Normal Probability Plot for Tank V-212 SNR Values in 2001

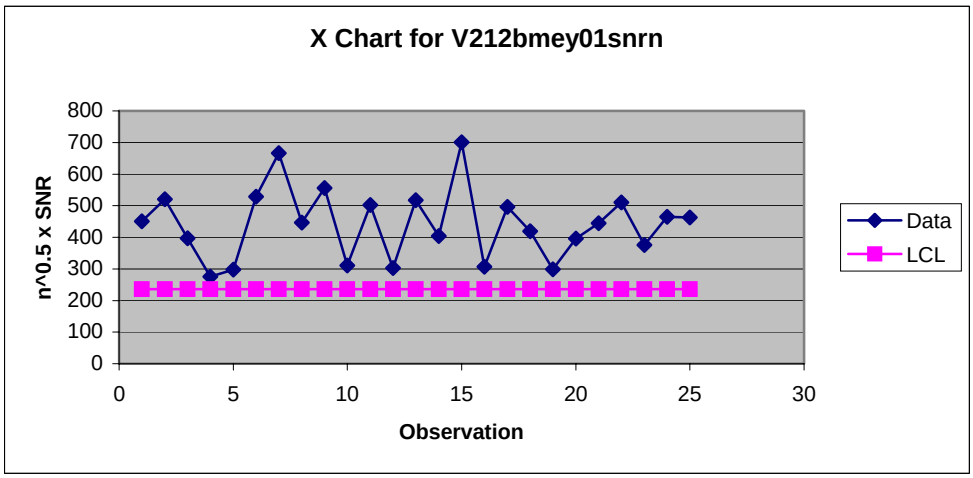


The normal probability plot of the data below indicates non-normality (as expected).

Delta3 was used to calculate the 0.001 critical value (δ = 415.6, 9 degrees of freedom) of 236.1 for the control charts for these data.

The SNR chart in Figure 4.5.1-4 plots the quantity $\sqrt{n}\ \bar{x}\ /\ s$ from the begin, middle, and end data from the first 25 batches manufactured during the year 2001 for Line #2 (Tank V-212):

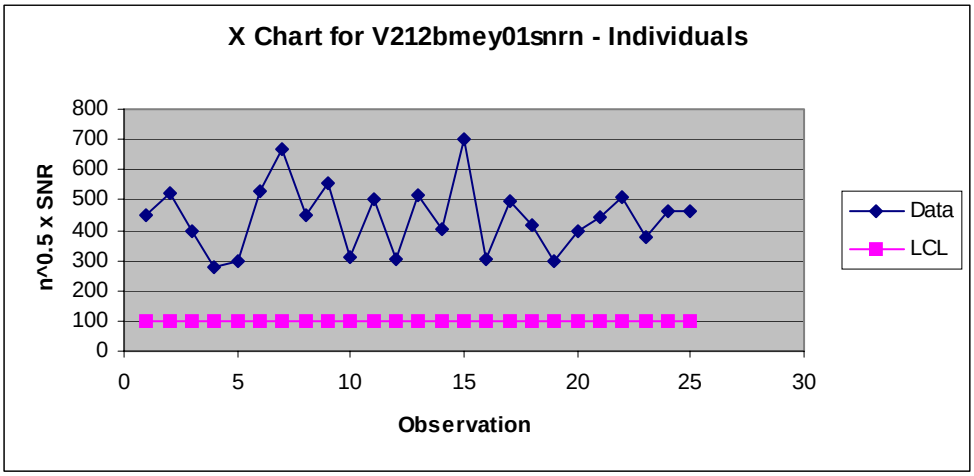
Figure 4.5.2-4: SNR Chart for Tank V-212 SNR Values in 2001



All $\sqrt{n}\ \bar{x}\ /\ s$ values were above the one-sided lower control limit (level 0.001 critical value = 236.1) indicating that the process appears to be in control. Values ranged from 275 to 701.

An Individuals Control Chart for the same SNR values charted in Figure 4.5.2-4 is shown in Figure 4.5.2-5. The lower 0.001 control limit was calculated to be 97.3 based on the summary statistics of the SNR values. The lower 0.001 control limit was calculated by assuming normality (e.g. limit = $\hat{\mu} - 3.1\,\hat{\sigma}$).

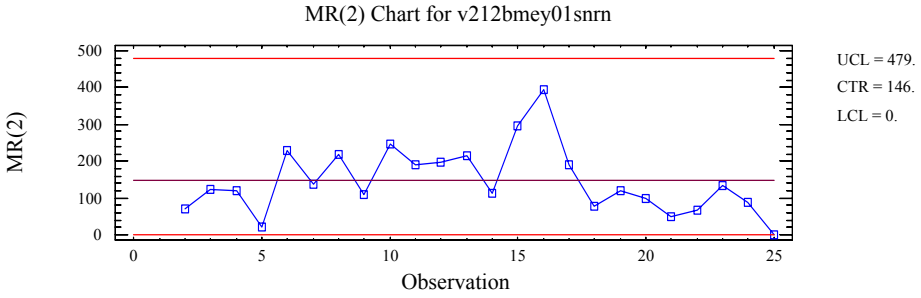
Figure 4.5.2-5: Individuals Chart for Tank V-212 SNR Values in 2001



All $\sqrt{n}\ \bar{x}\ /\ s$ values were above the one-sided lower control limit (level 0.001 critical value = 97.3) indicating that the process appears to be in control. Values ranged from 275 to 701. The limit in this chart is smaller than the limit in the SNR chart, indicating less sensitivity to out of control signals.

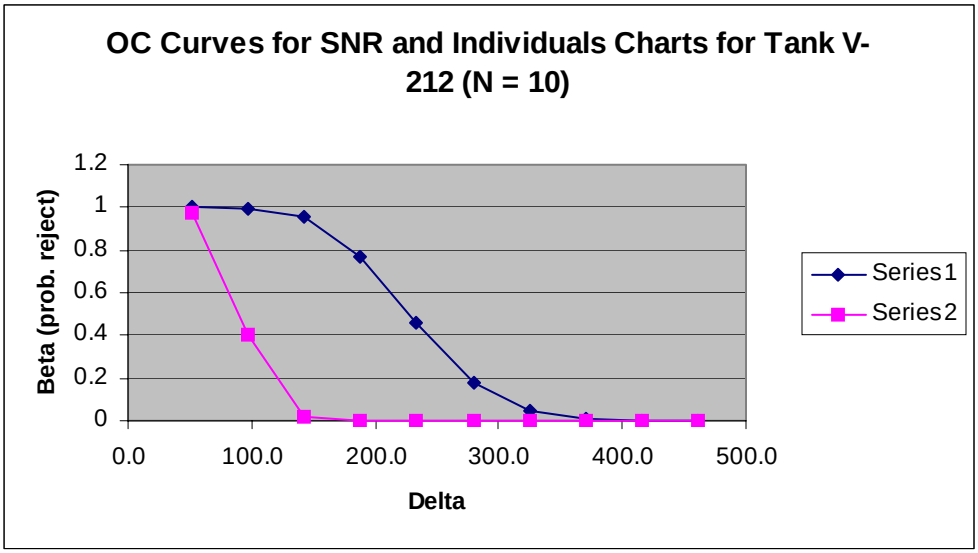
The chart below is an individuals MR(2) chart with control limits based on +/- 3 σ limits of the range between two consecutive values. All MR(2) values plotted within the control limits indicating that the process is in control.

Figure 4.5.2-6: MR(2) Chart for Tank V-212 SNR Values in 2001



In order to compare the performance of the SNR and individuals charts, the plot below (Figure 4.5.2-7) shows Operating Characteristic (OC) curves for the SNR and individuals control charts shown in Figures 4.5.2-4 and 4.5.2-5. The curves estimate the probability of the SNR values plotting outside the control limits for a given value of δ.

Figure 4.5.2-7: OC Curves for SNR and Individuals Charts for Tank V-211 (N=10)



The data resulted in a non-centrality parameter of 415.6 and a 0.001 critical value of 236.1 (9 degrees of freedom). This point is plotted on the chart and indicates that there is a 0.001 probability of the process (with a non-centrality parameter of 415.6) plotting below the control limit.

Based on a review of the chart above, the SNR Chart OC Curve displays superior performance to the Individuals Chart OC Curve. For smaller values of δ, the SNR chart has a much higher probability of detecting out of control data points. The Individuals Chart has an extremely low probability (near 0) of detecting an out of control signal for δ values as low as 150. In contrast, the SNR chart has an extremely high probability (near 1) of detecting an out of control signal for δ values as low as 150.

4.6. Performance of the SNR Chart

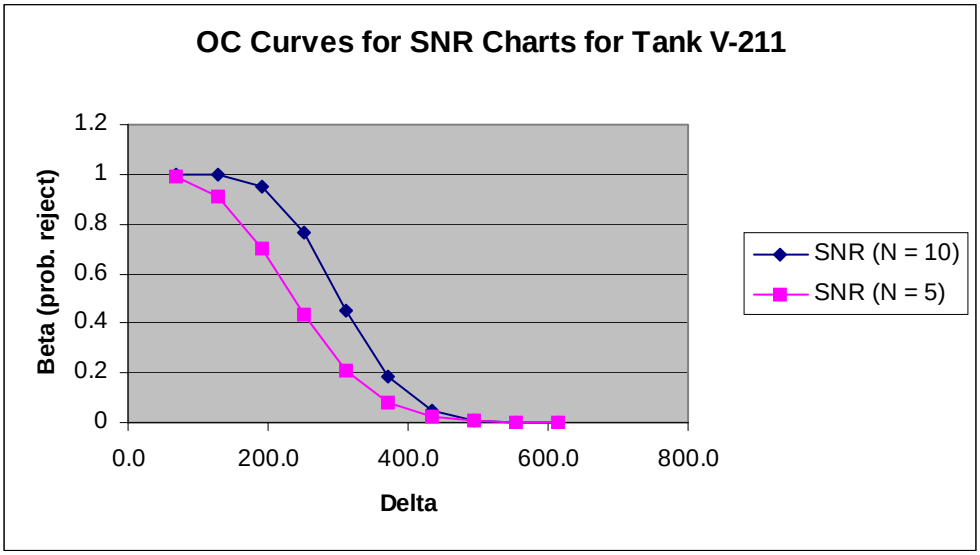
4.6.1. Comparison of the SNR Chart to the Individuals Chart

Recall Figure 4.5.1-7 which shows the OC curves for the SNR and individuals control charts based on the Tank V-211 data. The chart indicates that the SNR Chart is superior to the Individuals Chart.

4.6.2. Performance of the SNR Chart (N=10 vs N=5)

To further explore the performance of the SNR chart, OC curves were computed for SNR data from samples of size 5. Figure 4.6-1 shows the comparison of the SNR Chart for samples sizes of 5 and 10. As expected, the SNR Chart with N=10 displays superior performance to the SNR Chart (N = 5).

Figure 4.6.2-1: OC Curves for SNR Charts for Tank V-211 (N=10 vs N=5)



4.6.3. SNR Chart Performance In Terms of The Non-centrality Parameter

To further explore the performance of the SNR chart, OC curves were computed for SNR. The control limits for an SNR chart are computed from the non-central t distribution with given values of degrees of freedom and non-centrality parameter. We will denote this "null" value of the non-centrality parameter by δ_0. To explore the performance of the SNR chart as δ_0 varies, we created OC curves for level 0.001 charts with 9 degrees of freedom and non-centrality parameters δ based on different δ_0 values using the formula.

$$\delta = \delta_0 + k*std(\delta_0), \qquad (6)$$

where,

δ_0 = initial value of δ based on real data
k = constant varying over (-1.0,0.2) by 0.1 increments
$std(\delta_0)$ = standard deviation of δ_0 based on (2)

The curves in the chart in Figure 4.6-2 were based on computations for different values of δ. For a particular value of δ and a sample size of 10, a 0.001 critical value was determined based on the non-central t distribution. Then, equation (6) was used to determine which multiples of the subject value of δ would be plotted in the chart. For the new values of δ, the corresponding probability of a value falling below the 0.001 critical value was calculated and these points were plotted on the chart. This procedure was repeated for the different values of δ.

Based on a review of the chart below, the OC curves corresponding to higher values of δ_0 display clearly superior performance up to δ_0 values of 50 or so. Once the δ_0 values exceed 50 however, the effect appears to peak and larger values of δ_0 do not improve performance.

Figure 4.6.3-1: OC Curves for SNR Chart for Tank V-211 (N=10)

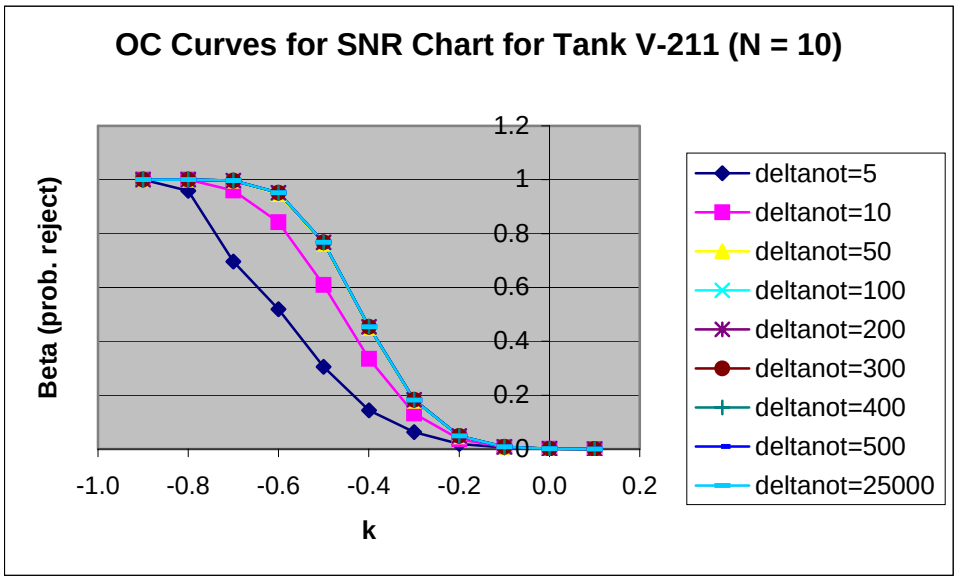
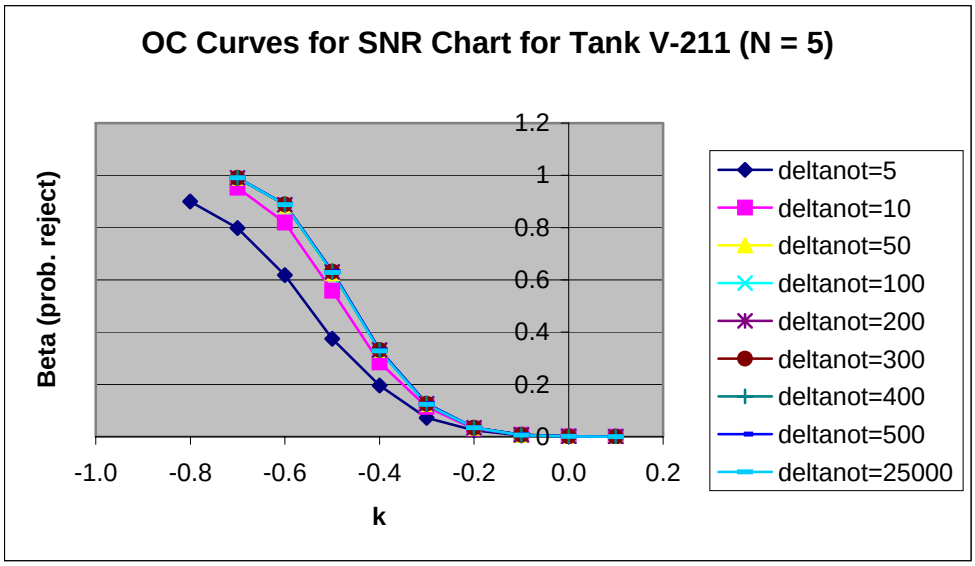


Figure 4.6.3-2: OC Curves for SNR Chart for Tank V-211 (N=5)



The same procedures as described above were used to create the chart in Figure 4.6-3, except that the sample size of 5 was used.

Although both charts indicate that higher values of δ display superior performance up to a certain point, the difference in performance between the different values of δ is smaller for the N = 5 case, as compared to the N = 10 case.

5. Conclusion

The objective of this project was to study the content uniformity assay values for the 0.25 mg/mL (0.5 mg) manufacturing process through the application of statistical techniques. The study focused on three separate tasks: (1) Monitoring process behavior for content uniformity assay values, (2) Ascertaining the equivalence of batches manufactured on Line 1 vs Line 2, and (3) Monitoring the signal to noise ratio of the content uniformity assay values

In order to accomplish the three tasks above, the following statistical techniques were applied: (1) Control chart techniques were applied to the data, including standard control chart techniques ($\bar{x}$ and S), individuals control chart techniques, and modified limits, (2) An equivalence test for the means of the two processes was conducted, and (3) A new control chart, the SNR chart, was developed and implemented

The traditional statistical process control techniques ($\bar{x}$ and S) that were applied in Section 2 resulted in the determination that the process was not in statistical process control. One reason for the process not being in statistical control is the potential for mixing to occur during the sampling process. The patterns that are shown on the $\bar{x}$ charts may be symptomatic of mixing. Since there are 10 independent filling needles on the Blow-Fill-Seal Machine and samples of 4 (begin), 3 (middle), and 3 (end) are randomly collected for each batch, the performance of a specific filling needle is not being consistently represented. This phenomenon could be a subject for future study. Unfortunately, this issue could not be addressed appropriately in this document since the author did not have ownership of the process.

The fact that between-batch variation was large relative to within-batch variation, and that the units were filled on multiple nozzles suggested that mixing was at least partly to blame for the poor patterns observed on the $\bar{x}$ charts. Since there was no record of which nozzles filled which of the tested units, a more appropriate method was sought for charting the process level. By eliminating within-batch variation as the measure of the size of between-batch variation, Individual Measurement Control Charts provide an alternative, and perhaps, fairer measure of process control. The charts were created by treating each batch as a single observation (e.g. the content uniformity values from begin, middle, and end portions of each batch were averaged to obtain one data point). The control procedure uses the moving range of two successive observations to estimate the process variability. This technique also resulted in the conclusion that the process was not in statistical process control, but was quite close to being so.

Despite the fact that traditional $\bar{x}$ and S control charts and Individual Measurement control charts indicated that the Line 1 and 2 processes in 2001 were not in statistical control, the natural variability of the processes was much smaller than the range of the specification limits. 0.237 – 0.262 mg/mL. In such situations, Modified Control Charts are appropriate for process monitoring if some drift in the process mean is acceptable. The modified $\bar{x}$ control chart is concerned only with detecting whether the true process

mean μ is located such that the process is producing a fraction non-conforming in excess of some specified value δ. This approach resulted in the conclusion that the process was in statistical process control with respect to the modified limits.

The equivalence test that was developed in Section 3 was a practical method to determine if the finished product results from the subject lines were equivalent. Prior to conducting this analysis, the process was verified to be in statistical process control with respect to the modified control limits. Based on the equivalence method outlined, the Line 1 and 2 content uniformity data from 2001 were statistically equivalent.

In order to chart and determine whether or not the SNR values were in statistical process control, a new chart (SNR) was developed. The SNR Chart was based on the assumption that the quantity ($\sqrt{n}\ \bar{x}$ / s) was non-central t distributed. Based on this assumption, a lower one-sided control limit was calculated. Review of the SNR charts for both compounding tanks revealed that all $\sqrt{n}\ \bar{x}$ / s values were above the one-sided lower control limit indicating that the processes were in statistical control.

The non-central t distribution characterized the SNR values well. Three different methods for estimating the non-centrality parameter δ were considered [using (1) $\bar{x}_{SNR}$ (2) s^2_{SNR} (3) setting $\delta = \frac{\bar{\bar{x}}\sqrt{n}}{\bar{s}}$]. In order to ascertain the best of these three estimation methods, a small simulation study was conducted and method 3 appeared to be most optimal.

In order to compare the performance of the SNR Chart to the Individuals Chart, both theoretical calculations and Monte Carlo simulations using the SAS macro *simulation* (Appendix 1) were utilized. The SNR control chart displayed superior performance to the Individuals control chart as it has a shorter response time to detect out of control signals.

We looked at operating characteristic curves for two sample sizes: N = 10 and N = 5. Although in both cases, the SNR Chart was superior to the Individuals Chart, the difference in performance between the SNR and Individuals Charts appeared to be smaller for the N = 5 case, as compared to the N = 10 case. As expected, the SNR Chart (N = 10) was superior to the SNR Chart (N = 5).

Based on a review of the OC Curves for different values of δ_0, higher values of δ_0 appear to display superior performance, but the improvement in performance was negligible for δ_0 values above 50.

As a result of this study, a great deal has been learned about the process. The process changes corresponding to these results have the potential to effect process improvements.

Appendix 1: SAS macro (*simulation*)

```
/* User specifies # of subsets, subset sample size */

%macro simulate(sigsize,noisize, sigsize1, noisize1);

/* Creates a column of values (1 to noisize, sigsize times) */

data one;
 %do i=1 %to &sigsize.;
  %do j=1 %to &noisize.;
   sample=&i.;
  output;
  %end;
 %end;
run;

/* Creates a matrix (sigsize x noisize = R by 1002) by sample.  Generates a 1000 columns
of R
values by sample which are normally distributed with mean and sigma specified by the
user.
Generates a column of random z values. */

data one;
 set one;
 array y{1000} y1-y1000;
 %do i=1 %to 1000;
  z=rannor(-1);
  y{&i.}=0*sample+0.248+.003*z;
 %end;
run;

/* Creates a matrix (sigsize by 1003) by sample.  Generates 1000 columns of sigsize
coefficent of variation values by sample. Generates a column of type and frequency. */

proc means data=one noprint;
 var y1-y1000;
 by sample;
 output out = sumstat cv=cv1-cv1000;
run;

/* Creates a row of type, frequency, 1000 values of N, 1000 values of mean, and 1000
values of std. */

proc means data=one noprint;
 var y1-y1000;
 output out = sumstat1 n=n1-n1000 mean=mean1-mean1000 std=std1-std1000;
run;

/* Creates a matrix (1000 x 5) of sample, N, Mean, STD, and Delta3 values.
Delta3 is calculated with the following formuala:
(((&noisize.)**(0.5))*mean) / std */

data signal3(keep=sample n mean std);
  set sumstat1;
  %do i=1 %to 1000;
    sample=&i.;
    n=n&i.;
    mean=mean&i.;
    std=std&i.;
    output;
  %end;
run;

data signal3;
  set signal3;
  delta3 = (((&noisize.)**(0.5))*mean) / std;
run;

/* Creates a row of type, frequency, mdelta3 = means of delta3 values. */

proc means data=signal3;
 var delta3;
 output out=del3 mean=mdelta3;
 run;

data _null_;
 set del3;
```

```
  call symput('mdelta3',trim(left(put(mdelta3,best10.))));
run;

proc print data=signal3;
run;

/* Creates a matrix (sigsize by 1001) by sample.  Generates 1000 columns of sigsize
signal to noise ratio values by sample. */

data signal(drop=_TYPE_ _FREQ_);
  set sumstat;
  %do i = 1 %to 1000;
    snr&i. = 100/ cv&i.;
    drop cv&i.;
  %end;
run;

/* Creates a row of type, frequency, 1000 values of N, 1000 values of mean, and 1000
values of var. */

proc means data=signal noprint;
  var snr1-snr1000;
  output out=sumstat2 n=n1-n1000 mean=mean1-mean1000 var=var1-var1000;
run;

/* Creates a matrix (1000 x 8) of sample, N, MEAN, VAR, GOFF, GOFF2, DELTA1,
and DELTA2 values.  GOFF, GOFF2, DELTA1, and DELTA2 are calculated with the
following formualas:
  goff = gamma((&noisize.-2)/2) / gamma((&noisize.-1)/2);
  goff2 = goff**2;
  delta1 = ((((&noisize.)**(0.5))*mean) / ((((&noisize.-1)/2)**(0.5))*goff);
  delta2 = ((((&noisize.)*var)-((&noisize.-1)/(&noisize.-3))) /
(((&noisize.-1)/(&noisize.-3))*(1-(((&noisize.-3)/2)*goff2))))**(0.5) */

data signal2(keep=sample n mean var);
  set sumstat2;
  %do i=1 %to 1000;
    sample=&i.;
    n=n&i.;
    mean=mean&i.;
    var=var&i.;
    output;
  %end;
run;

data signal2;
  set signal2;
  goff = gamma((&noisize.-2)/2) / gamma((&noisize.-1)/2);
  goff2 = goff**2;
  delta1 = ((((&noisize.)**(0.5))*mean) / ((((&noisize.-1)/2)**(0.5))*goff);
  delta2 = ((((&noisize.)*var)-((&noisize.-1)/(&noisize.-3))) /
(((&noisize.-1)/(&noisize.-3))*(1-(((&noisize.-3)/2)*goff2))))**(0.5);
run;

/* Creates a row of type, frequency, mdelta1 = means of delta1, and
mdelta2 = means of delta2 values. */

proc means data=signal2;
 var delta1 delta2;
 output out=mdelta3 mean=mdelta1 mdelta2;
run;

data _null_;
  set mdelta3;
   call symput('mdelta1',trim(left(put(mdelta1,best10.))));
   call symput('mdelta2',trim(left(put(mdelta2,best10.))));
run;

proc print data=signal2;
run;

/* Creates a matrix (sigsize by 1001) by sample.  Generates 1000 columns of sigsize
signal to noise ratio values by sample which have been multiplied by the values:
((&noisize.)**(0.5)). */

data signalc;
  set signal;
  %do i = 1 %to 1000;
    snrc&i. = ((&noisize.)**(0.5))*snr&i.;
    drop snr&i.;
  %end;
```

```
run;

/* Creates a row of type, frequency, and minimum values = to the minimum
values of the 1000 columns of SNRC values. */

proc summary data=signalc print min;
   var snrc1-snrc1000;
   output out=minsum min=min1-min1000;
run;

/* Creates a row of X1, 1000 minimum values, and a sum.  The X1 term is equal
to the 0.001 critical value of the non-central t distribution based on a
non-centrality parameter = average of the delta1 values (mdelta1).  The critical
value is compared to the minimum snrc value (SNR x n^0.5) and a 1 is returned if
the minimum value is greater than the critical value.  The sum term is equal to
the sums of the ones and zeros to be able to determine the # of minimums which
were below the critical value.  This procedure is repeated for delta2 and delta3
values. */

data compare1(drop=_TYPE_ _FREQ_ min1-min1000);
  set minsum;
   x1=tinv(.001,(&noisize.-1), &mdelta1);
   %do i = 1 %to 1000;
    c&i. = (min&i.>x1);
   %end;
run;

data compare2(drop=_TYPE_ _FREQ_ min1-min1000);
  set minsum;
   x2=tinv(.001,(&noisize.-1), &mdelta2);
   %do i = 1 %to 1000;
    c&i. = (min&i.>x2);
   %end;
run;

data compare3(drop=_TYPE_ _FREQ_ min1-min1000);
  set minsum;
   x3=tinv(.001,(&noisize.-1), &mdelta3);
   %do i = 1 %to 1000;
     c&i. = (min&i.>x3);
   %end;
run;

data compare1;
 set compare1;
 sum=sum(of c1-c1000);
run;

data compare2;
 set compare2;
 sum=sum(of c1-c1000);
run;

data compare3;
 set compare3;
 sum=sum(of c1-c1000);
run;

/* NEW PART 1 */

/* Creates a column of values (1 to noisize1, sigsize1 times) */

data two;
 %do i=1 %to &sigsize1.;
  %do j=1 %to &noisize1.;
   sample=&i.;
  output;
  %end;
 %end;
run;

/* Creates a matrix (sigsize1 x noisize1 = R by 1002) by sample.  Generates a 1000
columns of R
values by sample which are normally distributed with mean and sigma specified by the
user.
Generates a column of random z values. */

data two;
 set two;
 array y{1000} y1-y1000;
 %do i=1 %to 1000;
```

```
  z=rannor(-1);
  y{&i.}=0*sample+0.248+.003*z;
 %end;
run;

/* Creates a matrix (sigsize1 by 1003) by sample.  Generates 1000 columns of sigsize1
coefficent of variation values by sample. Generates a column of type and frequency. */

proc means data=two noprint;
 var y1-y1000;
 by sample;
 output out = sumstat3 cv=cv1-cv1000;
run;

data signal(keep=sample cval1-cval1000 low1-low1000);
  set sumstat3;
    %do i = 1 %to 1000;
      snr&i. = 100 / cv&i.;
      cval&i. = snr&i. * (&noisize.**.5);
      low&i. = 0;
      if cval&i. < 200 then low&i. = 1;
    %end;
run;

data signal(keep=sample rlength cval);
  set signal;
  %do i = 1 %to 1000;
      if low&i. = 1 then do;
      rlength = &i.;
      cval = cval&i.;
      i = 1000;
      output;
      end;
  %end;
run;

proc print;
run;

proc sort data=signal;
  by sample rlength;
run;

proc print;
run;

data signal2;
  set signal;
  by sample rlength;
  if first.sample;
run;


proc print;
run;

proc means data = signal2;
  var rlength;
  output out = sumstat4 mean=mean n=n std=std uclm=uclm lclm=lclm;
run;

%mend simulate;
%simulate(25,10,50,10);


%macro simulate(sigsize,noisize,inputmn,inputsd);

data three;
 %do i=1 %to &sigsize.;
  %do j=1 %to &noisize.;
   sample=&i.;
  output;
  %end;
 %end;
run;

data three;
 set three;
 array y{1000} y1-y1000;
 %do i=1 %to 1000;
  z=rannor(-1);
```

```
  y{&i.}=0*sample+ &inputmn. + &inputsd.*z;
 %end;
run;

/* Creates a matrix (sigsize by 1003) by sample.  Generates 1000 columns of
sigsize
coefficent of variation values by sample. Generates a column of type and
frequency. */

proc means data=three noprint;
 var y1-y1000;
 by sample;
 output out = sumstat5 cv=cv1-cv1000;
run;

data signal5(keep=sample cval1-cval1000 low1-low1000);
  set sumstat5;
    %do i = 1 %to 1000;
      snr&i. = 100 / cv&i.;
      cval&i. = snr&i. * (&noisize.**.5);
      low&i. = 0;
      if cval&i. < 200 then low&i. = 1;
    %end;
run;

data signal5(keep=sample rlength cval);
  set signal5;
  %do i = 1 %to 1000;
      if low&i. = 1 then do;
      rlength = &i.;
      cval = cval&i.;
      i = 1000;
      output;
      end;
  %end;
run;

proc sort data=signal5;
  by sample rlength;
run;

data signal6;
  set signal5;
  by sample rlength;
  if first.sample;
run;

proc means data = signal6;
  var rlength;
  output out = sumstat6 mean=mean n=n std=std uclm=uclm lclm=lclm;
run;

proc print;
run;


%mend simulate;
%simulate(25,10,0.243,0.003);
```

Appendix 4a: Results of the SAS Macro Simulation

OBS	SAMPLE	RLENGTH	CVAL
1	1	577	315.239
2	2	84	297.696
3	3	310	303.142
4	4	128	314.280
5	5	311	311.428
6	6	98	298.315
7	7	607	314.429
8	8	32	303.638
9	9	184	314.358
10	10	313	295.903
11	11	108	263.752
12	12	59	258.182
13	13	567	293.137
14	14	36	294.070
15	15	554	309.796
16	16	244	302.268
17	17	41	302.637
18	18	158	302.823
19	19	324	312.360
20	20	114	311.173
21	21	536	307.524
22	22	36	294.123
23	23	331	298.152
24	24	418	302.742
25	25	388	310.545

Appendix 4b: Results of the SAS Macro Simulation

OBS	SAMPLE	RLENGTH	CVAL
1	1	538	312.320
2	2	226	299.408
3	3	77	314.086
4	4	657	313.702
5	5	118	291.494
6	6	61	308.911
7	7	9	305.978
8	8	36	293.112
9	9	321	308.216
10	10	501	288.800
11	11	49	310.281
12	12	221	308.512
13	13	36	287.262
14	14	176	311.470
15	15	82	298.419
16	16	21	280.864
17	17	140	310.850
18	18	189	285.706
19	19	54	297.879
20	20	13	294.721
21	21	191	298.382
22	22	188	300.971
23	23	80	306.944
24	24	19	304.597
25	25	122	281.537

Appendix 4c: Results of the SAS Macro Simulation

OBS	SAMPLE	RLENGTH	CVAL
1	1	94	306.764
2	2	52	305.806
3	3	28	301.862
4	4	110	312.290
5	5	68	305.994
6	6	70	290.817
7	7	5	286.204
8	8	94	297.611
9	9	90	302.768
10	10	29	312.318
11	11	265	306.495
12	12	111	313.896
13	13	284	313.339
14	14	53	310.468
15	15	46	295.730
16	16	130	294.269
17	17	17	311.678
18	18	157	240.423
19	19	23	312.382
20	20	12	311.799
21	21	20	297.985
22	22	5	306.719
23	23	46	284.855
24	24	21	295.019
25	25	330	300.879

Appendix 4d: Results of the SAS Macro Simulation (.15, .00154)

OBS	SAMPLE	RLENGTH	CVAL
1	1	10	180.880
2	2	22	225.464
3	3	23	213.579
4	4	7	206.961
5	5	21	206.137
6	6	10	190.072
7	7	16	205.700
8	8	14	224.027
9	9	49	215.243
10	10	2	218.775
11	11	68	204.320
12	12	12	205.048
13	13	1	225.524
14	14	8	162.694
15	15	7	170.905
16	16	5	219.862
17	17	18	222.593
18	18	1	218.098
19	19	8	223.262
20	20	64	221.040
21	21	3	208.550
22	22	18	217.482
23	23	16	225.510
24	24	7	207.594
25	25	6	203.791

Appendix 4e: Results of the SAS Macro Simulation (.2465, .00252) - Individuals

OBS	SAMPLE	RLENGTH	CVAL
1	1	6	224.429
2	2	16	211.386
3	3	38	211.972
4	4	10	225.774
5	5	14	224.423
6	6	6	214.853
7	7	1	200.429
8	8	4	208.209
9	9	25	208.078
10	10	7	220.246
11	11	8	165.243
12	12	53	205.592
13	13	43	219.336
14	14	2	211.440
15	15	28	214.511
16	16	27	206.586
17	17	1	203.968
18	18	3	215.003
19	19	48	212.680
20	20	49	213.759
21	21	14	204.318
22	22	50	216.616
23	23	11	219.664
24	24	8	208.055
25	25	19	203.637

Appendix 4f: Results of the SAS Macro Simulation (.2465, .00252) - SNR

OBS	SAMPLE	RLENGTH	CVAL
1	1	1	302.729
2	2	1	314.335
3	3	2	225.513
4	4	2	310.738
5	5	1	280.214
6	6	1	221.630
7	7	3	257.498
8	8	1	310.422
9	9	2	257.279
10	10	1	313.199
11	11	1	278.952
12	12	2	303.364
13	13	3	299.858
14	14	2	238.105
15	15	1	292.074
16	16	1	252.944
17	17	2	292.694
18	18	1	307.353
19	19	1	252.173
20	20	1	288.266
21	21	1	309.825
22	22	5	288.109
23	23	4	314.306
24	24	1	295.453
25	25	1	229.127

BIBLIOGRAPHY

Basic Ideas of Scientific Sampling, Alan Stuart, Hafner Publishing Company, New York, 1962

Density Estimation for Statistics and Data Analysis, B.W. Silverman, Chapman and Hall, New York, 1986

Distribution-free Statistical Methods, 2nd Edition, J.S. Maritz, Chapman and Hall, New York, 1995

Fine Particle Measurement, Clyde Orr, The Macmillan Company, New York, 1959

Handbook of Powder Science and Technology, M. E. Fayad and L. Otten, Van Nostrand Reinhold Company Inc., 1984

Introduction to Statistical Quality Control, 3rd Ed., Douglas C. Montgomery, John Wiley and Sons Inc., New York, 1996

Lognormal Distributions Theory and Applications, edited by Edwin L. Crow and Kunio Shimizu, Marcel Dekker, Inc., New York, 1988

Nonparametric Probability Density Estimation, Richard A. Tapia and James R. Thompson, The Johns Hopkins University Press, Baltimore, 1978

Optimization &Variation Reduction in Quality, Wayne A. Taylor, McGraw-Hill, Inc., 1991

Particle Size Analysis in Industrial Hygiene, Leslie Silverman, Charles Billings, and Melvin First, Academic Press, New York, 1971

Particle Size Analysis in Pharmaceutics and Other Industries, Clive Washington, Ellis Horwood Limited, New York, 1992

Particle Size Determination, R.D. Cadle, Interscience Publishers, New York, 1955

Particle Size: Measurement, Interpretation, and Application, Riyad Irami and Clayton Callis, John Wiley and Sons Inc., New York, 1963

Particulate Science and Technology, John Keith Beddow, Chemical Publishing Co., Inc., New York, 1980

Pharmaceutical Statistics: Practical and Clinical Applications, Sanford Bolton, Dekker, New York, 1990

Sampling, Steven K. Thompson, John Wiley and Sons Inc., New York, 1992

Statistical Power Analysis, Kevin R. Murphy and Brett Myors, Lawrence Erlbaum Associates, Inc., Publishers, New Jersey, 1998

Statistics in the Pharmaceutical Industry, C. Ralph Buncher and Jia-Yeong Tsay, M. Dekker, New York, 1994

How to Choose the Proper Sample Size (Volume 12), Gary G. Brush, American Society for Quality Control, Wisconsin, 1988