PHARMACEUTICAL ONLINE

Guest Column | March 3, 2017

How To Establish The Number Of Runs Required For Process Validation

By Mark Durivage, ASQ Fellow

The first article in this series, *Risk-Based Approaches To Establishing Sample Sizes For Process Validation* (June 2016), provided and established the relationship between risk and sample size. Subsequent articles have introduced different methods for determining sample sizes to ensure validation activities will yield valid results. This article will demonstrate the how to establish the number of runs required for process validation.

U.S. Food and Drug Administration (FDA) regulations, International Organization for Standardization (ISO) standards, and Global Harmonization Task Force (GHTF) guidance documents do not prescribe the number of runs required for process validation activities. Industry has typically used three batches during the process performance qualification (PPQ) phase to demonstrate that a process is capable of consistently delivering quality product.

However, the "rule of three" batches or runs is no longer appropriate for process validation activities. FDA's guidance for industry *Process Validation: General Principles and Practices* (2011) recommends that a PPQ protocol should include the sampling plan, "including sampling points, number of samples, and the frequency of sampling for each unit operation and attribute. The number of samples should be adequate to provide sufficient statistical confidence of quality both within a batch and between batches. The confidence level selected can be based on risk analysis as it relates to the particular attribute under examination. Sampling during this stage should be more extensive than is typical during routine production."

Even those with a limited knowledge of statistics will recognize that a run of three is not statistically significant. And those who have a mastery of statistical techniques already know that a sample of three is not significant. If a manufacturer were to use the success-run theorem to determine the number of process validation runs, using 95% confidence and 90% reliability, it would require 30 runs. However, 30 runs is neither practical nor cost effective. Additionally, some products may be produced only occasionally.

Using a DOE Scheme to Justify PPQ Runs

One method that can be used to justify the number of PPQ runs is based upon the design of experiments (DOE) used to identify and characterize the process parameters. Generally, a screening DOE will be used to separate the "vital few from the trivial many" (Pareto analysis). For example, a machine has seven parameters that can be set, and each setting has two levels. A full factorial DOE would require 2^7 = 128 runs to fully determine the main effects and interactions. The same 2^7 run as a fractional factorial would require eight runs. However, with eight runs, interactions will not be identified.

Using the example, a machine has seven parameters that can be set, and each setting has two levels (2^7 128 runs). The engineering team decided to use a fractional factorial experiment that requires eight runs. The result of the experiment indicated that factors A, B, and C were significant. (Factors D, E, F, and G were insignificant.) To further explore the effects of factors A, B, and C, the team will run a full factorial 2^3 experiment, which requires eight runs. The results of the 2^3 experiment are shown in Figure 1.

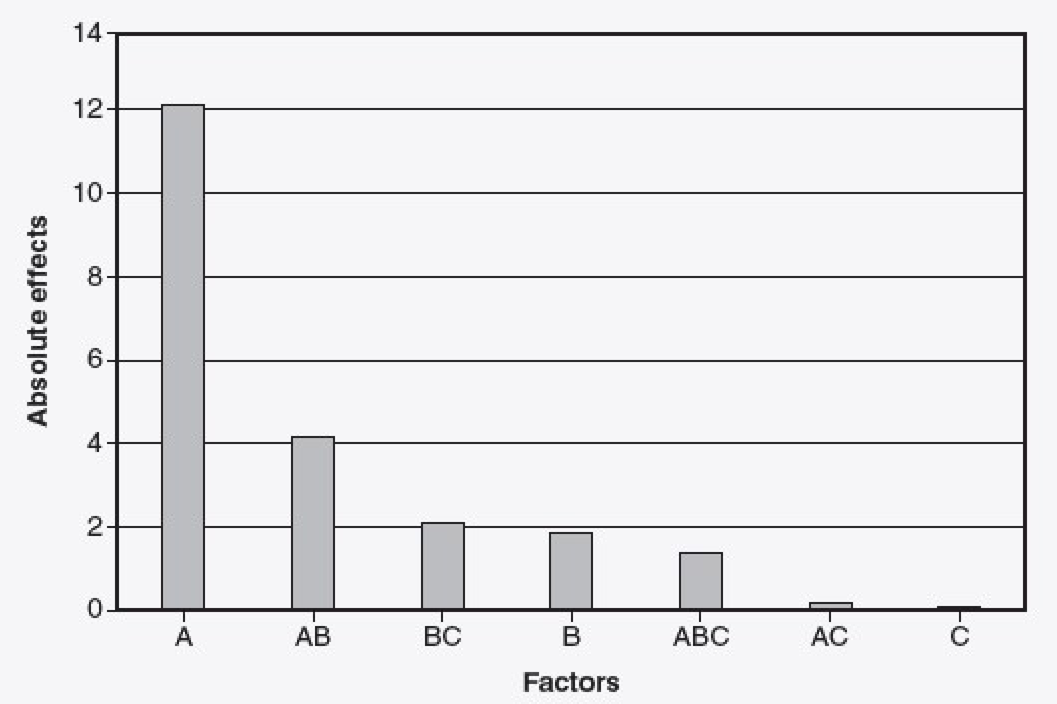


Figure 1*: Pareto chart of the absolute values of the effects*

Factor is significant. The interaction AB is also significant. Due to the hierarchy of model rule, factor B is also included. We have now gone from a 2^7 fractional factorial with eight runs to a 2^2 full factorial experiment requiring four runs. Due to the hierarchy of model rule, factors A and B must be used. Four runs could now be justified for the number of PPQ runs.

Note: Factors C, D, E, F, and G are not significant and can be set solely based on cost, productivity, or convenience.

Shifts and Suppliers Scheme

The purpose of PPQ runs is to establish that a process is capable of consistently delivering quality product by introducing as much variation into the process as is expected during production operations. The two main sources of variation (disregarding the process itself) will relate to the number of shifts producing the product and the number of suppliers suppling a particular raw material or component used in the production process (i.e., two suppliers suppling a key chemical or component).

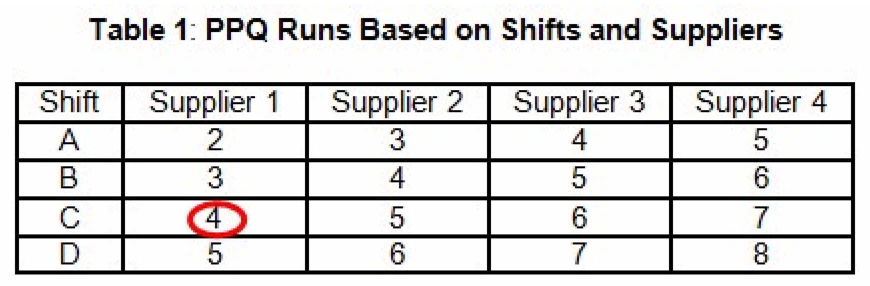


Table 1: PPQ Runs Based on Shifts and Suppliers

Shift	Supplier 1	Supplier 2	Supplier 3	Supplier 4
A	2	3	4	5
B	3	4	5	6
C	4	5	6	7
D	5	6	7	8

Table 1 provides an example that uses a shift and supplier scheme to determine the number of required PPQ runs. For example, a manufacturing facility maintains three shifts and uses a sole supplier for each chemical used in a blending process. According to Table 1, four PPQ runs will be required. It should be obvious that this method is risk-based and not statistically based or statistically significant.

FMEA Scheme

Before we begin, we must establish our definitions of risk and the minimum required number of PPW runs. These definitions can and should vary based upon organizational needs. A good place to determine the risk level is from the failure modes and effects analysis (FMEA). FMEA (design, process, user) is a systematic group of activities designed to recognize, document, and evaluate the potential failure of a product or process, and its effects. FMEA uses a risk priority number (RPN), which is comprised of frequency, detection, and severity. The higher the RPN, the higher the risk. However, a high severity in conjunction with a low probability of occurrence and high probability of detection may still necessitate the appropriate controls for high risk. Figure 2 depicts an example FMEA with the associated risk levels. Once the risk level has been determined (low, medium, high), the appropriate number of PPQ runs can be selected.

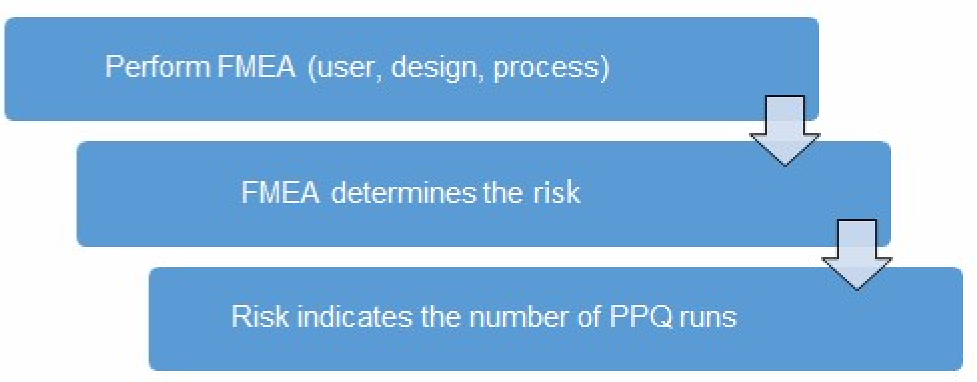
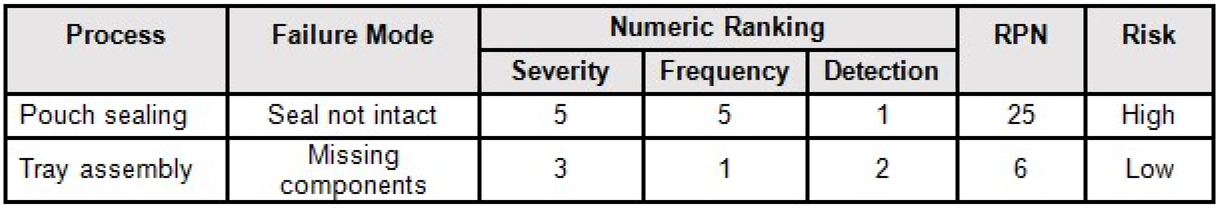

Perform FMEA (user, design, process)
FMEA determines the risk
Risk indicates the number of PPQ runs

***Figure 2**: Risk process for determining the appropriate number of PPQ runs*

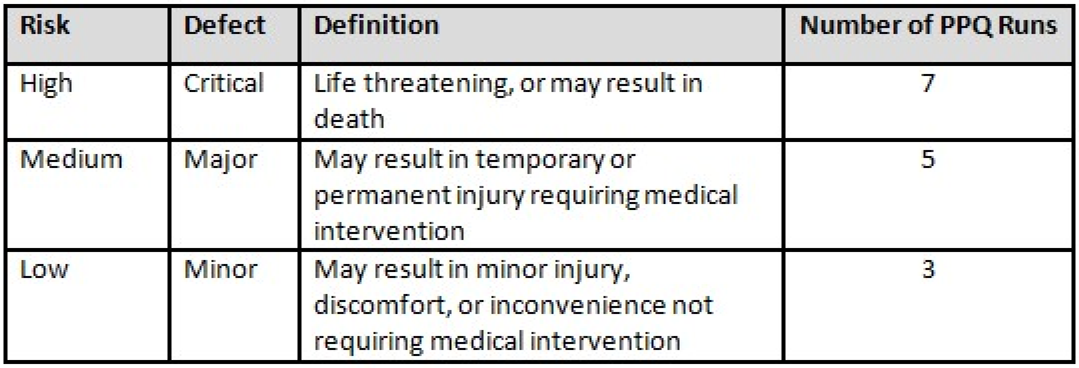
Table 2: Example FMEA



Process	Failure Mode	Numeric Ranking			RPN	Risk
		Severity	Frequency	Detection		
Pouch sealing	Seal not intact	5	5	1	25	High
Tray assembly	Missing components	3	1	2	6	Low

Table 3 shows an example of risk level definitions with accompanying defect classifications. These definitions can and will vary based upon the product(s) produced and their intended and unintended uses.

Table 3: Example of Risk Level Definitions and Number of PPQ Runs Required



Risk	Defect	Definition	Number of PPQ Runs
High	Critical	Life threatening, or may result in death	7
Medium	Major	May result in temporary or permanent injury requiring medical intervention	5
Low	Minor	May result in minor injury, discomfort, or inconvenience not requiring medical intervention	3

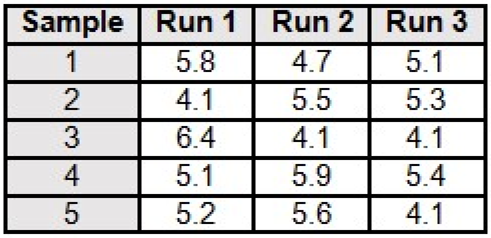
Demonstrating Statistical Confidence Between Runs

Demonstrating sufficient statistical confidence between runs can be accomplished using the non-parametric Kruskal-Wallis test and Leven's test. These tests will help determine if there is sufficient evidence to conclude whether two or more means or variances respectively are equal with a significance level α (usually 90%, 95%, or 99%). Both tests allow the use of unequal sample sizes. Please be aware that the original data **must pass** the stated validation criteria to perform these tests.

A manufacturing process is being validated. The validation team determines that three PPQ validation runs would be necessary based upon the risk assessment. The data for the three PPQ validation runs is shown in Table 4. The non-parametric Kruskal-Wallis Means Test and Leven's Variance Test will be used to assess if there is sufficient evidence to conclude whether the means and variances are equal with a significance level α 95%.

Note: Each run only has five data points, to simplify the example calculations.

Table 4: PPQ Validation Data



Sample	Run 1	Run 2	Run 3
1	5.8	4.7	5.1
2	4.1	5.5	5.3
3	6.4	4.1	4.1
4	5.1	5.9	5.4
5	5.2	5.6	4.1

Kruskal-Wallis Test

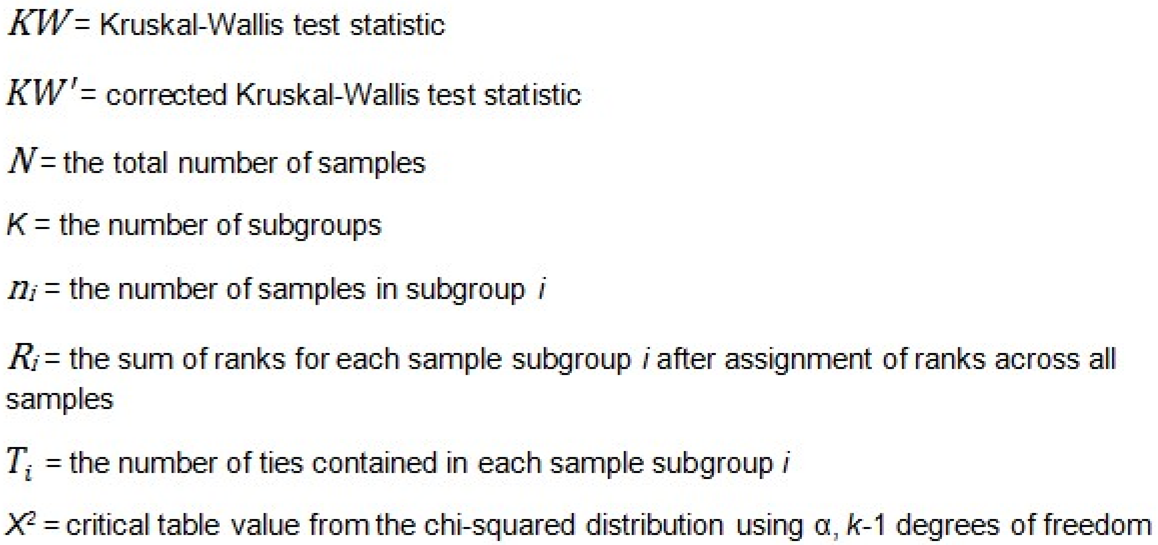
The Kruskal-Wallis test is a non-parametric test used to determine if two or more samples originate from the same distribution by evaluating the means. The Kruskal-Wallis test assigns ranks to the data points, replacing the original data points. The test statistic KW is compared a critical □X^2 value to determine if:

H_o: $\mu_1 = \mu_2 = \ldots = \mu_k$ or H_A: $\mu_i \neq \mu_k$ for at least one pair.

The Kruskal-Wallis test statistic (KW) is calculated using:

$$KW = \left[\frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i}\right] - 3(N+1)$$

Where:



KW = Kruskal-Wallis test statistic

KW' = corrected Kruskal-Wallis test statistic

N = the total number of samples

K = the number of subgroups

n_i = the number of samples in subgroup i

R_i = the sum of ranks for each sample subgroup i after assignment of ranks across all samples

T_i = the number of ties contained in each sample subgroup i

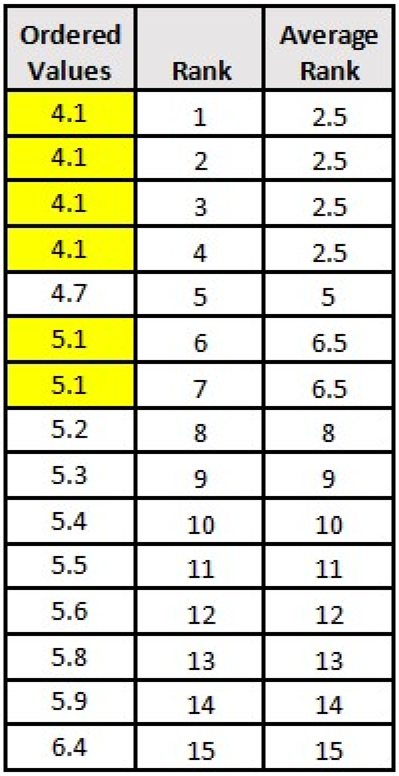
X^2 = critical table value from the chi-squared distribution using α, k-1 degrees of freedom

When there are many ties (≈50% or greater), a correction factor for *KW* is made by:

$$KW' = \frac{KW}{1 - \left[\sum_{i=1}^{k} \frac{(R_i^3 - R_i)}{N^3 - N}\right]}$$

To calculate the KW test statistic, we must order the data points from Table 4 and assign ranks to each value. When values are repeated, the average rank is assigned (see the highlighted values in Table 5).

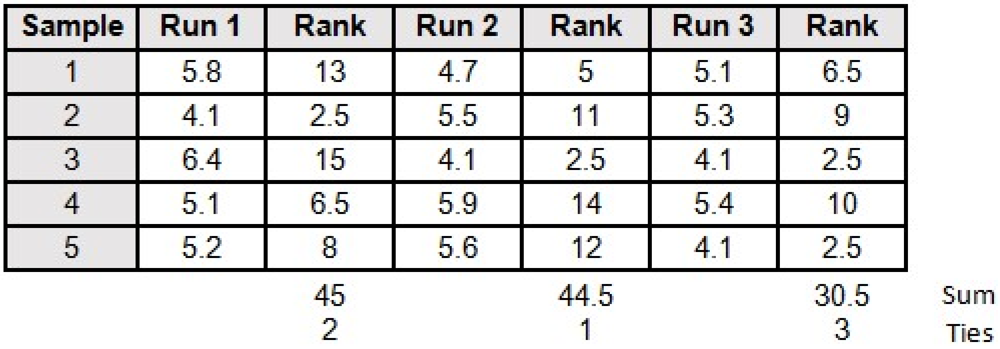
Table 5: Ordered and Ranked Data



Ordered Values	Rank	Average Rank
4.1	1	2.5
4.1	2	2.5
4.1	3	2.5
4.1	4	2.5
4.7	5	5
5.1	6	6.5
5.1	7	6.5
5.2	8	8
5.3	9	9
5.4	10	10
5.5	11	11
5.6	12	12
5.8	13	13
5.9	14	14
6.4	15	15

The average ranks are then assigned to the data in the original runs (see Table 6).

Table 6: Ranks Assigned to Data in Original Runs



Sample	Run 1	Rank	Run 2	Rank	Run 3	Rank	
1	5.8	13	4.7	5	5.1	6.5	
2	4.1	2.5	5.5	11	5.3	9	
3	6.4	15	4.1	2.5	4.1	2.5	
4	5.1	6.5	5.9	14	5.4	10	
5	5.2	8	5.6	12	4.1	2.5	
		45		44.5		30.5	Sum
		2		1		3	Ties

$$KW = \left[\frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i}\right] - 3(N+1)$$

$$KW = \left[\frac{12}{15(15+1)}\frac{45^2}{5} + \frac{44.5^2}{5} + \frac{30.5^2}{5}\right] - 3(15+1) = 1.355$$

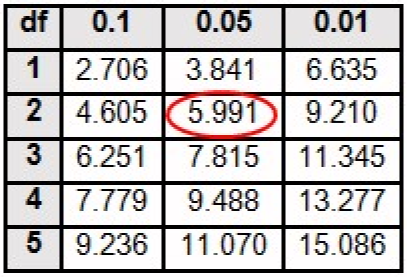
For the purpose of this example, we will calculate *KW*'

$$KW' = \frac{KW}{1 - \left[\sum_{i=1}^{k} \frac{(T_i^3 - T_i)}{N^3 - N}\right]}$$

$$KW' = \frac{1.355}{1 - \left[\frac{(2^3-2)}{15^3-15} + \frac{(1^3-1)}{15^3-15} + \frac{(3^3-3)}{15^3-15}\right]} = 1.367$$

The critical chi-squared distribution value is found using α and *k*-1degrees of freedom. For this example, α of 0.05 will be used.

Table 7: Chi-Square Distribution Table (partial)



df	0.1	0.05	0.01
1	2.706	3.841	6.635
2	4.605	5.991	9.210
3	6.251	7.815	11.345
4	7.779	9.488	13.277
5	9.236	11.070	15.086

$KW'_{\text{calculated}} = 1.367$

$KW'_{\text{critical}} = 5.991$

Since *KW'* calculated is less than *KW'* critical at the 0.05 α (alpha) level, there is insufficient evidence to reject the null hypothesis that H0: μ1 = μ2 = μ3. Therefore, we have demonstrated sufficient statistical confidence for the validation runs (for the means). Please note that if *KW* (1.355) was used, the conclusion would be the same.

Levene's Test

Levene's test is a non-parametric test used to determine if two or more samples have the same variance. The Levene's test uses deviations replacing the original data points. The test statistic *L* is compared a critical □*F* value to determine if:

H_o: $\sigma_1 = \sigma_2 = \ldots = \sigma_k$ or H_A: $\sigma_i \neq \sigma_k$ for at least one pair

$$L = \frac{\sum_{i=1}^{k} n_i(\overline{D}_i - \overline{D})^2}{k-1} \Big/ \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\overline{D}_{ij} - \overline{D}_i\right)^2}{N-k}$$

Where:

L = Levene's test statistic

N = the total number of samples

K = the number of subgroups

n_i = the number of samples in subgroup i

$D_{ij} = \left|y_{ij} - \tilde{y}_i\right|$ = the absolute deviation of observation j from treatment i median

$\overline{D}_i$ = average of the n_i absolute deviations from treatment i

$\overline{D}$ = average of all N absolute deviations

$F = \alpha, (df_1 = k - 1), (df_2 = n – k)$

Table 8: **Median ($\tilde{y}$)**

Sample	Run 1	Run 2	Run 3	
1	5.8	4.7	5.1	
2	4.1	5.5	5.3	
3	6.4	4.1	4.1	
4	5.1	5.9	5.4	
5	5.2	5.6	4.1	
	5.2	5.5	5.1	median ($\tilde{y}$)

Table 9: $(\bar{\boldsymbol{D}}_{\boldsymbol{i}})$ **and** $(\bar{\boldsymbol{D}})$

Sample	Run 1	$\lvert y-5.2\rvert$	Run 2	$\lvert y-5.5\rvert$	Run 3	$\lvert y-5.1\rvert$
1	5.8	0.6	4.7	0.5	5.1	0.1
2	4.1	1.1	5.5	0.3	5.3	0.1
3	6.4	1.2	4.1	1.1	4.1	1.1
4	5.1	0.1	5.9	0.7	5.4	0.2
5	5.2	0	5.6	0.4	4.1	1.1
		0.60 $(\bar{D}_i)$ avg		0.54 $(\bar{D}_i)$ avg		0.50 $(\bar{D}_i)$ avg

0.55 $(\bar{D})$ avg

Table 10: $\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\bar{\boldsymbol{D}}_{ij}-\bar{\boldsymbol{D}}_{i}\right)^2$

Sample	Run 1	$\left(\bar{D}_{ij}-\bar{D}_i\right)^2$	Run 2	$\left(\bar{D}_{ij}-\bar{D}_i\right)^2$	Run 3	$\left(\bar{D}_{ij}-\bar{D}_i\right)^2$
1	0.6	0.00	0.8	0.07	0	0.36
2	1.1	0.25	0	0.29	0.2	0.16
3	1.2	0.36	1.4	0.74	1	0.16
4	0.1	0.25	0.4	0.02	0.3	0.09
5	0	0.36	0.1	0.19	1	0.16
		1.22 sum		1.31 sum		0.88 sum

3.41 sum

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\bar{D}_{ij}-\bar{D}_i\right)^2 = 3.41$$

Table 11: $\sum_{i=1}^{k}\boldsymbol{n}_{\boldsymbol{i}}(\bar{\boldsymbol{D}}_{\boldsymbol{i}}-\bar{\boldsymbol{D}})^2$**z**

	$(\bar{D}_i)$	$n_i(\bar{D}_i-\bar{D})^2$
Run 1	0.06	0.014
Run 2	0.54	0.000
Run 3	0.5	0.011
		0.025

sum

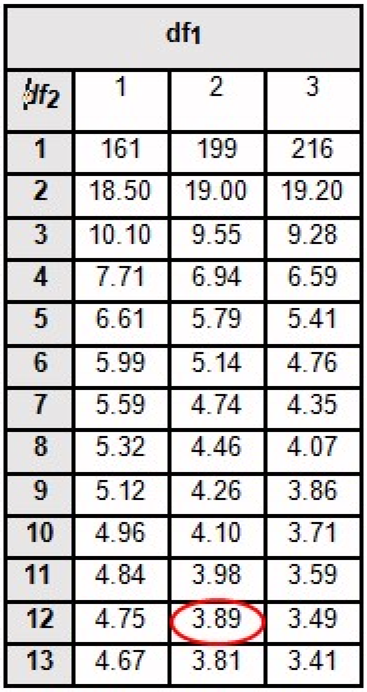
$$\sum_{i=1}^{k} n_i(\overline{D}_i - \overline{D})^2 = 0.04$$

$$L = \frac{\sum_{i=1}^{k} n_i(\overline{D}_i - \overline{D})^2}{k-1} / \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\overline{D}_{ij} - \overline{D}_i\right)^2}{N-k}$$

$$L = \frac{0.025}{3-1} / \frac{3.41}{15-3} = 0.04$$

The critical *F* distribution value is found using α,(df1 = k - 1), (df2 = n – k). For this example, α of 0.05 will be used.

Table 12: *F* Distribution α 0.05 Table (partial)



	df1		
df2	1	2	3
1	161	199	216
2	18.50	19.00	19.20
3	10.10	9.55	9.28
4	7.71	6.94	6.59
5	6.61	5.79	5.41
6	5.99	5.14	4.76
7	5.59	4.74	4.35
8	5.32	4.46	4.07
9	5.12	4.26	3.86
10	4.96	4.10	3.71
11	4.84	3.98	3.59
12	4.75	3.89	3.49
13	4.67	3.81	3.41

$L_{calculated} = 0.04$

$F_{critical} = 3.89$

Since *L* calculated is less than *F* critical at the 0.05 α (alpha) level, there is insufficient evidence to reject the null hypothesis that H0: σ1 = σ2 = σ3. Therefore, we have demonstrated sufficient statistical confidence for the validation runs (for the variances).

The examples provided used three runs with five data points each. If a PPQ required five runs each with 60 data points, the calculations will quickly become very cumbersome. Hopefully it is apparent that the use of spreadsheets and statistical software will facilitate the calculations for the Kruskal-Wallis and Leven's tests.

Choosing the Right Requirements for your Process

I want to reinforce that selecting the number of PPQ runs should be based upon an organization's risk acceptance determination threshold, industry practice, guidance documents, and regulatory requirements.

Additional considerations should be reviewed and justified for "family of parts." Choosing the worst case, and providing why the selected items are worst case, is extremely important.

Validation activities should be supported by laboratory testing (physical and analytical), continuous process monitoring; first article layouts (FAL); first article inspections (FAI); design of experiments (DOE); test method validation (TMV); measurement systems analysis (MSA); incoming, in-process, and final (release) inspections; calibration; and training. A sound sampling scheme will help demonstrate sufficient statistical confidence of quality both within a run and between runs. Validation activities should not be performed in isolation, but rather as part of a process supported by the activities just listed.

This article series has introduced several different methods for establishing sample sizes for process validation. The articles in the series include:

- Risk-Based Approaches To Establishing Sample Sizes For Process Validation
- How To Establish Sample Sizes For Process Validation Using The Success-Run Theorem
- How To Use Reliability-Based Life Testing Sampling For Process Validation
- How To Establish Sample Sizes For Process Validation Using C=0 Sampling Plans
- How To Establish Sample Sizes For Process Validation Using Statistical Tolerance Intervals
- How To Establish Sample Sizes For Process Validation Using Variable Sampling Plans

- How To Establish Sample Sizes For Process Validation Using LTPD Sampling
- How To Establish Sample Sizes For Process Validation When Destructive or Expensive Testing Is Required

References:

1. Durivage, M.A., 2016, *Practical Design of Experiments (DOE)*, Milwaukee, ASQ Quality Press
2. Durivage, M.A., 2014, *Practical Engineering, Process, and Reliability Statistics*, Milwaukee, ASQ Quality Press
3. Durivage, M.A. and Mehta B., 2016, *Practical Process Validation*, Milwaukee, ASQ Quality Press
4. Durivage, M.A., 2016, *Risk-Based Approaches To Establishing Sample Sizes For Process Validation,* Life Science Connect
5. *FDA Guidance for Industry: Process Validation: General Principles and Practices.* FDA online. Accessed December 27, 2016. http://www.fda.gov/downloads/Drugs/Guidances/ UCM070336.pdf.
6. The Global Harmonization Task Force (GHTF). 2004. SG3 *Quality Management Systems— Process Validation Guidance.* 2nd ed. GHTF.

About the Author

Mark Allen Durivage is the managing principal consultant at Quality Systems Compliance LLC and an author of several quality-related books. He earned a B.A.S. in computer aided machining from Siena Heights University and a M.S. in quality management from Eastern Michigan University. Durivage is an ASQ Fellow and holds several ASQ certifications including CQM/OE, CRE, CQE, CQA, CHA, CBA, CPGP, and CSSBB. He also is a Certified Tissue Bank Specialist (CTBS) and holds a Global Regulatory Affairs Certification (RAC). Durivage resides in Lambertville, Michigan. Please feel free to email him at mark.durivage@qscompliance.com with any questions or comments, or connect with him on LinkedIn.