FOI Services Teleconference TC161130

Determining a Statistically Valid Sample Size: What Does FDA Expect to See?

Presented by:

Steven Walfish

When:

November 30, 2016

Eastern Standard Time: 1:00pm – 2:30pm (GMT-5)

Central Time: 12:00pm – 1:30 pm

Mountain Time: 11:00am – 12:30pm

Pacific Time: 10:00am – 11:30am

Prepare:

Please provide each listener with access to this handout; it is online at www.foiservices.com/safesam4166.htm

Black & white photocopies can be used.

Connect:

Call in 5-10 minutes before the start time:

1-866-614-1121 (Toll-from free US & Canada)

Conference ID: 17414166

Your unique PIN is on your invoice & your dial-in email.

Registrants from outside the US/Canada should contact FOI for customized instructions.

Please Note:

Each registration you purchase covers one connection. You are welcome to put the conference on a speakerphone and invite as many listeners as you like for your registration. If you need to dial-in from an additional line, another registration with a different PIN is required; call +1-301-975-9400 for more information.

Do not use a conferencing feature to add remote locations to the one registered location. The quality of the connection will be greatly reduced and several features, such as the ability to ask a question, are disabled.

This is an audio-format presentation. For the best connection, use a corded telephone on a land-line. You will be connected for 90 minutes or more; relying on a battery-powered phone may not be wise.

If you ask a question during the live Q&A session, please try to avoid using a speakerphone.

This presentation is recorded. If you register for the live conference, you can order the audio package of an mp3 file and all of the handouts at a 50% discount; contact FOI for details.

Produced by:

FOI Services, Inc.
23219 Stringtown Road; Suite 240
Clarksburg, MD 20871-9563 • USA
Phone: +1-301-975-9400
Email: teleconferences@foiservices.com
Web: www.foiservices.com

***Important Notice**: The information provided in this presentation by each instructor is personal opinion and does not necessarily represent the opinions of FOI, Inc. Companies relying on the information do so at their own risk and assume responsibility for any subsequent consequences and liability. The information provided does not constitute legal advice.*

 TC161130

Determining a Statistically Valid Sample Size: What Does FDA Expect to See?

Steven Walfish, MBA MS
Principal Scientific Liaison, USP

An FOI Services Teleconference

November 2016

Agenda

- Basic risk management principles
- The relationship of risk to sample size
- Tools for continuous data
  - One-sample
  - Individual values
  - One proportion
  - Capability index
  - Non-Inferiority
- Tools for attribute data
  - ANSI/ASQ Z1.4
  - C=0
  - Square root N plus 1
- A case study and examples

What is *Unbiased & Representative*?

- The word bias is thrown around in the statistical literature. What does it mean?
  - "Bias is a quantitative term describing the difference between the average of measurements made on the same object and its true value."
  - The concept of unbiased means the sample is representative of the population.
- The problem is that an inadequate sample, or a poorly selected sample will induce bias.

Sampling Plans are Poorly Written

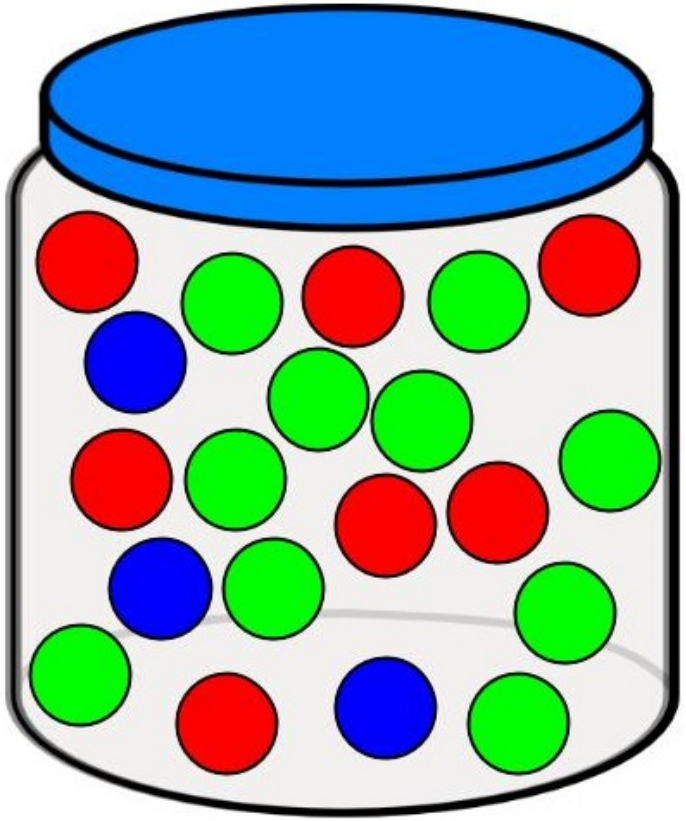
- Most documents that detail a sampling plan state the sample size, but not the sampling method.
  - "Measure the pH of ten samples"
  - "Inspect thirty labels"
- Should be written:
  - "Measure the pH of ten samples throughout the process"
  - "Inspect thirty labels, ten each from the beginning, middle and end."
- Sampling plans should also reflect the acceptance levels and risks levels.
  - Inspect thirty labels, ten each from the beginning, middle and end with no defects. This reflects 95% confidence with 90% reliability.

Principles of Risk Management

- Two primary principles of quality risk management are:
  - The evaluation of the risk to quality should be based on scientific knowledge and ultimately link to the protection of the patient; and
  - The level of effort, formality and documentation of the quality risk management process should be commensurate with the level of risk.
- Sample size is a function of risk

Representative Sampling

- Decisions are often based on our analysis of a sample.
- How we conduct a sample is very important.
  - Want:
    - Minimize bias
    - Sample reflects the characteristics of the lot or batch
    - Economical sample size



Sampling Selection

- Simple random sampling (SRS) ensures that all units are selected independently with equal probability (i.e. raffle).
- Stratified random sampling ensures that each strata (subgroup) are represented in the sample.
- Composite sampling combines several samples into a single sample unit (10 tablets used to make a single sample prep).
- Systematic sampling is a convenient sampling method ensuring that items from the beginning, middle and end are sampled.

Which is the Best?

- **As all good statisticians say…**
  - **It depends!**
- The selection of a sample is based on the answer needed to the question to be asked
- It is not a one size fits all problem

Sampling Error

- Sampling has inherent risks and potential error.
- The number of samples should be sufficient to minimize the risks.
- Cost versus benefit

Decision	*Reality*		
		Accept	**Reject**
	Accept	Correct Decision (green)	Type II Error (β) Consumer Risk (red)
	Reject	Type I Error (α) Producer Risk (red)	Correct Decision (green)

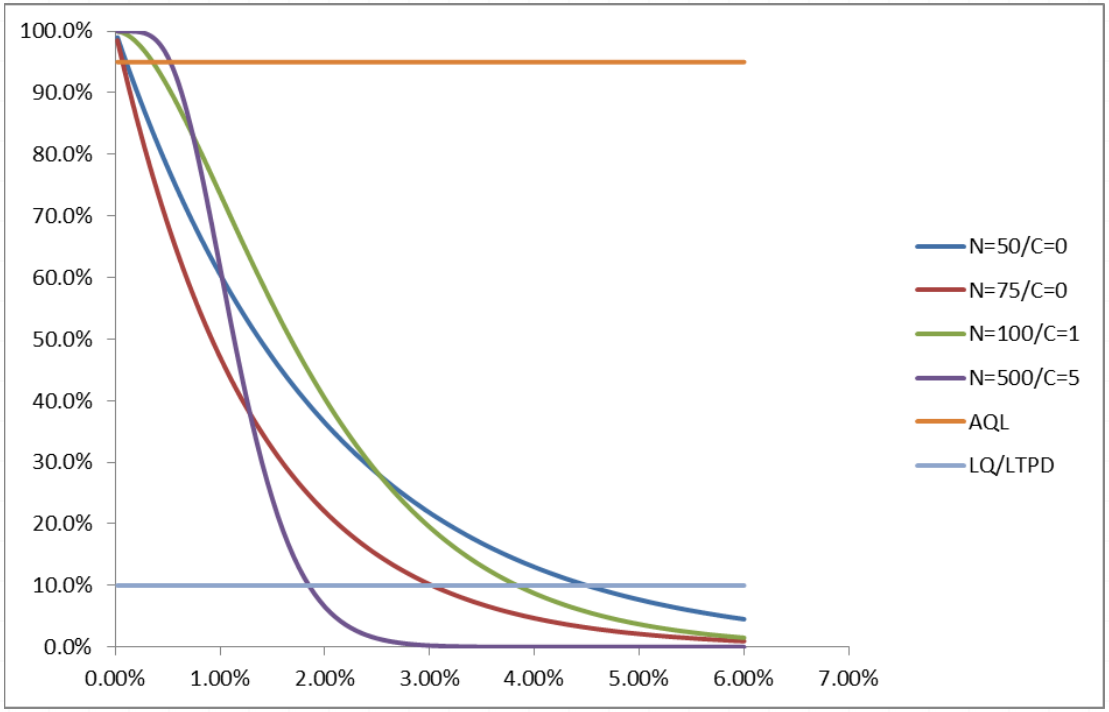
Definitions

- AQL – Acceptable Quality Level is the process average percent defective we will accept with high probability (usually set at 95%)
- LQ – Limiting Quality is the process average percent defective we will accept with low probability (usually set at 10%)
- Accept Number – Maximum number of defectives allowable in our sample and still accept the "lot or batch"
- Reject Number – Minimum number of defectives allowable in our sample that will cause "lot or batch" to be rejected
- Sample Size – The number of units to be tested from the lot or batch.
- Sampling Plan – A combination of sample size, accept/reject number with AQL and LQ levels.

OC Curve

- The operating characteristic (OC) curve depicts the power of an acceptance sampling plan.
- Its a plot of the probabilities for accepting a lot given a fraction defective.
- Every sampling plan has an OC Curve associated with it.

OC Curve



N	Accept Number (C)	AQL	LQ
50	0	0.101%	4.507%
75	0	0.069%	3.025%
100	1	0.355%	3.833%
500	5	0.526%	1.851%

Reportable Value

- A reportable value is an average value of readings from one or more units of a test solution.
- Taking three values and averaging them into a single reportable value has a smaller error.
- This plays an important role in determining sample size.

Reportable Value - Example

- Taking 6 individual samples with a standard deviation 0.58 has a 95% confidence interval of +/- 0.61
- Taking 6 sets of three value reduces the width of the interval.
- Determining which replication strategy to use is a function of the sources of variability seen in obtaining a reportable value.

Sampling & Measurement Error

Two sources of "error"

1) The variability of the sample statistic around the population parameter – standard deviation.

2) The variability of the measurement itself due to the instrument we are using.
   - Measuring the same unit repeatedly.

Sampling & Measurement Error

- Minimizing the variation:
  - To get a more precise estimate of the population parameter, take a larger sample.
    (i.e., more individual sampling units)
  - To obtain a more precise measurement, measure the same individual sampling unit multiple times (replicates) and take the average.

Sample Size for Mean

$$n = \frac{\left(Z_\alpha + Z_\beta\right)^2 S^2}{\Delta^2}$$

- Z_α and Z_β are the Type I and Type II error rates.
- You can substitute the t-value for smaller sample sizes.
- Differs slightly from the method used by software such as JMP™.

Example

- Type I error rate is 5% (95% Confidence)
  - $Z\alpha=1.96$
- Type II error rate is 1% (99% Power)
  - $Z\beta = 2.326$
- Standard deviation of 1 and a delta of 0.5.

$$n = \frac{(1.96 + 2.326)^2 (1)^2}{0.5^2}$$

$$n = 74$$

Sample Size for Individual Values

- Can develop sample sizes for individual values instead of the mean.
- Requires a confidence level and percent of future values expected to be in the interval (coverage or reliability).
- This approach is for all future observations (beta-content approach)

$$\overline{X} + k * S < U$$

$$\overline{X} - k * S > L$$

- Two-sided tolerance limits for normal populations-Some improvements. Howe, W. G. 1969, Journal of the American Statistical Association, Vol. 64, pp. 610-620)

Example

- How large a sample do I need to have 95% confidence with 95% reliability for a mean of 10.0 and standard deviation of 0.55?

- The specification are 8.5 to 12.0

- Since the mean is closest to the lower specification

$$10 - k * 0.55 > 8.5$$

$$k = 2.727$$

Sample size is approximately 21

What is the Difference?

- We collected 20 results for purity of a supplement
  - Mean = 95.1%
  - Std Dev. = 6.97
  - LSL = 90
  - Min = 92
  - Max= 95

Which is Right?

- Want to test if the MEAN is greater than 90.
  - The 95% Lower Bound on the mean of 95.1 is 92.4. This implies if we sample 20 units from a population whose mean is 95.1 with SD=6.97, the average will be greater than 92.4 approximately 95% of the time.

- Want to test if the INDIVIDUAL purity is greater than 95.
  - The lower bound with 95% confidence and 99% coverage (reliability) is on the distribution of the individuals is 72.1. This implies if we sample 20 units from a population whose mean is 95.1 with SD=6.97, 99% of the individual values would be greater than 72.1.

Which is Right

- The mean of 20 results would meet the specification.
- A sample of 20 results would have a high probability of failure.
- Must be careful to specify what is the reportable value for the testing.

Sample Size for Proportions

$$n = \left(\frac{z_{1-\frac{\alpha}{2}} \sqrt{p_0 q_0} + z_{1-\beta} \sqrt{p_1 q_1}}{p_1 - p_0} \right)^2$$

Where

- α = alpha level of the test (two-sided)
- 1 – β = power of the test
- p_0 = proportion under the null hypothesis
- p_1 = proportion under the alternative hypothesis

Example

How large a sample is needed to test H_0: $p = .21$ against H_a: $p = .31$ at $\alpha = 0.05$ (two-sided) with 90% power?

$$n = \left(\frac{1.96\sqrt{(0.21)(0.79)} + 1.28\sqrt{(0.31)(0.69)}}{0.31 - 0.21} \right)^2$$

$$= 193.3 \Rightarrow 194$$

Note: The symbol $\Rightarrow$ means round up to ensure stated power

Power Calculation

$$1-\beta = \Phi\left(\frac{|p_1 - p_0|\sqrt{n} - z_{1-\frac{\alpha}{2}}\sqrt{p_0 q}_0}{\sqrt{p_1 q}_1}\right)$$

Where:

- α = alpha level of the test (two-sided)
- n = sample size
- p_0 = proportion under the null hypothesis
- p_1 = proportion under the alternative hypothesis

Power - Example

What is the power of testing H_0: $p = .21$ against H_a: $p = .31$ at $\alpha = 0.05$ (two-sided) when $n = 57$?

$$1-\beta = \Phi\left(\frac{|0.31-0.21|\sqrt{57}-1.96\sqrt{(0.21)(0.79)}}{\sqrt{(0.31)(0.69)}}\right)$$

$$= \Phi(-0.09)$$

$$= 0.4641 \text{ (or about 46\%)}$$

Sample Size for Capability

- Some companies are calculating sample size based on capability indices.
- This can only be done using a confidence interval approach.
- Have to set a lower limit on the capability index and an observed capability index.

$$\hat{C}pk\left[1-Z_{\alpha/2}\sqrt{\frac{1}{9n(\hat{C}pk)^2}+\frac{1}{2(n-1)}}\right] \le Cpk \le \hat{C}pk\left[1+Z_{\alpha/2}\sqrt{\frac{1}{9n(\hat{C}pk)^2}+\frac{1}{2(n-1)}}\right]$$

Capability Sample Size Example

$$\text{Lower Limit on Cpk} = \text{Observed } C_{pk} - Z_{\alpha}\sqrt{\frac{1}{9n} + \frac{\text{Observed } C_{pk}^2}{2n-2}}$$

95% Confidence that the "True" Cpk is no worse than 1.0

Sample Size	Observed Cpk
3	5.69
4	3.10
5	2.44
6	2.13
7	1.95
8	1.83
9	1.73
10	1.67

Non-Inferiority

- The “at least good as” criterion.
- A one-sided significance test to reject the null hypothesis that standard therapy is better than experimental therapy by a clinically acceptable amount.
- To demonstrate that a new device is 'at least as good as' an existing device, a statistical test or confidence interval procedure must rule out clinical inferiority with a high probability.

Sample Size - NonInferiority

The noninferiority hypothesis and sample size are attributable to Blackwelder (1982). The sample size computation is given below for equal sample size in each treatment group.

$$n \geq \left(\frac{(Z_{1-\alpha} + Z_{1-\beta})}{(P_T - P_C - \delta)} \right)^2 \left[P_T(1-P_T) + P_C(1-P_C) \right]$$

Blackwelder, W. "Proving the null hypothesis in clinical trials." *Controlled Clinical Trials* 3: 345-353, 1982.

Example

In a group of patients using an approved sleeping apparatus, 80% claim an improvement in their insomnia.

The new device is proved to be non-inferior to the predicate if the rate of insomnia is no worse than 75% (a=0.05, Power=80%).

$$790 \geq \left(\frac{(1.645 + 0.841}{\left(0.80 - 0.80 - 0.05\right)} \right)^2 \left[0.8(1 - 0.8) + 0.8(1 - 0.8) \right]$$

Attribute Sampling Plans When All Units Pass

- At 95% Confidence and 90% Reliability
- $n = \frac{\ln(1-Confidence)}{\ln(Reliability)} = \frac{\ln(1-0.95)}{\ln(0.90)} = 28$

Confidence	Reliability	Sample Size
80%	80%	8
80%	90%	16
80%	95%	32
80%	99%	161
90%	80%	11
90%	90%	22
90%	95%	45
90%	99%	230
95%	80%	14
95%	90%	29
95%	95%	59
95%	99%	299
99%	80%	21
99%	90%	44
99%	95%	90
99%	99%	459

ANSI Z1.4

- Z1.4 is a standard developed for incoming inspection where the attribute is pass/fail.
- ANSI Z1.4 system is a collection of sampling plans with switching rules.
- Plans are intended primarily to be used for a continuing series of lots or batches.

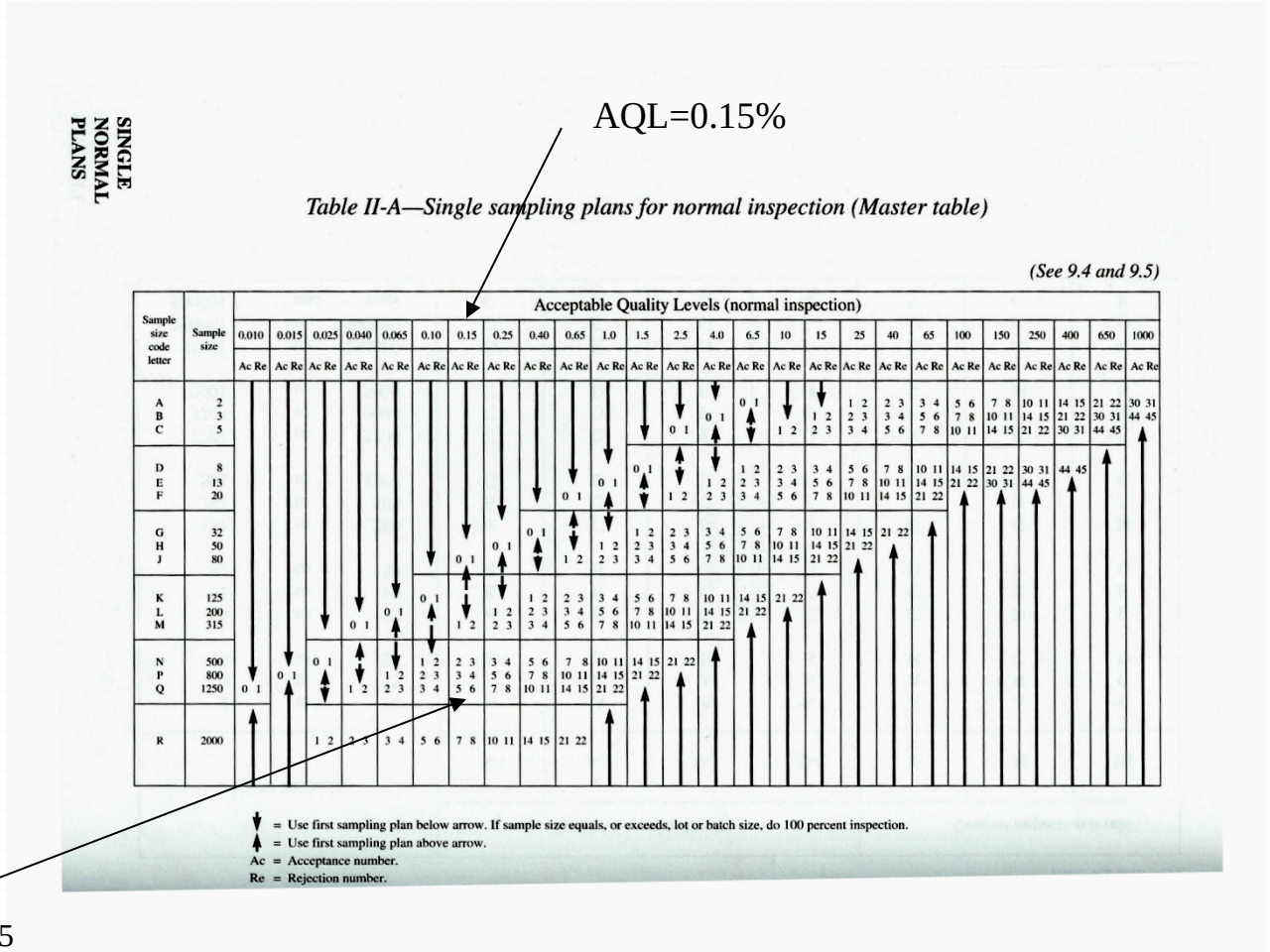
Attribute Sampling Plans

AQL: Acceptable Quality Level "is the maximum percent nonconforming (or the maximum number of nonconformities per hundred units) that, for purposes of sampling inspection, can be considered satisfactory as a process average." §4.2

Note: AQL is not lot or batch specific but rather a process average.

AQL is stated in the standard as a percent:
an AQL = 0.15 is a rate of 0.15 nonconforming units per 100 units or 0.15%.

ANSI Z1.4 – Normal Inspection



SINGLE NORMAL PLANS

AQL=0.15%

Table II-A—Single sampling plans for normal inspection (Master table)

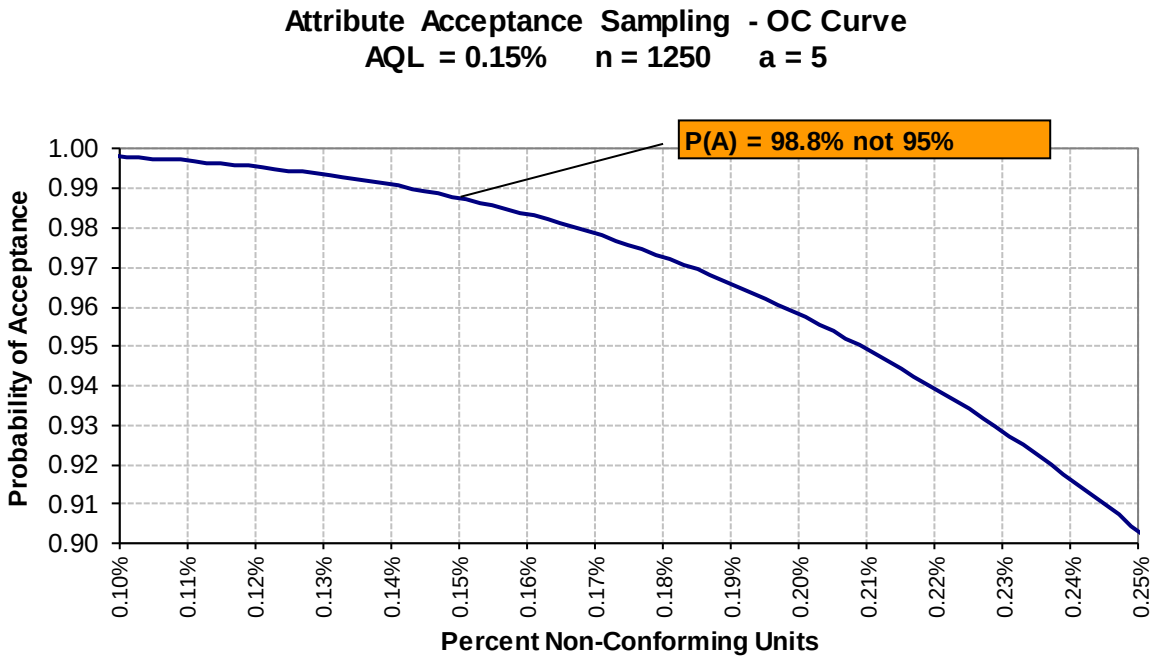
(See 9.4 and 9.5)

Sample size code letter	Sample size	Acceptable Quality Levels (normal inspection)																													
		0.010	0.015	0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1000
		Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A	2	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	0 1	↑	↓	1 2	1 2	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31
B	3	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45	↑
C	5	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	5 6	7 8	10 11	14 15	21 22	30 31	44 45	↑	↑
D	8	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45	↑	↑	↑	↑
E	13	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑	↑	↑
F	20	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑	↑	↑
G	32	↓	↓	↓	↓	↓	↓	↓	↓	0 1	0 1	0 1	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
H	50	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑	↑	↑
J	80	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑	↑	↑
K	125	↓	↓	↓	↓	↓	0 1	0 1	0 1	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
L	200	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑	↑	↑
M	315	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑	↑	↑
N	500	↓	↓	0 1	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑	↑	↑
P	800	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑	↑	↑
Q	1250	0 1	↑	↓	1 2	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↑	↑	↑	↑
R	2000	↑	↑	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑

↓ = Use first sampling plan below arrow. If sample size equals, or exceeds, lot or batch size, do 100 percent inspection.
↑ = Use first sampling plan above arrow.
Ac = Acceptance number.
Re = Rejection number.

N=1250, acc=5

Power of Attribute Sampling Plans



$\sqrt{N} + 1$

- $\sqrt{N} + 1$ was an unpublished report by the AOAC in 1927.
- It was intended to be a quick rule of thumb for inspection of foods.
- Since it was unpublished, there was not a description of the statistical basis of it.

$\sqrt{N} + 1$

- There is no known statistical justification for the use of the square root of n plus one' sampling plan.

- "Despite the fact that there is no statistical basis for a 'square root of n plus one' sampling plan, most firms utilize this approach for incoming raw materials."

  Henson, E., A Pocket Guide to CGMP Sampling, IVT.

Compare the Plans

- ANSI/ASQ Z1.4
- Lot Size N=1000
- Sample size n=32
- Acceptance Ac=0
- Rejection Re=1
- AQL=0.160%
- LQ = 6.94%

- Square root N plus one
- Lot Size N=1000
- Sample size n=33
- Acceptance Ac=0
- Rejection Re=1
- AQL=0.153%
- LQ = 6.63%

Case Study

- Have a new product that requires potency of 45± 5. During development, the average was found to be 45.2 with a SD of 3.2.
- What sample size do we need to have 95% confidence with 99% reliability that the average will meet the requirement?
- What sample size do we need to have 95% confidence with 99% reliability that the individual values will be the requirement?

Calculations (Mean)

$$N = \frac{\left(z_\alpha + z_\beta\right)^2 s^2}{\Delta^2}$$

$$N = \frac{(1.96 + 2.326)^2 * 3.2^2}{(50 - 45.2)^2}$$

$$N = 8.2$$

Round up to 9

Calculations (Individual)

$$\overline{X} + k * S < U$$

$$45.2 + k * 3.2 < 50$$

$$k = 1.5$$

$$N > \infty$$

Why the Difference?

- A mean difference of 4.8 is different than having all the individual values between 40 and 50.
- The capability analysis would show a Ppk of 0.50

Parting Thoughts

- There is no RIGHT answer, just different options. It depends on the risks you are willing to take
- Sampling plans can be developed for continuous or attribute characteristics
- It is important to know what you want your sample to tell you.
- Reportable units what sampling size calculations are based.

Thank You! Questions?

All questions are welcome
To protect your privacy, anyone asking a question will be announced only by the first name of the person reported to the operator at initial dial-in. However, anyone in your group is welcome to ask a question. Please try to avoid using a speakerphone while asking your question.

If you think of a question later, or prefer to research an issue privately, you are invited to contact the speaker at steven@statisticaloutsourcingservices.com

We value your feedback
Please complete and return the evaluation on the next page to be entered into a drawing for a $100 amazon.com gift card and/or arrange to receive a certificate verifying your participation in the 1.5 hours of this educational session.

Handouts & audio files
For an online copy of these slides, including live links to email addresses and cited websites, go to **www.foiservices.com/safesam4166.htm** This site also describes how to order an audio file of this presentation at a 50% attendee discount.

Training Evaluation Form

TC161130

Determining a Statisically Valid Sample Size:
What Does FDA Expect to See
Steven Walfish - November 30, 2016

Win a $100 Amazon.com Gift Certificate from FOI!

Just complete this form and fax it to +1-301-569-7506 by December 7, 2016
and you will be entered in a drawing for an Amazon.com Gift Certificate

Please indicate your primary job responsibility (circle one):

Regulatory Affairs QA/QC Statistics Medical Writing Electronic Submissions General Management
Other (please specify):_____

Please rate various aspects of this course

(1 = Poor; 2=Fair; 3=Good; 4=Very Good; 5=Excellent – please circle your answers)

What is your overall rating of the course?	1	2	3	4	5
How do you rate the content of the course?	1	2	3	4	5
Were course materials clear and understandable?	1	2	3	4	5
Was the length of the course adequate to cover the content?	1	2	3	4	5
Was the instructor knowledgeable about the subject matter?	1	2	3	4	5
How would you rate the instructor overall?	1	2	3	4	5

Other comments:

Please feel free to comment on any aspect of this course, including the instructor, content, and technical arrangements:

_____

_____

_____

_____

_____

_____

_____

☐ **Check here to receive a Certification of Attendance verifying your 1.5 training contact hours for this educational session.** Please provide your legibly printed name with your signature next to it, email address, and fax this form to the number below. Your certificate will be emailed to you.

By providing the information below and faxing this page to +1-301-569-7506 by December 7, 2016 you will be entered in a drawing for a $100 Amazon.com Gift Certificate from FOI.

Your Name: _____

Company Name: _____

Email Address: _____