Designing a Statistically Sound Sampling Plan

Presented by:
Steven Walfish
President, Statistical Outsourcing Services
steven@statisticaloutsourcingservices.com
http://www.statisticaloutsourcingservices.com

Purpose and Objectives

- Objective:
  - Define different types of sampling including random, stratified and composite.
  - Create and justify your sampling plan.
  - Account for sampling and measurement error.
  - Determine the relationship between sample size, statistical power and statistical precision
  - Sampling plans for attribute data.

Sampling Plans

- Decisions are often based on our analysis of a sample.
- How we conduct a sample is very important.
  - Minimize bias
  - Representative sample
  - Sufficient size.

Sampling Plans

- Simple Random Sample
  - Each sampling unit has an equal probability of being sampled with each selection.
  - Can perform simple random sampling if:
    - Enumerate every unit of the population
    - Randomly select n of the numbers and the sample consists of the units with those IDs
    - One way to do this is to use a random number table or random number generator

Sampling Plans

- Stratified Random Sampling:
  - Population strata which may have a different distribution of variable.
  - Strata must be known, non-overlapping and together they comprise the entire population.
  - Examples:
    - Measuring Heights: Stratify on Gender
      - Strata are Male, Female
    - Clinical study: stratify on stage of cancer
    - Measuring Income: Stratify on education or years of experience

Sampling Plans

- Composite Sampling:
  - Sample n units at random
  - Form a composite of n/k units for k composite-samples; mix well
  - Take the measurement on each of the k composite-samples
  - For binary outcome (positive or negative; success or failure; yes or no, etc) with rare probability of one of the two possible outcomes then forming composites can save a lot of testing.
    - For blood screening, pool the samples from x individuals and test for rare disease. If the test is negative for disease then all x blood draws are negative. If the test is positive then test all x individually.

Sampling Methods

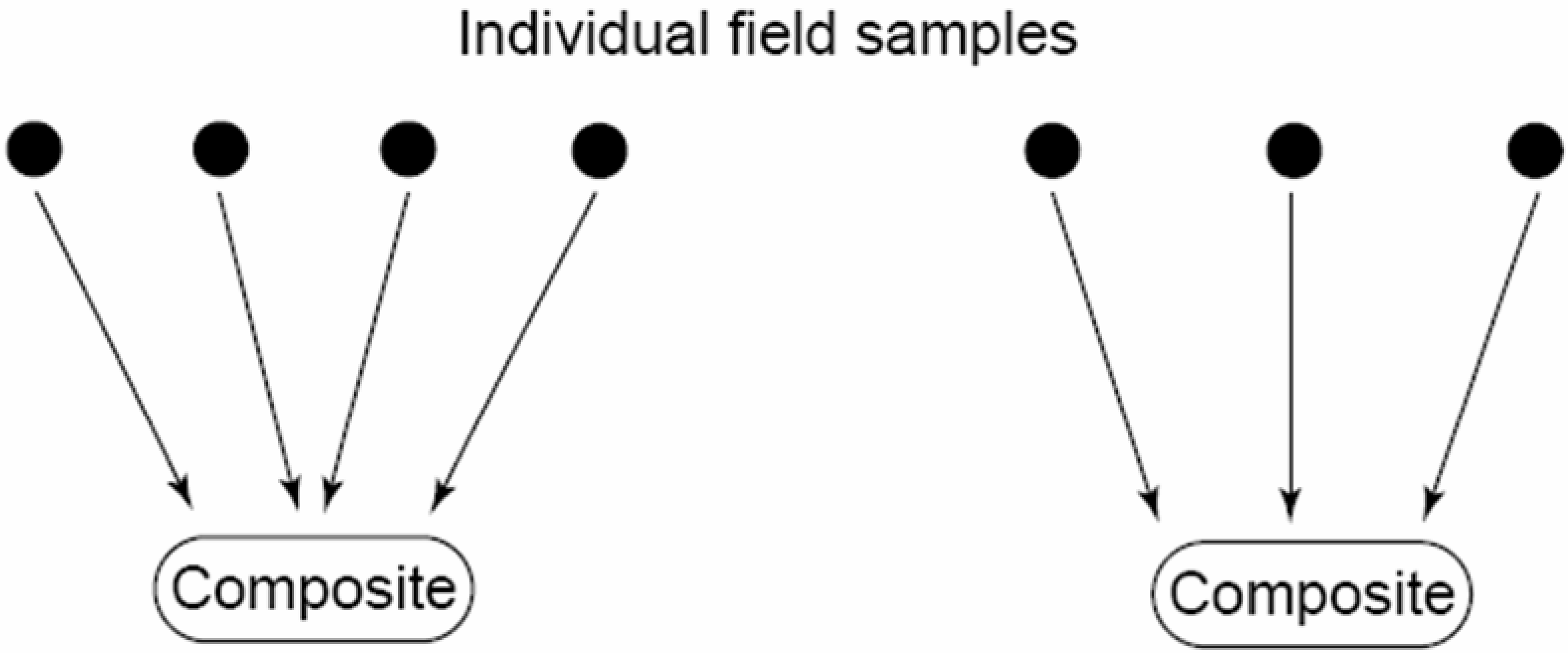


Figure 1 Forming composite samples from individual samples

Sampling Plans

- Systematic Sampling
  - Population has N units, plan to sample n units and N/n = k.
  - Line-up all N units
  - Randomly select a number between 1 and k (call it j)
  - Select the j^{th} unit and every k^{th} unit after that
  - Each unit has an equally likely chance of being selected

Sampling Plans

- Reasons for using different sampling plans:
  - Simple random sampling (SRS) ensures that all samples of size n are equally likely to be selected – units are selected independently – can use standard statistics
  - Stratified random sampling ensures that each of the strata are represented in the sample and we can construct the sample to either minimize variability of the estimator or to minimize cost
  - Composite sampling can save costs making sampling more efficient but you lose information about the individual sampling units.
  - Systematic sampling is a convenient sampling method for items coming off a line – ensures that items from the beginning, middle and end of production are sampled

Sampling Plans

- **<u>SRS uses basic statistics</u>**; estimates and standard error estimates need to be adjusted for the other sampling methods
  - For Simple Random Sampling and estimating the population mean:

$$\bar{x}_{srs} = \frac{\sum x}{n}$$

With variance (standard error squared): $$s^2(\bar{x}_{srs}) = \frac{s^2}{n}$$

Sampling Plans

- For stratified sampling and estimating the population mean:

$$\bar{x}_{st} = \frac{\sum N_h \cdot \bar{x}_h}{N}$$

With variance (standard error squared):

$$s^2(\bar{x}_{st}) = \frac{1}{N^2} \sum N_h (N_h - n_h) \frac{s_h^2}{n_h}$$

- Note that you need to know how many units are in each strata (N_h).

Sampling and Measurement Error

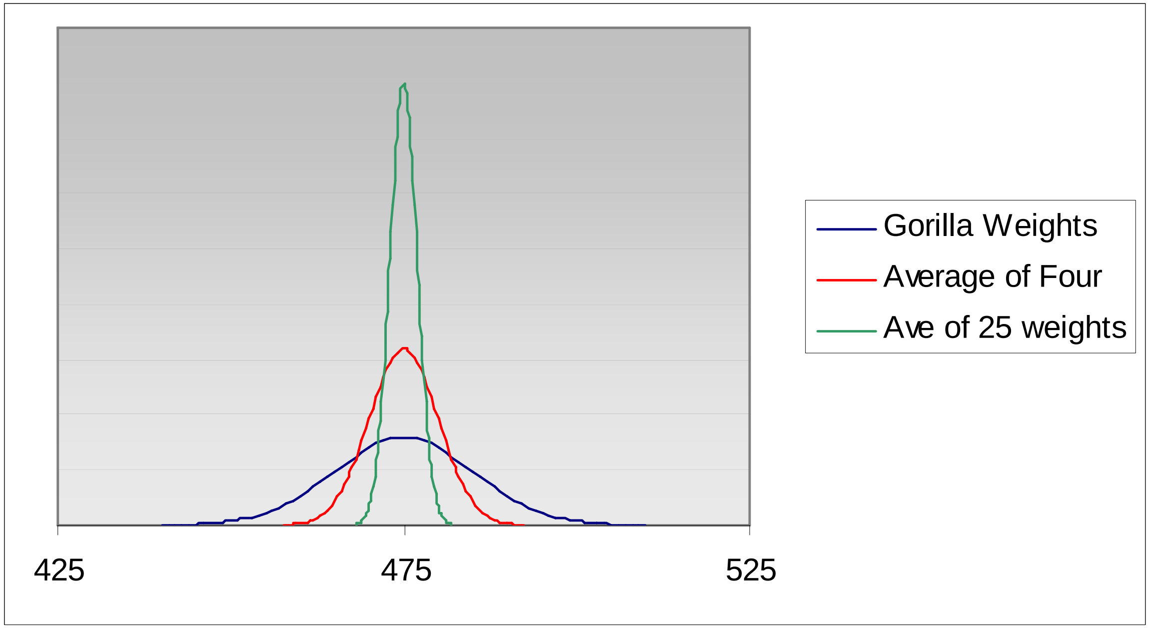
- Two sources of "error":
  - The variability of the sample statistic around the population parameter – standard deviation.
  - The variability of the measurement itself due to the instrument we are using.
    - Measuring the same unit repeatedly.

Sampling and 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, Statistical Precision, and Statistical Power

- Increasing the sample size increases the precision of the sample estimate
- If we take a large sample then the sample mean is closer (in distribution) to the population mean



Sample Size, Statistical Precision, and Statistical Power

- Increasing the sample size decreases the standard error of your estimate.

- Example: Estimating the population mean:
  - Point Estimate:

$$\bar{x}_{srs} = \frac{\sum x}{n}$$

  - 95% Confidence Interval:

$$\bar{x}_{srs} \pm t \cdot \frac{s}{\sqrt{n}}$$

Sample Size, Statistical Precision, and Statistical Power

- Standard Error is $\frac{s}{\sqrt{n}}$
- 95% Margin of error is $t \cdot \frac{s}{\sqrt{n}}$

  where t has n-1 df
  and is for 95%

- Width of confidence interval is $2 \cdot t \cdot \frac{s}{\sqrt{n}}$
- Increasing n makes each of these smaller.
  Increase sample size for better precision.

Sample Size, Statistical Precision, and Statistical Power

- Hypothesis Testing and Types of Errors

		REALITY	
		H_0 is True	H_0 is False & H_A is True
DECISION	Accept H_0	Correct Decision	Type II error with Probability β (Depends on true value of μ)
	Reject H_0	Type I error with Probability α (we get to specify α)	Correct Decision with Probability 1-β (1-β is called Power)

Statistical Power

- Statistical power is defined as the ability to detect effects when the effect is present.
- It is the probability of rejecting the null hypothesis when the alternative hypothesis is true.

Sample Size, Statistical Precision, and Statistical Power

- For a specific alternative (H_1: $\mu = 650$ in example) we can estimate the probability of deciding "Reject H_0" based on the standard error of the estimator.
- Power:
  - Increases with increase in sample size
  - Increases with increase in probability of Type I error
  - Increases as the specific alternative claim moves away from the claim of H_0.

Sample Size, Statistical Precision, and Statistical Power

- Example: Calculating a sample size to detect a given difference:

  - From historical data we know σ is approximately 50 units; we'd like a 95% confidence interval that has a margin of error (m.e.) of 16 units.

  - Margin of error = $t \cdot \frac{s}{\sqrt{n}}$

  - Use algebra to solve for n:

$$n = \left(\frac{t \cdot s}{m.e.} \right)^2$$

Sample Size, Statistical Precision, and Statistical Power

- Example: Calculating a sample size:
  - The t-value will be somewhat bigger than 2 → use 2 to start with. We can solve for n:

$$n = \left(\frac{2 \cdot 50}{16}\right)^2 = 39.06 \Rightarrow 40$$

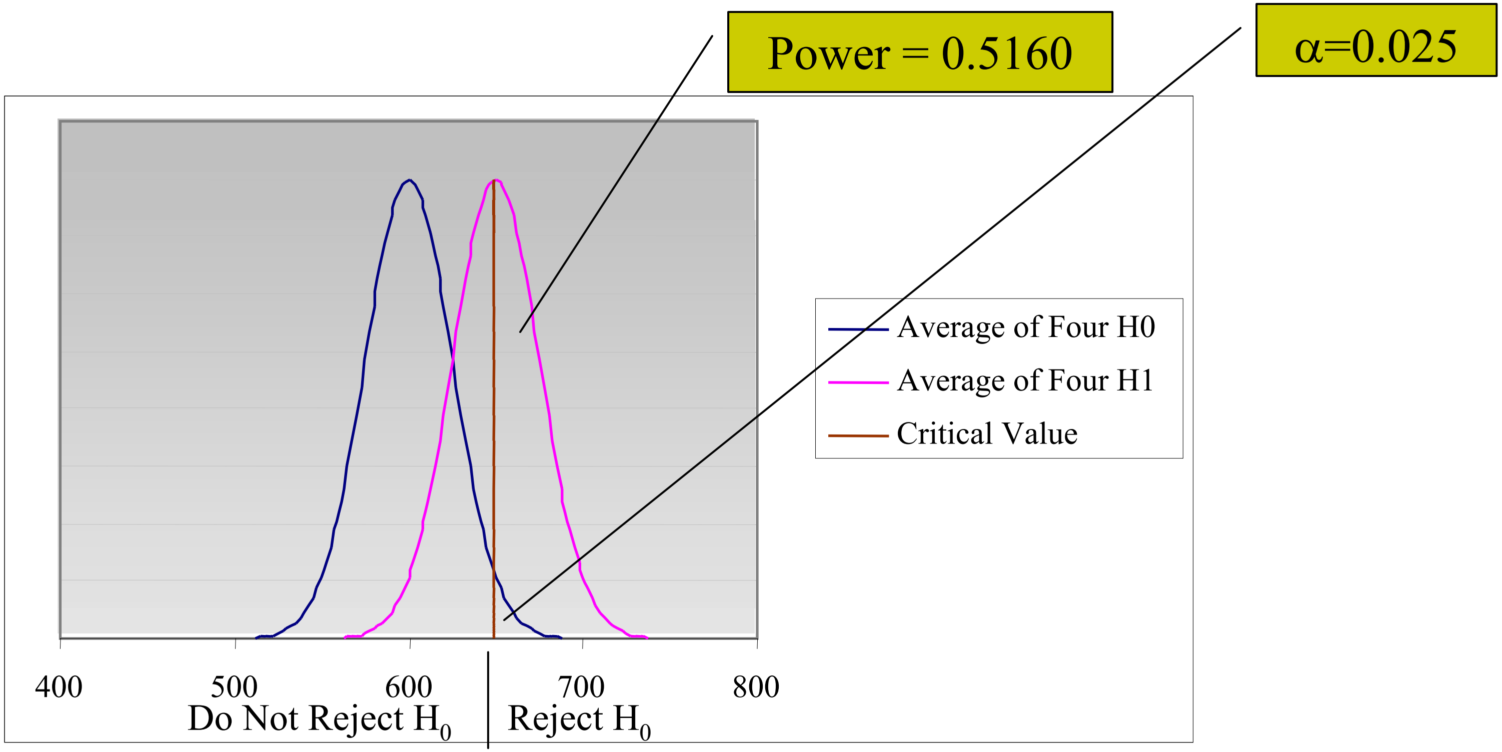
  - Then use $t_{39df} = 2.023$ and re-solve:

$$n = \left(\frac{2.023 \cdot 50}{16}\right)^2 = 39.97 \Rightarrow 40$$

  - To get the precision we'd like we need a random sample size of 40 (based on preliminary estimate of σ)

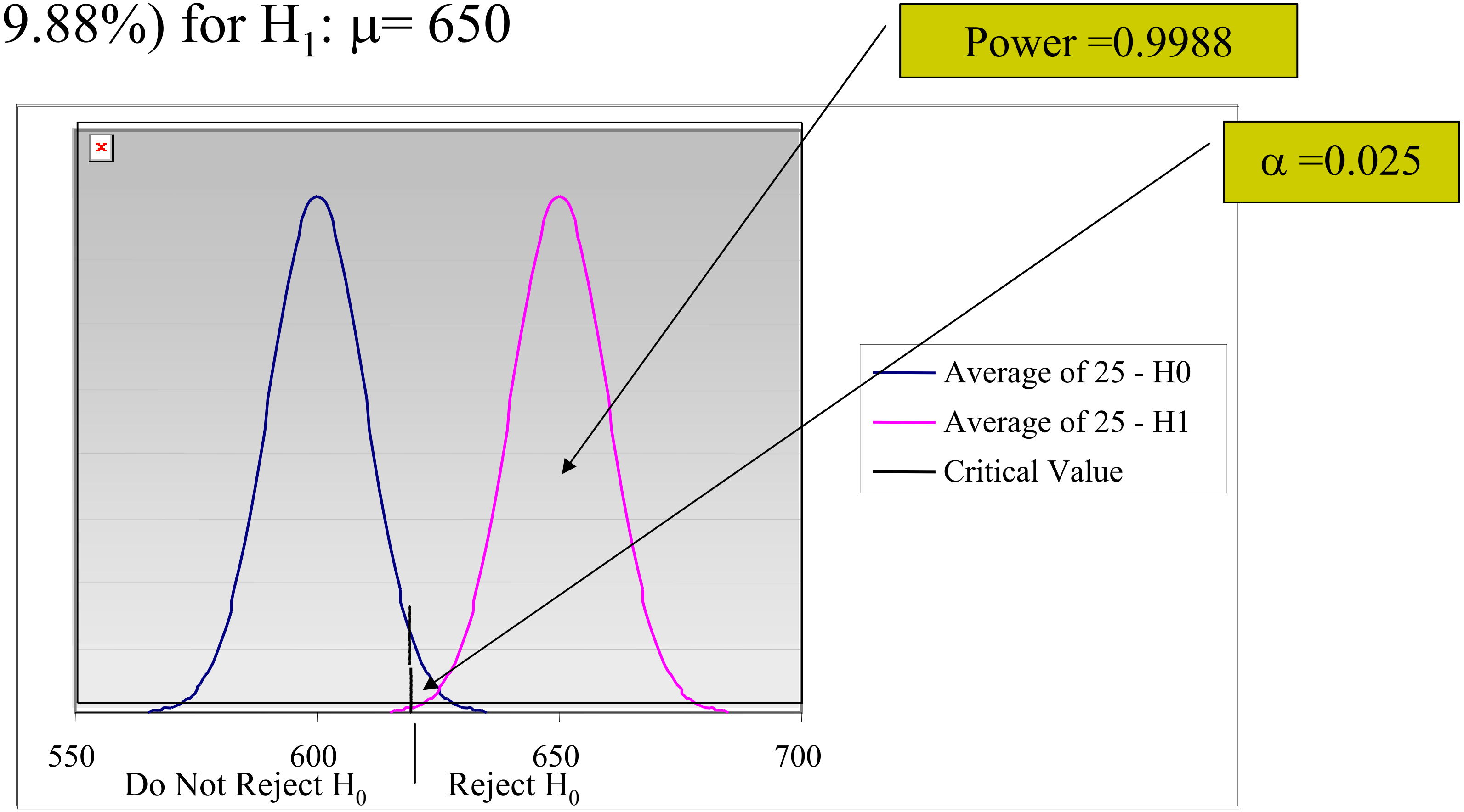
Sample Size, Statistical Precision, and Statistical Power

- H_0:μ=600,with n = 4 and α=0.025 we have power of 0.5160 (51.60%) for H_1: μ= 650



Sample Size, Statistical Precision, and Statistical Power

- H_0:μ=600,with n = 25 and α=0.025 we have power of 0.9988 (99.88%) for H_1: μ= 650



Sample Size, Statistical Precision, and Statistical Power

- With a large enough sample size we could get 90% power for a population average score of 605.

- But, is this a meaningful difference? Would it be worth throwing resources if we could prove that the new method's average test is around 605?

Attribute Sampling Plans

- ANSI sampling plans for attributes and relationship to statistical hypothesis testing
  - “Inspection by attributes is inspection whereby either the unit of product is classified simply as conforming or nonconforming, or the number of nonconformities in the unit of products is counted, with respect to a given requirement or set of requirements.” §1.4 ANSI\ASQC Z1.4 1993
  - ANSI Z1.4 system is a collection of sampling plans and switching rules.
  - “Plans are intended primarily to be used for a continuing series of lots or batches.” §1.2 ANSI\ASQC Z1.4 1993

Attribute Sampling Plans

- ANSI sampling plans for attributes and relationship to statistical hypothesis testing
  - 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%.

Attribute Sampling Plans

- ANSI sampling plans for attributes and relationship to statistical hypothesis testing
  - What you need to choose a sampling plan:
    - Lot or Batch Size
    - Inspection level
    - Single, Double or Multiple sampling
    - Normal, tightened or reduced inspection
    - AQL
  - Under AQL sampling plans if the process average is less than or equal to the AQL then each lot has a high probability of passing inspection

Attribute Sampling Plans

Lot or Batch Size	Special Inspection Levels				General Inspection Levels		
	S-1	S-2	S-3	S-4	I	II	III
2 to 8	A	A	A	A	A	A	B
9 to 15	A	A	A	A	A	B	C
16 to 25	A	A	B	B	B	C	D
26 to 50	A	B	B	C	C	D	E
51 to 90	B	B	C	C	C	E	F
91 to 150	B	B	C	D	D	F	G
151 to 280	B	C	D	E	E	G	H
281 to 500	B	C	D	E	F	H	J
501 to 1200	C	C	E	F	G	J	K
1201 to 3200	C	D	E	G	H	K	L
3201 to 10000	C	D	F	G	J	L	M
10001 to 35000	C	D	F	H	K	M	N
35001 to 150000	D	E	G	J	L	N	P
150001 to 500000	D	E	G	J	M	P	Q
500001 and over	D	E	H	K	N	Q	R

Normal Inspection

AQL=0.15%

SINGLE NORMAL PLANS

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

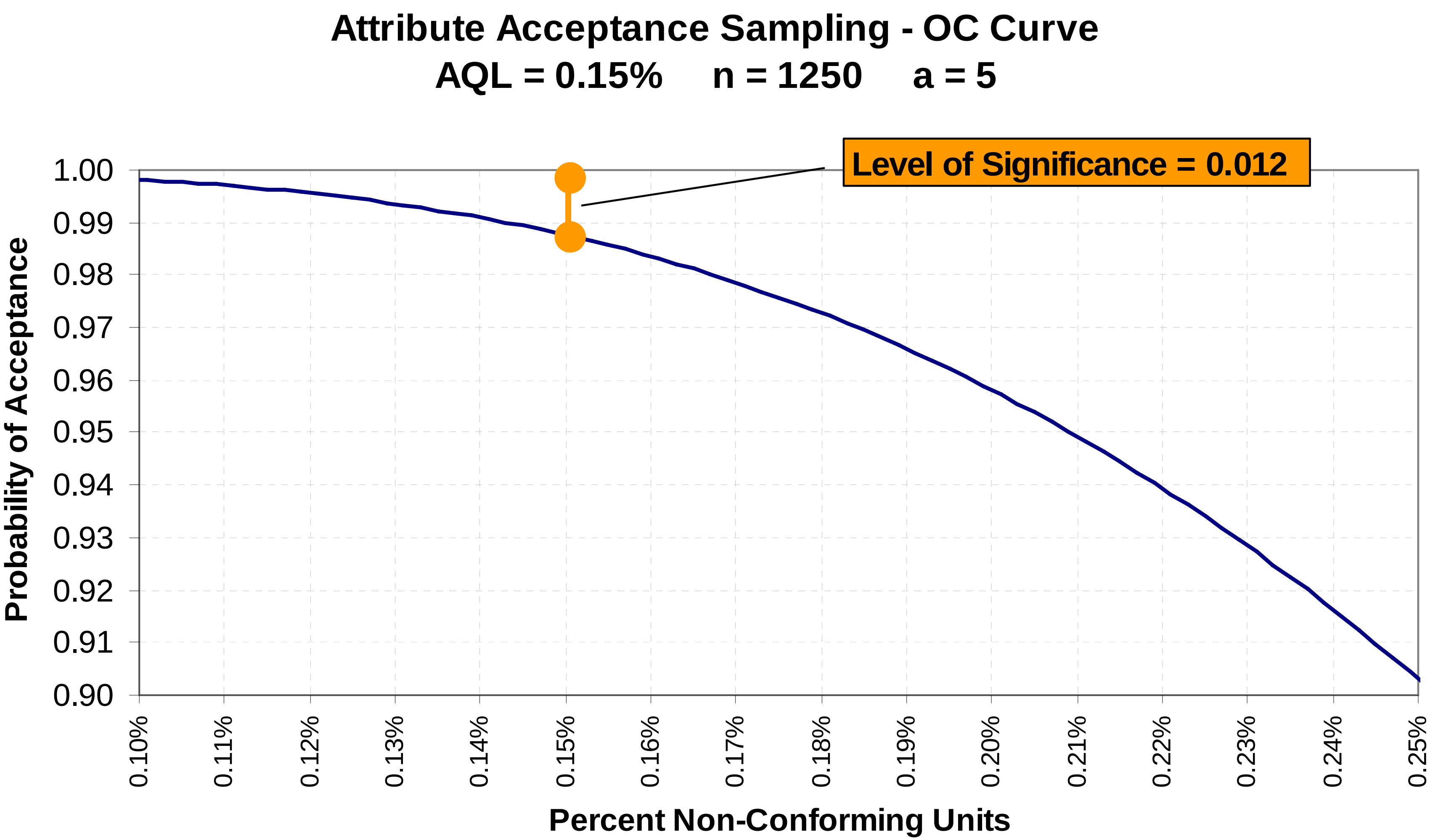
(See 9.4 and 9.5)

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	↓	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31
B	3	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45
C	5	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	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	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45	↑	↑	↑
F	20	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑
G	32	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑
H	50	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑
J	80	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑
K	125	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
L	200	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
M	315	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
N	500	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
P	800	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
Q	1250	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
R	2000	↑	↑	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑

N=1250, acc=5

↓ = 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.

Sample Size, Statistical Precision, and Statistical Power of Attribute Sampling Plans



Caveats

- Some of the caveats to look for are:
  - Lurking variables: These are variables that have an impact on the outcome variable but are not measured – often we may not even be aware of these.
  - Confounding variables: If two (or more) input variables are changing at the same time or near to the same time then it will be impossible to distinguish which variable has an impact on the outcome measure.

Caveats

- Another of the caveats to look for is:
  - Collinearity: When 2 input variables are highly correlated we have collinearity and the regression estimates are very unstable (highly variable). Although each input variable may seem to be measuring something different from a modeling perspective one of the variables is enough.
    - Extreme example: Regression of Y = weight on X = height in inches and Z = height in centimeters.
    - Less extreme example: Regression of Y= weight on X = height and Z = age for boys between the ages of 3 and 16.

Caveats

- Another of the caveats to look for is:
  - Interaction: Interaction occurs when the effect of one input variable depends on the value of another input variable. Ignoring an interaction can lead to erroneous conclusions.

Binomial Distribution

- Binomial Experiment:
  - n trials are conducted (n is known in advance) – here a trial is a unit inspected
  - on each trial there are only two possible outcomes, Success (what we're counting) and Failure – here Success is a nonconforming unit
  - on each trial, π, the probability of Success remains constant
  - the trials are independent (the outcome of any one trial does not depend on the outcome of any other trial)
  - (The last two are met by using random sampling)
- A binomial random variable is the number of successes out of n trials of a binomial experiment
- The probability of seeing x or less Successes in n trials is:

$$P(X \le x) = \sum_{k=0}^{x} \binom{n}{k} \cdot \pi^{k} \cdot (1 - \pi)^{(n-k)}$$

Binomial Distribution

- Excel has a Binomdist function to calculate these probabilities.
  - Probability of seeing 5 or less nonconforming units if the process rate is right at the AQL and we sample 1250 units is $P(X \leq 5 \mid n=1250, \pi = 0.15\%)$ and in Excel this is: =Binomdist(5,1250,0.0015,true) and function returns 0.988.
  - We have 98.8% probability of seeing LESS THAN 5 nonconforming unit if the true nonconforming rate is 0.15%; we have 1.2% probability of seeing 6 or more nonconforming units.
  - If the true process rate for nonconforming units is 0.15% (or 15 units out of 10,000 units) then for 100 lots approximately 99 lots will be accepted and approximately 1 lot will be rejected.
  - If the true process rate for nonconforming units is less than 0.15% then the probability of accepting the lot will be more than 98.8%

Summary

- Representative Sampling is critical to valid statistical inference
  - Biased Sampling can result in erroneous inference
  - If a sample is too small there may be too little information to draw any conclusions
- Sampling plans can accommodate the structure of the population
  - Stratified sampling
- Caution should be taken for lurking variables, confounding variables, collinearity and interactions especially when taking a sample
- Increasing the sample size increases
  - Statistical precision
  - Statistical power

➢ Caution: Statistical Significance does not necessarily imply a Meaningful (Practical) Result