Sample Size Estimation for BMP Verification

August 14, 2014

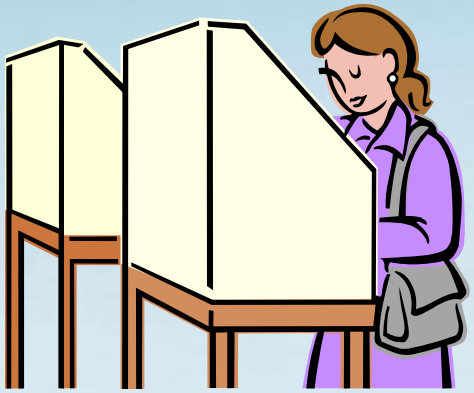
TETRA TECH

complex world | CLEAR SOLUTIONS™

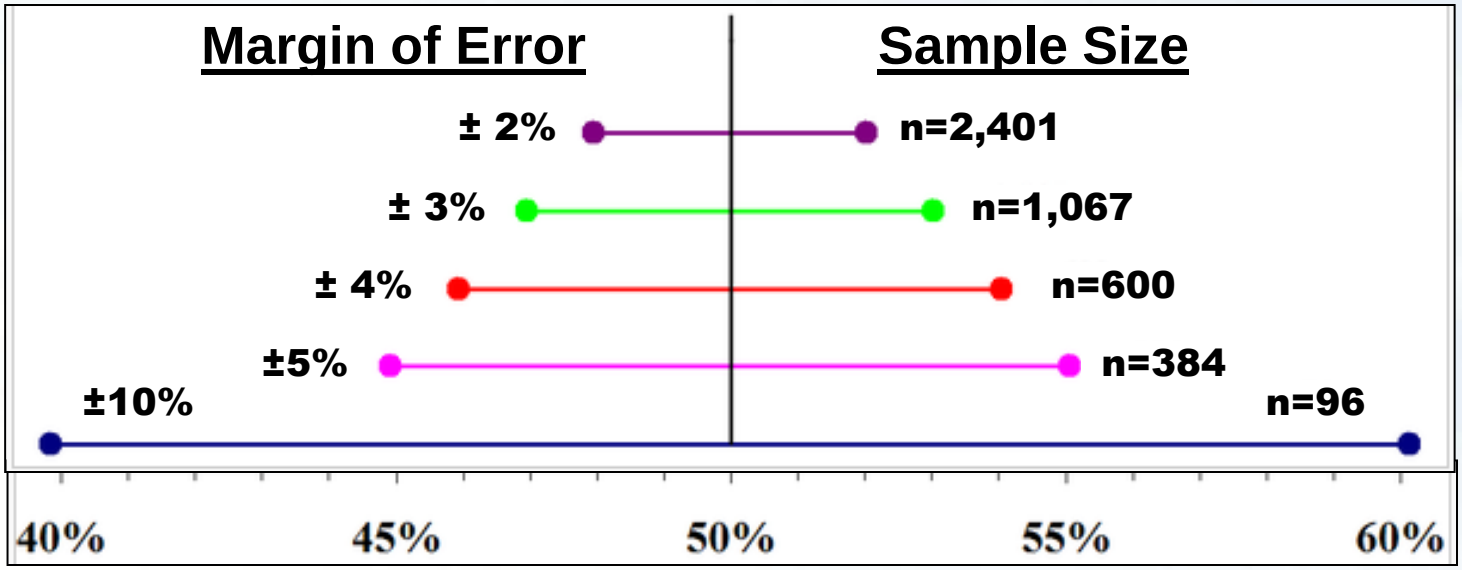
Sample Size Estimation

- Objective/Management Goal
  - Document the percentage of BMPs that are still in place and functioning properly
  - Inform Bay model simulation updates
- Target Populations
  - BMP implemented through state cost-share program
  - BMP implemented through CAFO permits
  - BMP implemented voluntarily without cost-share
  - Resource improvement practice

Binomial Distribution



- Binomial Distribution
  - Are the BMPs still there?
    - Yes/No
  - Are the BMPs still functioning properly?
    - Yes/No
- Sample Size—just like political polls



Source: http://en.wikipedia.org/wiki/Margin_of_error#Calculations_assuming_random_sampling

Some Common Terms

- N: total number of population units in sample population
  - County XYZ wrote 246 nutrient management plans (N=246)
- n: number of samples
  - County XYZ randomly inspected/reviewed the implementation of 30 nutrient management plans (n=30)
- n/N: Sampling level
  - i.e., 30/246 = 0.12 or 12%
- p: proportion of "yes" responses
  - County XYZ found that 26 of the nutrient management plans continue to be implemented as expected (26/30= 0.867 or 86.7%)
- Confidence level: certainty that p is within a certain range
  - e.g., 90% confidence level (significance level, α = 0.10)
- d: allowable error (margin of error)
  - 86.7% ±9.6% (87% ±10%) with 90% confidence
  - Upper and lower 90% confidence interval is 77-97%

Binomial Distribution

Standard Sample Size Equation

$$n_o = \frac{\left(Z_{1-\alpha/2}\right)^2 pq}{d^2}$$

Political Poll Example

$$96 = \frac{(1.96)^2(0.5)(0.5)}{(0.10)^2}$$

Finite Population Correction

$$n = \frac{n_o}{(1+\varphi)}$$

N = total number of population units in sample population

n_o = preliminary estimate of sample size (sample size for large N)

$Z_{1-\alpha/2}$ = value corresponding to cumulative area of 1-α/2 using the normal distribution

p = proportion of "yes" responses

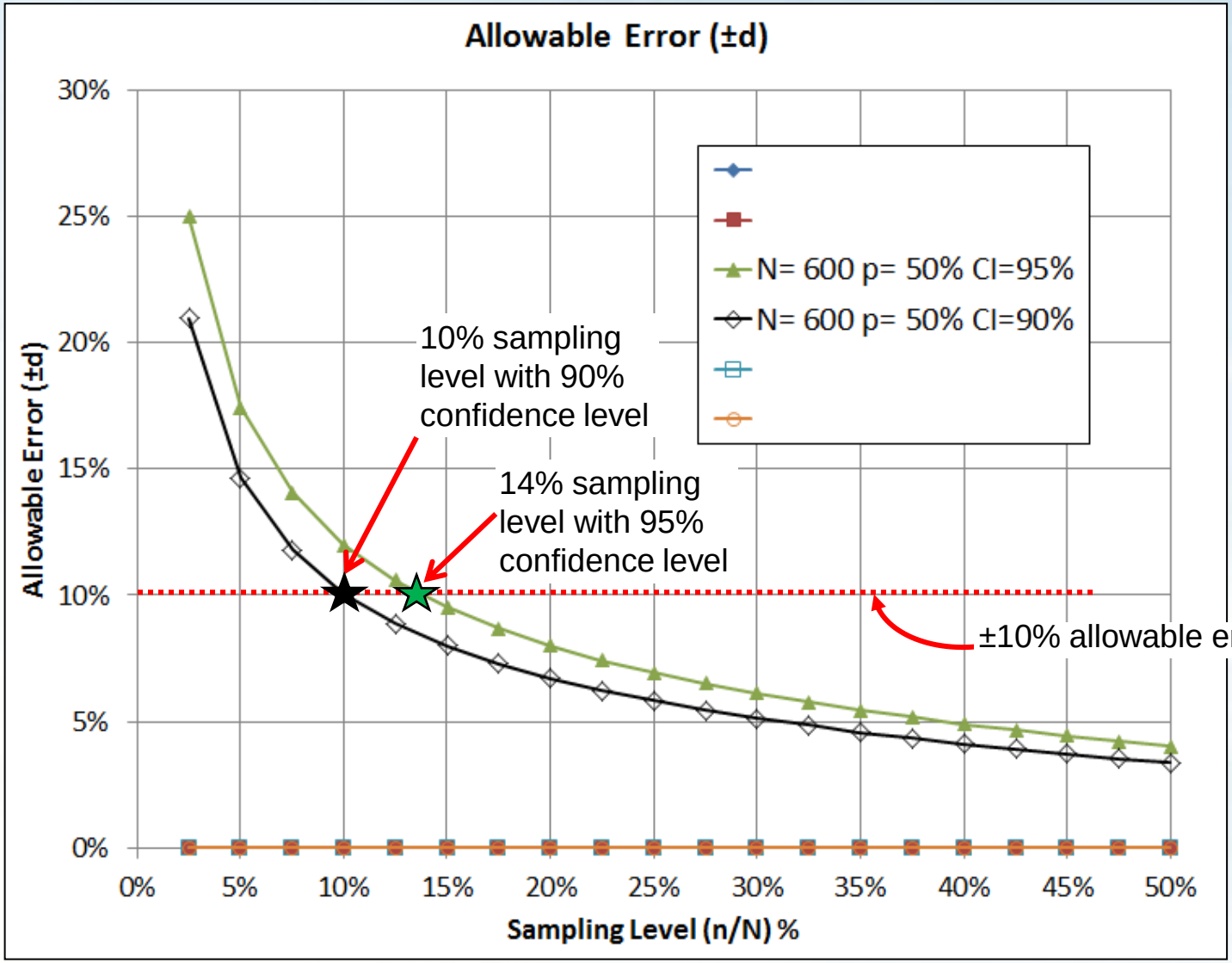
q = proportion of "no" responses (i.e., 1-p)

d = allowable error (margin of error)

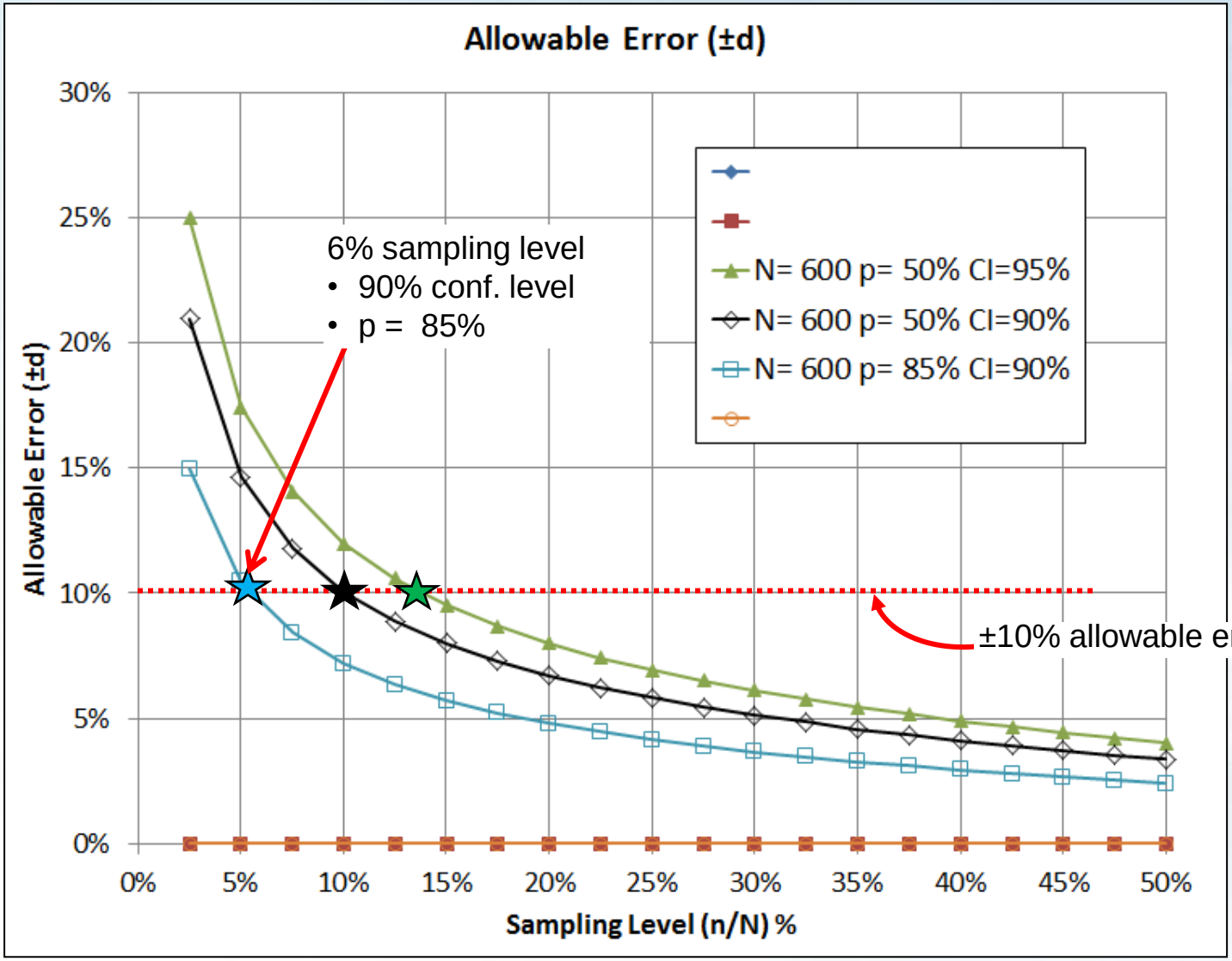
φ = n_0/N unless otherwise stated

n = number of samples (adjusted for finite population)

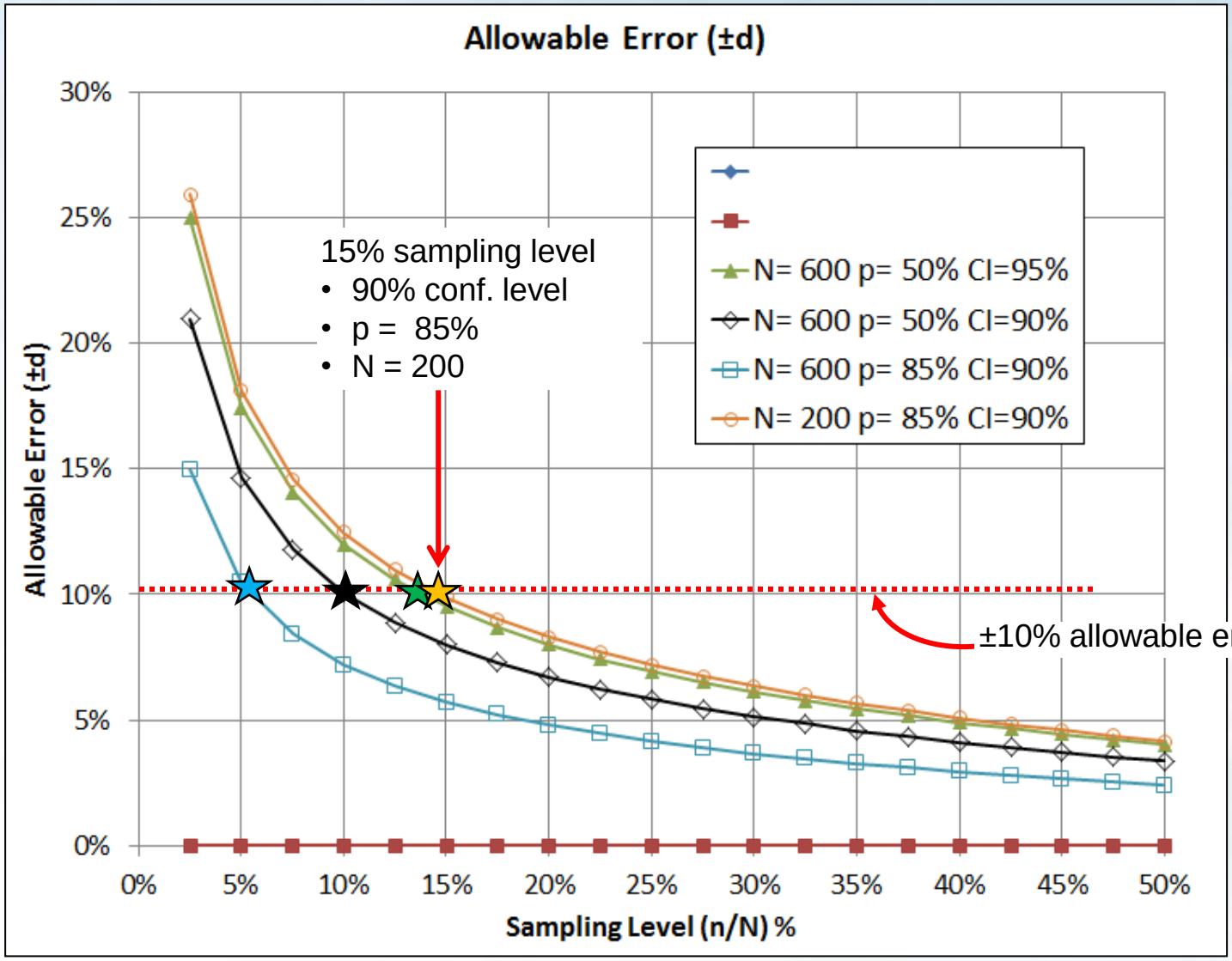
Allowable Error—Decreasing Conf. Level



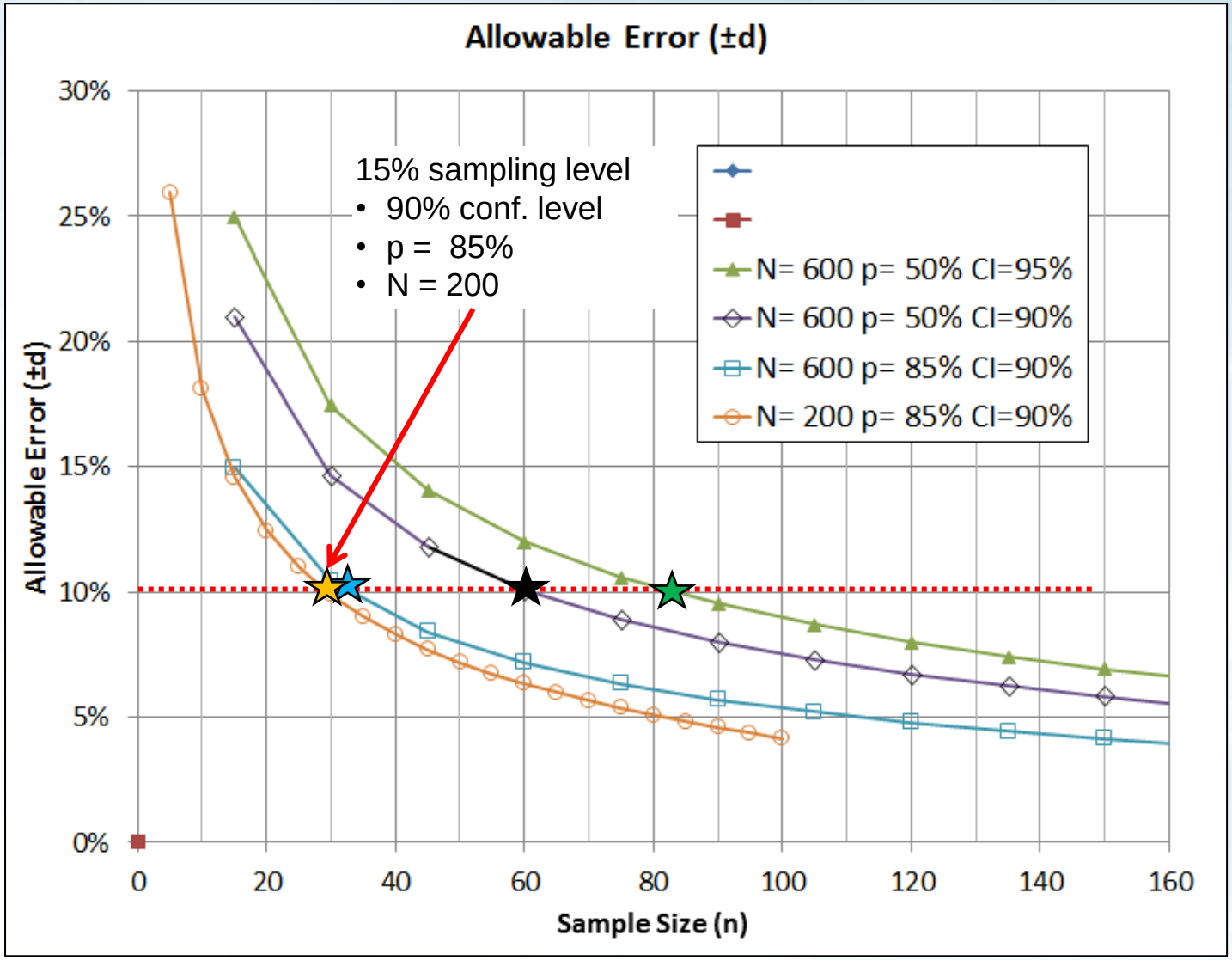
Allowable Error—A Priori Knowledge



Allowable Error—Smaller Population



Allowable Error—Smaller Population



Observations

- Improved precision
  - More sampling
- Reduce sampling costs
  - Lower confidence level (e.g., 95% CI → 90% CI)
  - Increased allowable error, d, (e.g., ±10% → ±15%)
- Less sampling is needed to maintain precision if the percentage of BMPs maintained is closer to 100%
  - A priori knowledge is important
  - 50% BMP maintenance is a conservative assumption
  - But don't overestimate
- Finite Populations
  - Sampling from small populations can result in large errors.

Potential Application

- Precision Statement
  - Estimate the percentage of BMPs maintained, **p**, to within **±d%** using a **X%** confidence level.
- Example:
  - The percentage of BMPs maintained is 85% ±10% with a 95% confidence level, or
  - The range of maintained BMPs is 75-95% with a 95% confidence interval.
- Worked Example
  - **p**: No information (50%), Good (70%), Excellent (85%)
  - **±d**: 5%, 10%, and 15%
  - **X%**: 90% and 95%

Work Example—Sample Size (n)

95% Confidence Level

p		±d	Large N	100	200	600	1000	1,500	2,000
No Information	50%	5%	385	80	132	235	278	307	323
	50%	10%	97	50	66	84	89	92	93
	50%	15%	43	31	36	41	42	42	43
Good Maintenance	70%	5%	323	77	124	210	245	266	279
	70%	10%	81	45	58	72	75	77	78
	70%	15%	36	27	31	34	35	36	36
Excellent	85%	5%	196	67	99	148	164	174	179
	85%	10%	49	33	40	46	47	48	48
	85%	15%	22	19	20	22	22	22	22

90% Confidence Level

p		±d	Large N	100	200	600	1000	1,500	2,000
No Information	50%	5%	271	74	116	187	214	230	239
	50%	10%	68	41	51	62	64	66	66
	50%	15%	31	24	27	30	31	31	31
Good Maintenance	70%	5%	228	70	107	166	186	198	205
	70%	10%	57	37	45	53	54	55	56
	70%	15%	26	21	24	25	26	26	26
Excellent	85%	5%	138	58	82	113	122	127	130
	85%	10%	35	26	30	34	34	35	35
	85%	15%	16	14	15	16	16	16	16

Work Example—Sample Level (n/N)

>20%	●
10-20%	●
5-10%	●
<5%	●

95% Confidence Level

p		±d	Large N	100	200	600	1000	1,500	2,000
No Information	50%	5%	385	80%	66%	39%	28%	20%	16%
	50%	10%	97	50%	33%	14%	9%	6%	5%
	50%	15%	43	31%	18%	7%	4%	3%	2%
Good Maintenance	70%	5%	323	77%	62%	35%	25%	18%	14%
	70%	10%	81	45%	29%	12%	8%	5%	4%
	70%	15%	36	27%	16%	6%	4%	2%	2%
Excellent	85%	5%	196	67%	50%	25%	16%	12%	9%
	85%	10%	49	33%	20%	8%	5%	3%	2%
	85%	15%	22	19%	10%	4%	2%	1%	1%

90% Confidence Level

p		±d	Large N	100	200	600	1000	1,500	2,000
No Information	50%	5%	271	74%	58%	31%	21%	15%	12%
	50%	10%	68	41%	26%	10%	6%	4%	3%
	50%	15%	31	24%	14%	5%	3%	2%	2%
Good Maintenance	70%	5%	228	70%	54%	28%	19%	13%	10%
	70%	10%	57	37%	23%	9%	5%	4%	3%
	70%	15%	26	21%	12%	4%	3%	2%	1%
Excellent	85%	5%	138	58%	41%	19%	12%	8%	7%
	85%	10%	35	26%	15%	6%	3%	2%	2%
	85%	15%	16	14%	8%	3%	2%	1%	1%