

## Number of Samples Required to Detect Violations with Predefined Probabilities

Scheduled sampling is conducted to provide some assurance of detection of a violation that affects a given percentage of the sample population.

The statistical table provides the calculated number of samples required to ensure detection of at least one violation that affects a given percentage of the sampled population. Statistically, for a binomial distribution with sample size “ $n$ ” and violation rate “ $v$ ” (in decimal), if  $v$  is the true violation rate in the population and  $n$  is the number of samples, the probability,  $p$ , of finding at least one violation among the  $n$  samples (assuming random sampling) is  $p = 1 - (1 - v)^n$

For example, if the true violation rate is 1% the probability of detecting at least one violation with sample sizes of 230,300,390,460, and 800 are 90%, 95%, 98%, 99%, and 99.97% respectively.

In the table below the probability of detecting at least one violation with a sample size of 800 is italicized and bolded.

Percentage % Violative in the population (v)	Number of samples required to detect at least one violation in (n) samples with a probability (p)				
	0.90	0.95	0.98	0.99	0.9997
	Sample Size required “n”				
10	22	29	37	44	77
5	45	59	76	90	158
1	230	300	389	459	<b><i>807</i></b>
0.57	403	525	684	<b><i>806</i></b>	1,419
0.50	460	598	<b><i>780</i></b>	919	1,618
0.37	620	<b><i>808</i></b>	1,055	1,242	2,188
0.29	<b><i>793</i></b>	1,032	1,347	1,586	2,793
0.10	2,302	2,995	3,910	4,603	8,108

The procedure to calculate the required sample size needed:

$p = 1 - (1 - v)^n$	← Probability of detecting at least one violation in n sample of binomial distribution with violation rate v
$1 - p = (1 - v)^n$	← Subtract one from both side of the equation. This gives the probability of detecting No violations in n samples
$\log(1 - p) = \log(1 - v)^n$	← Apply logarithmic function to both side of the equation
$\log(1 - p) = n * \log(1 - v)$	← A logarithmic function property
$n = \frac{\log(1-p)}{\log(1-v)}$	← Sample size based on violation rate (v) and probability of detecting (p)