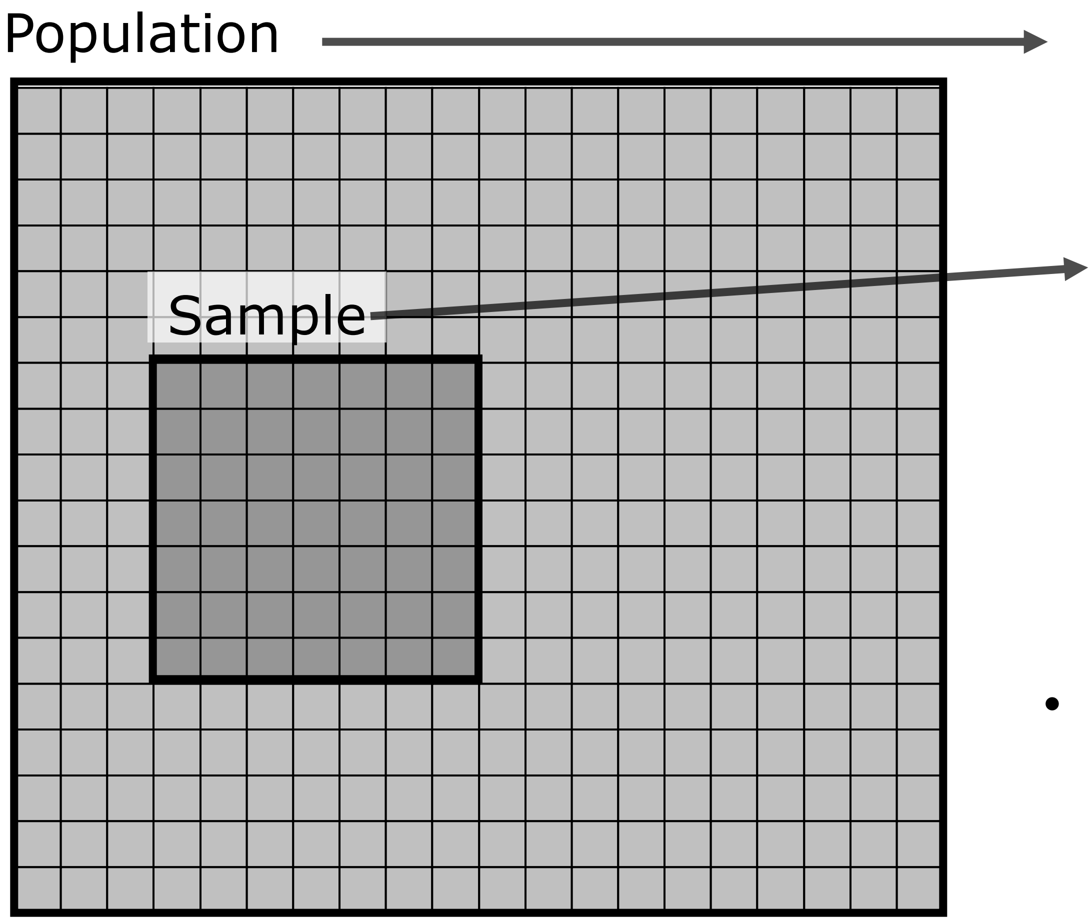
ECE 510 Lecture 6
Confidence Limits

Scott Johnson
Glenn Shirley

Concepts

Statistical Inference

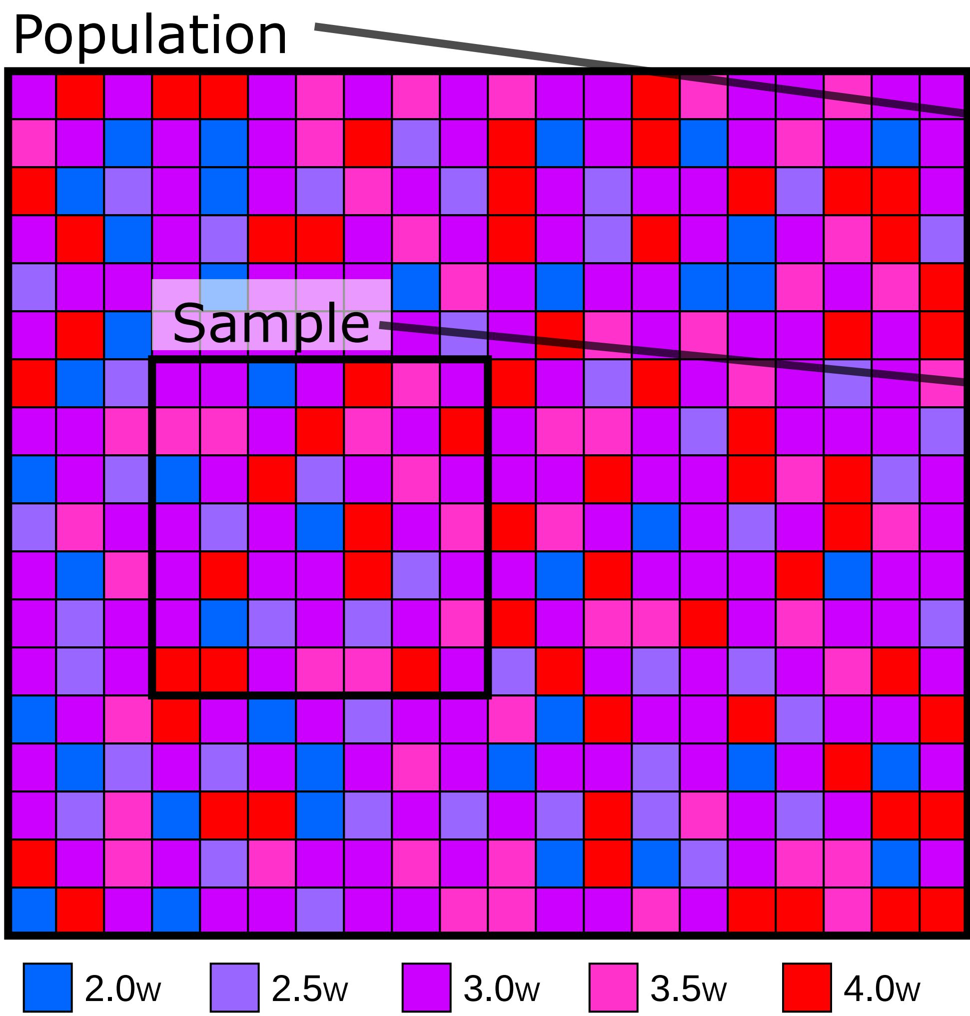


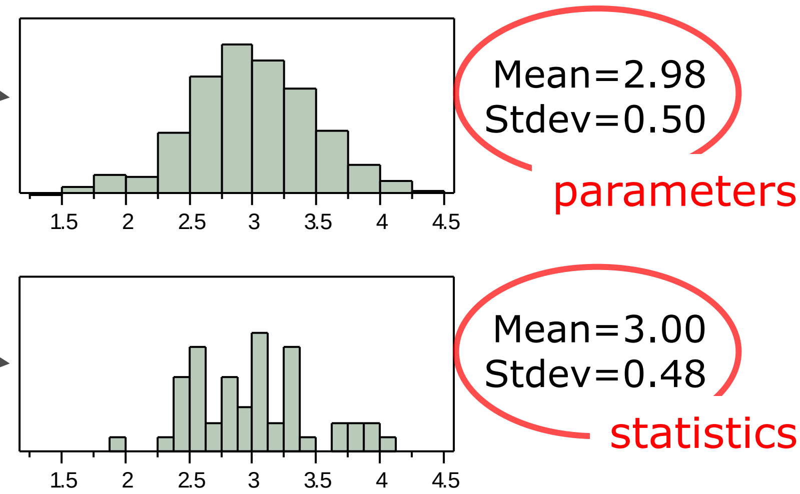
True (“population”) value
= parameter

Sample value
= statistic

- Use a sample ***statistic*** to estimate a population ***parameter***

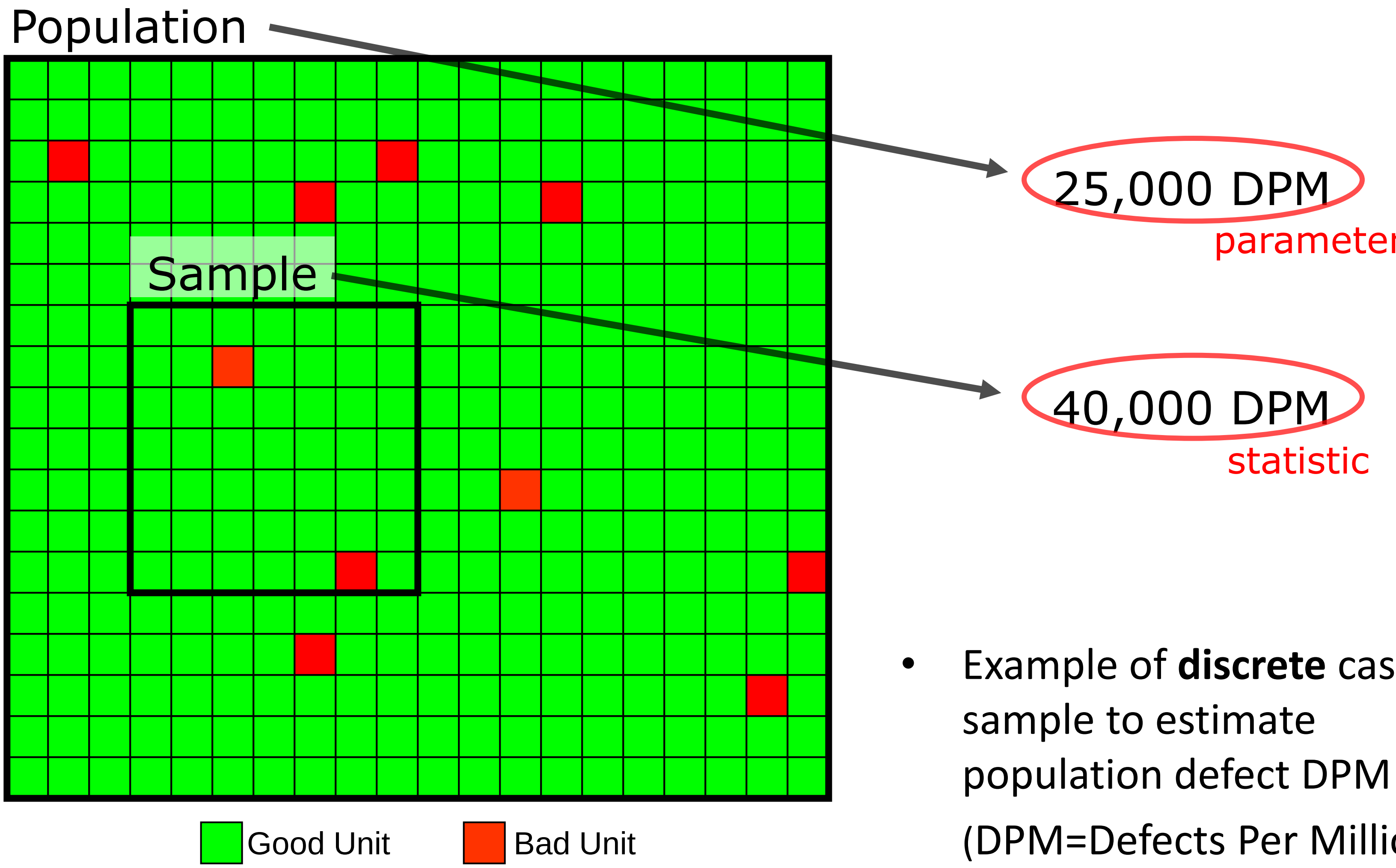
Statistical Inference (Continuous)





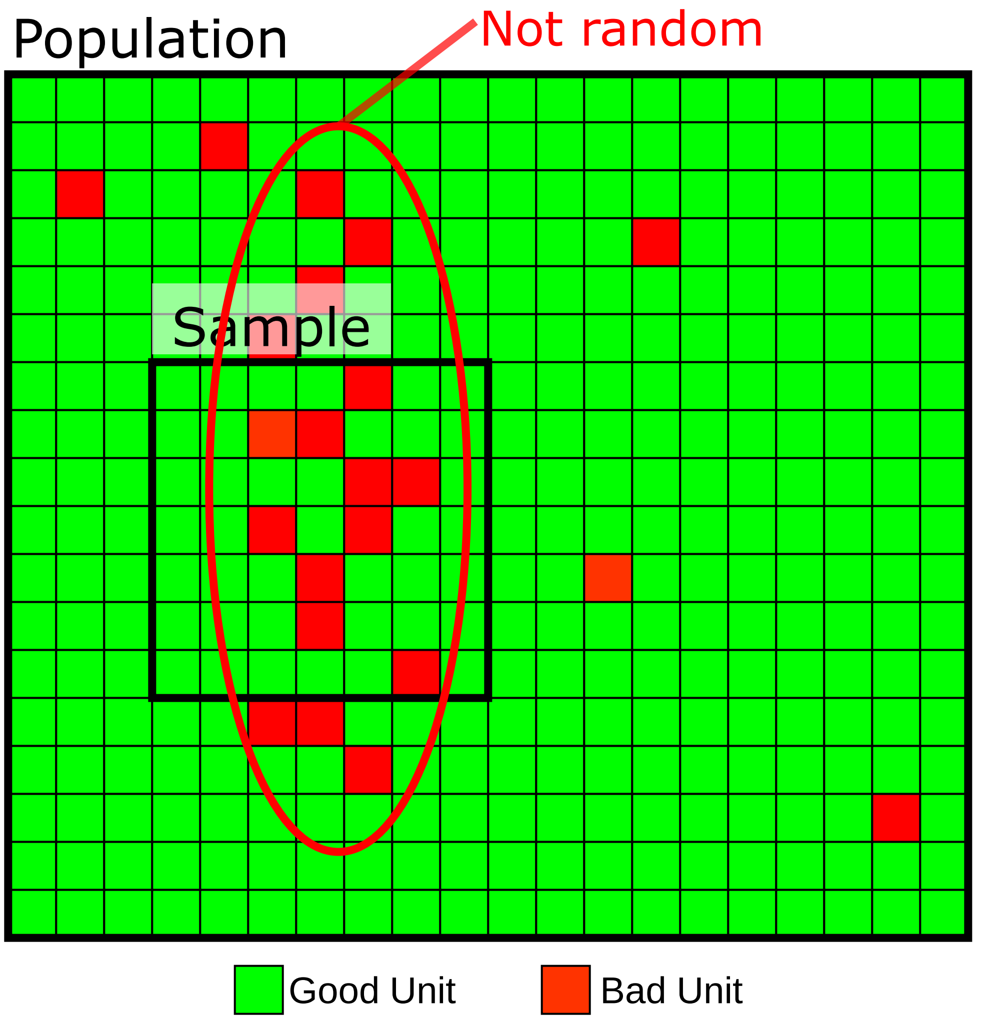
- Example of **continuous** case: Use sample to estimate population mean and standard deviation

Statistical Inference (Discrete)



- Example of **discrete** case: Use sample to estimate population defect DPM (DPM=Defects Per Million)

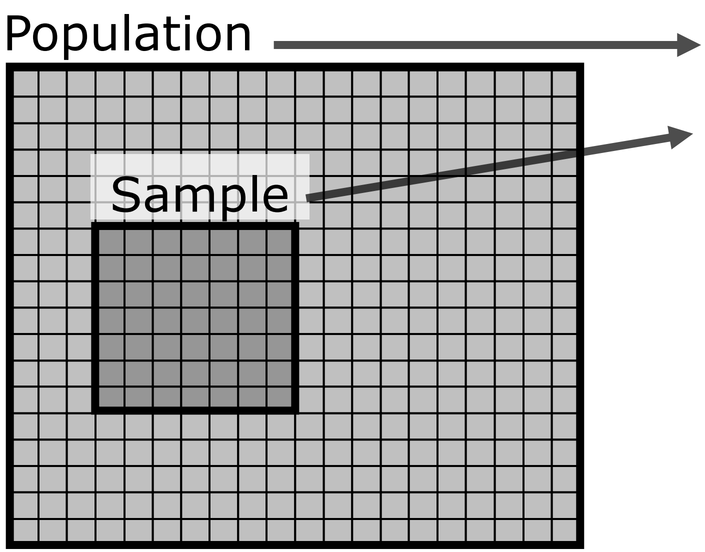
Note: Samples Must Be Random!



Population = 55,000 DPM
Sample = 204,000 DPM

- Samples must be representative of the entire population!
- Best to select samples truly randomly
  - Not the first lot available or other partly-random methods
- No statistical analysis can correct for non-random samples

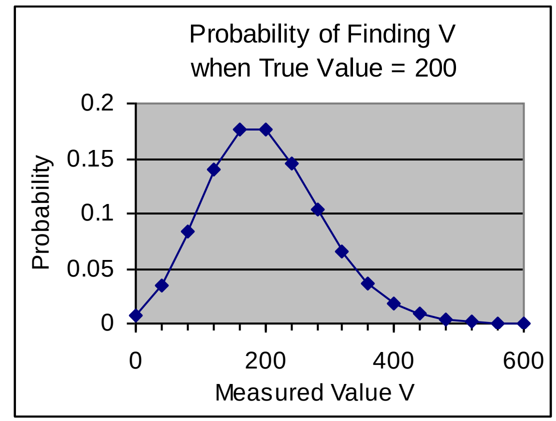
Distributions of Statistics



True ("population") value = parameter

Sample value = statistic

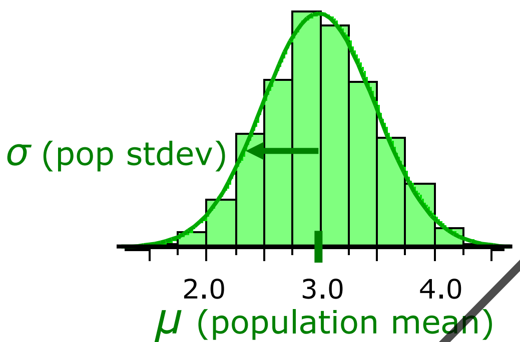
Distribution of statistic:

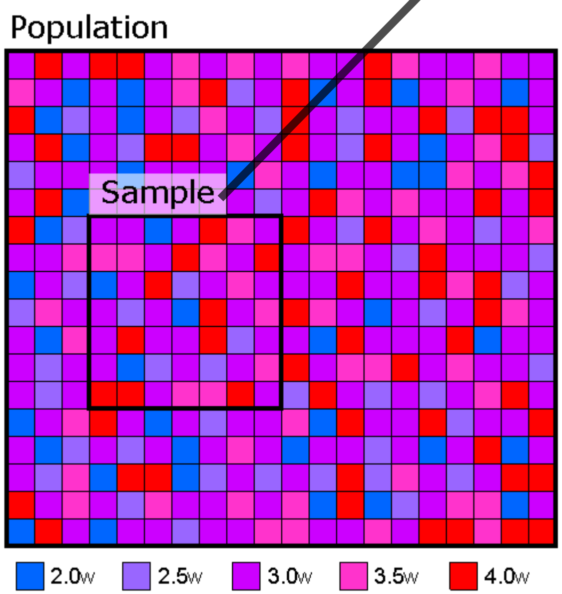


- Measured statistic is not enough
- Need to add either
  - Confidence interval or limits
  - Answer to a statistically-well-posed question ("hypothesis test")
- Calculated from distributions of statistics
  - If we looked at many samples from many identical populations, what values of the statistics might we get?

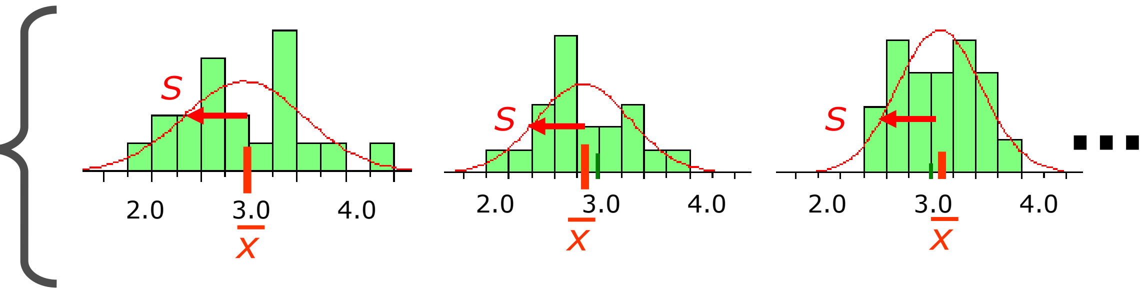
Distributions of Statistics (Continuous)

Population has one true distribution:

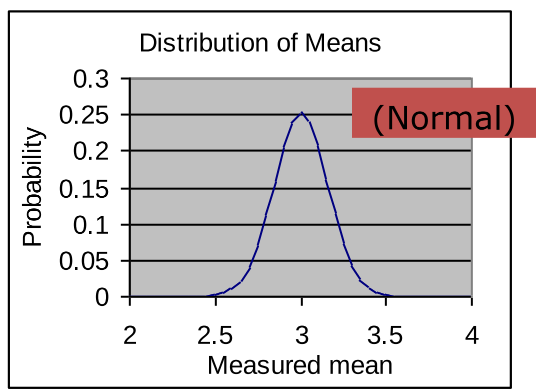


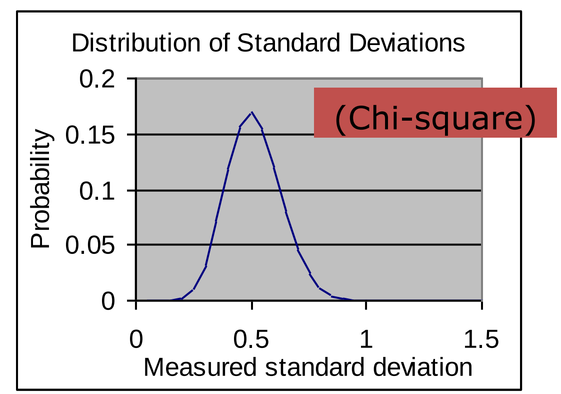


Different samples have different distributions:



Properties of sample distributions are *statistics*. We can calculate distributions of these statistics:





We get one value for each from our one sample.

Distributions of Statistics (Discrete)

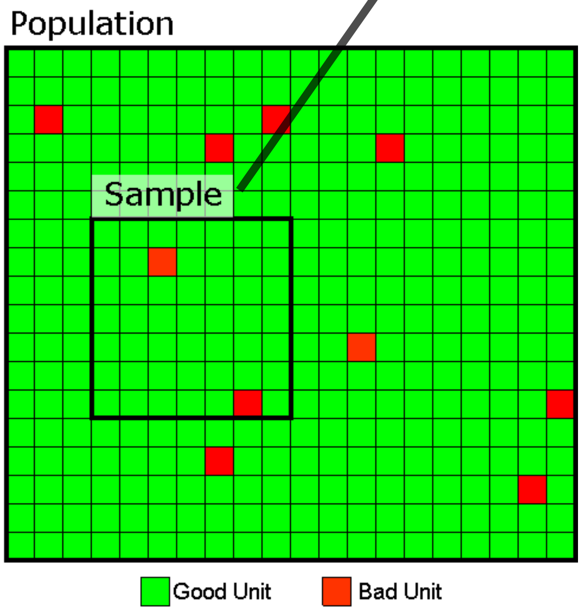
Population has one true DPM:

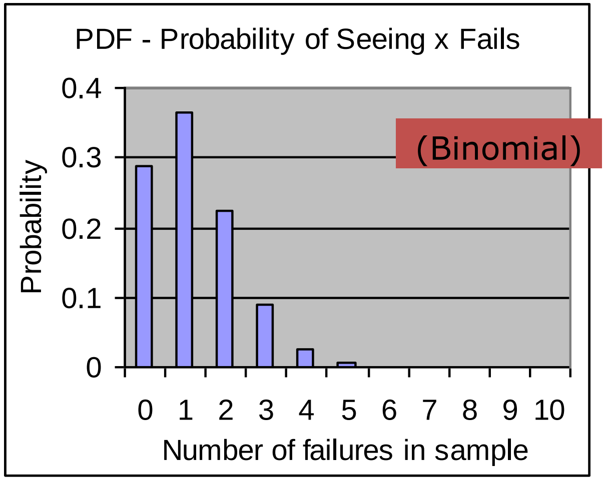
25,000 DPM

Different samples have different DPMs:

20,000 DPM (1 fail) 0 DPM (0 fail)
40,000 DPM (2 fail) 60,000 DPM (3 fail) ...
20,000 DPM (1 fail) 40,000 DPM (2 fail)

The measured sample DPM is a *statistic*.
We can calculate the distribution of this statistic:





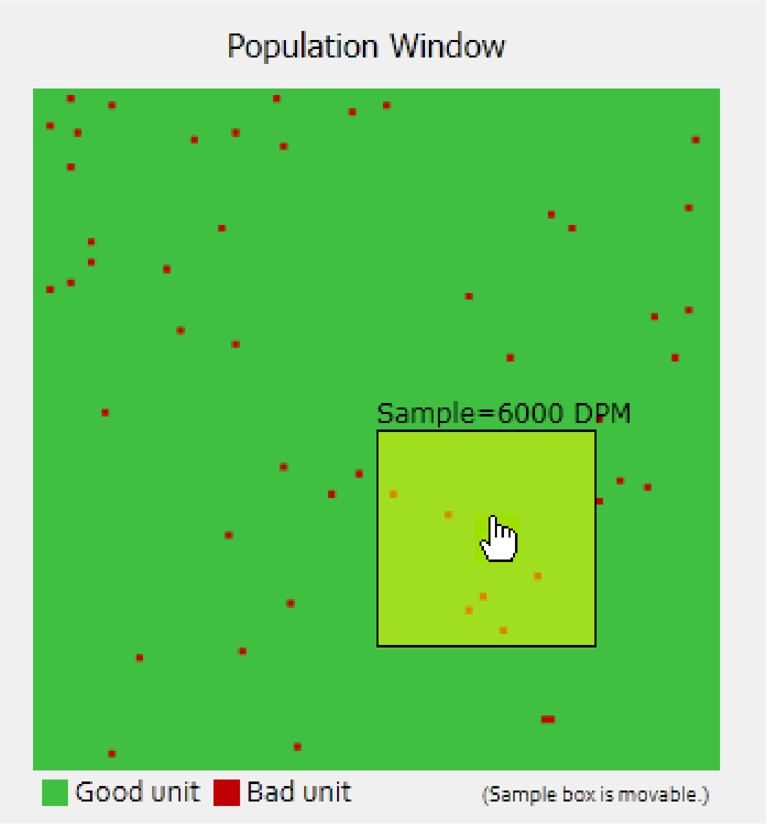
We get one value from our one sample.

DPM Simulation

Population Window

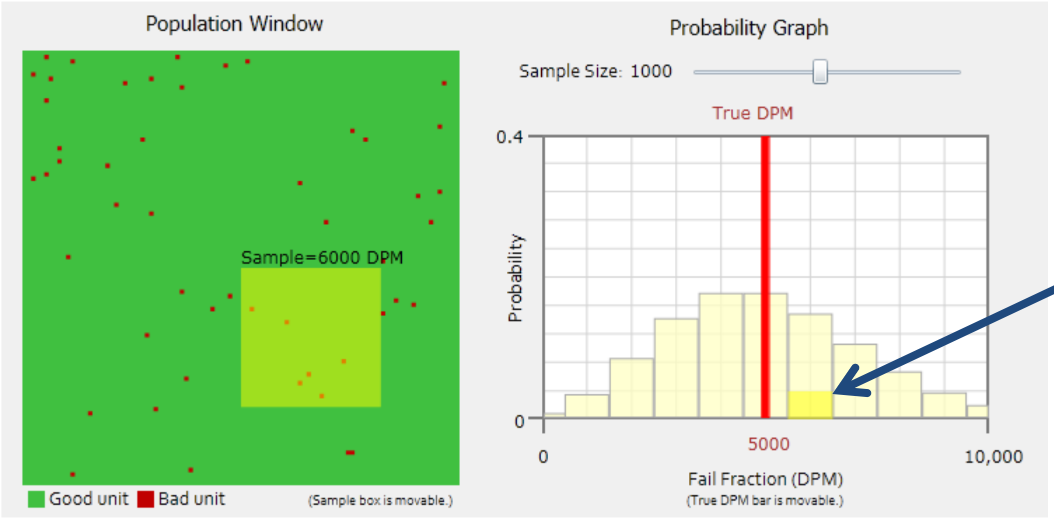
- Shows 10,000 units, most good, a few bad

The Sample

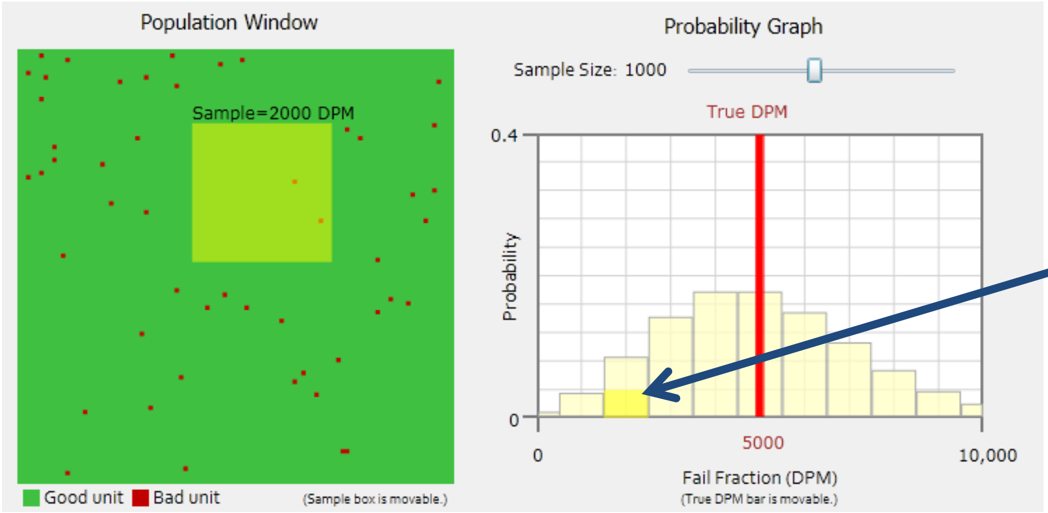


- You can move the sample box

DPM Indicator on DPM Histogram

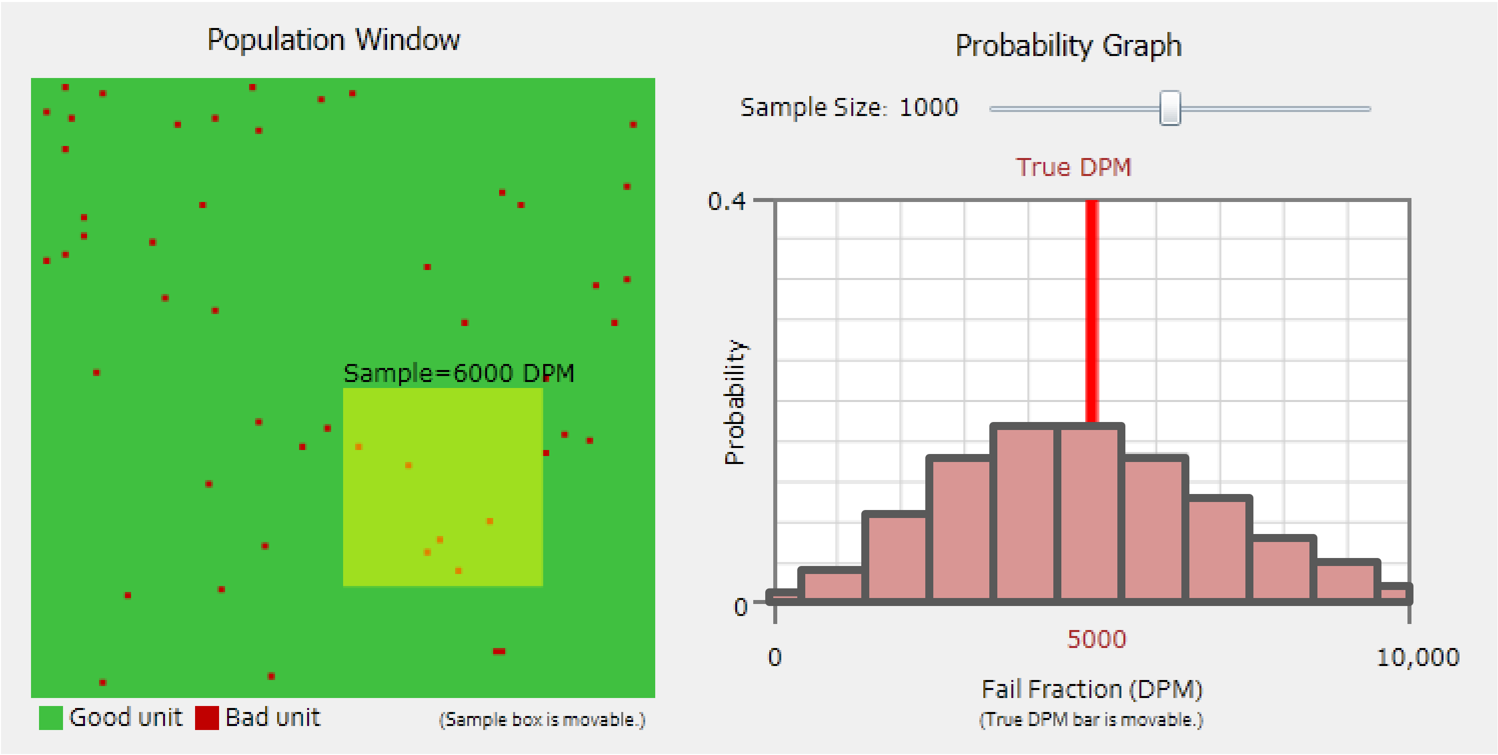


6000 DPM



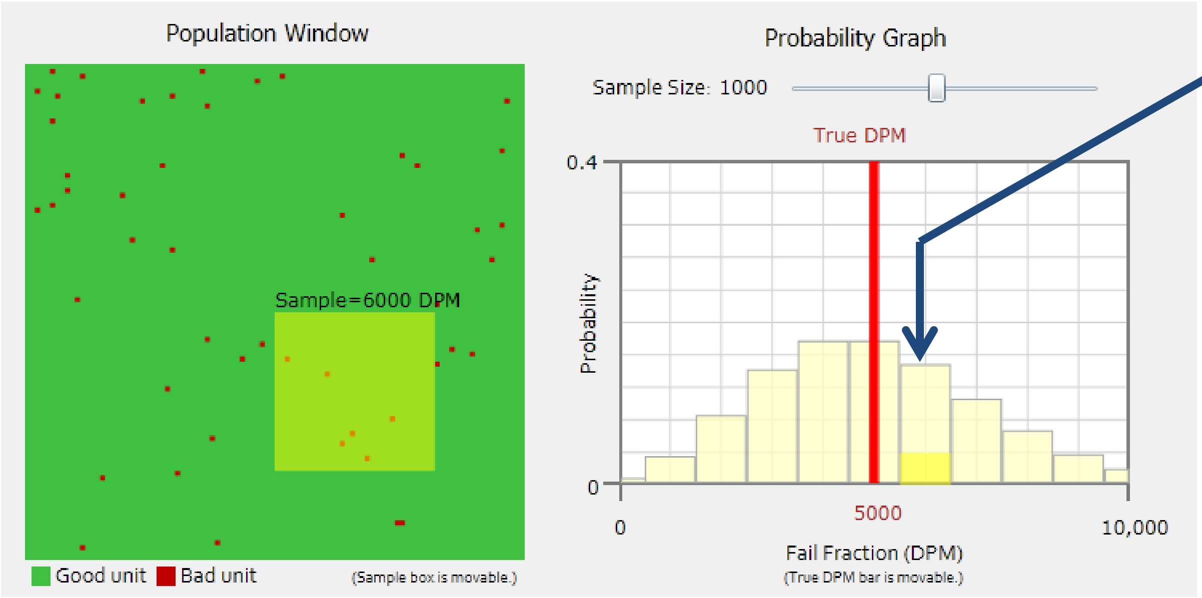
2000 DPM

Binomial Histogram



- Gives probability of getting each measurement given the true DPM

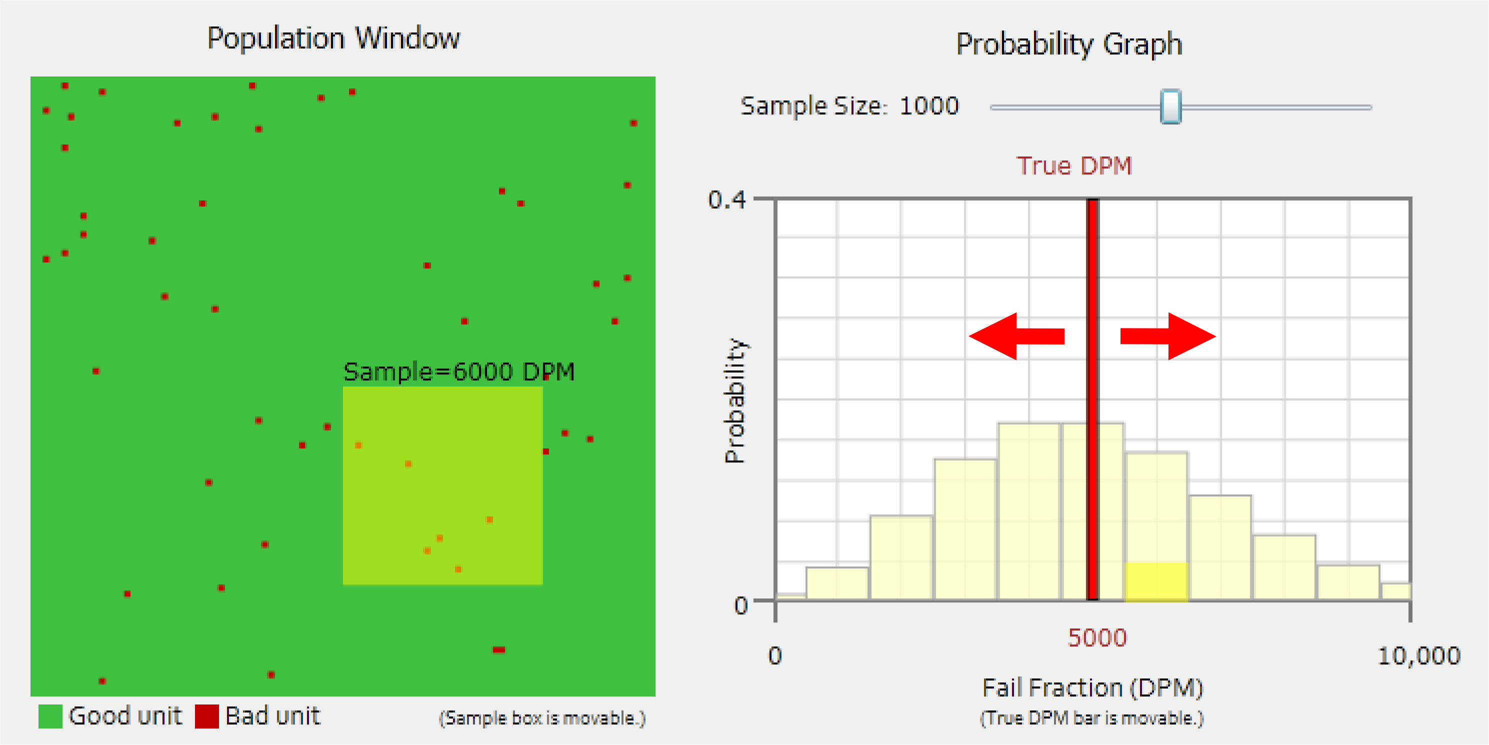
Binomial Distribution



=binomdist
(6, 1000, 0.005, false)
6 fails
1000 samples
5000 DPM
Not cumulative

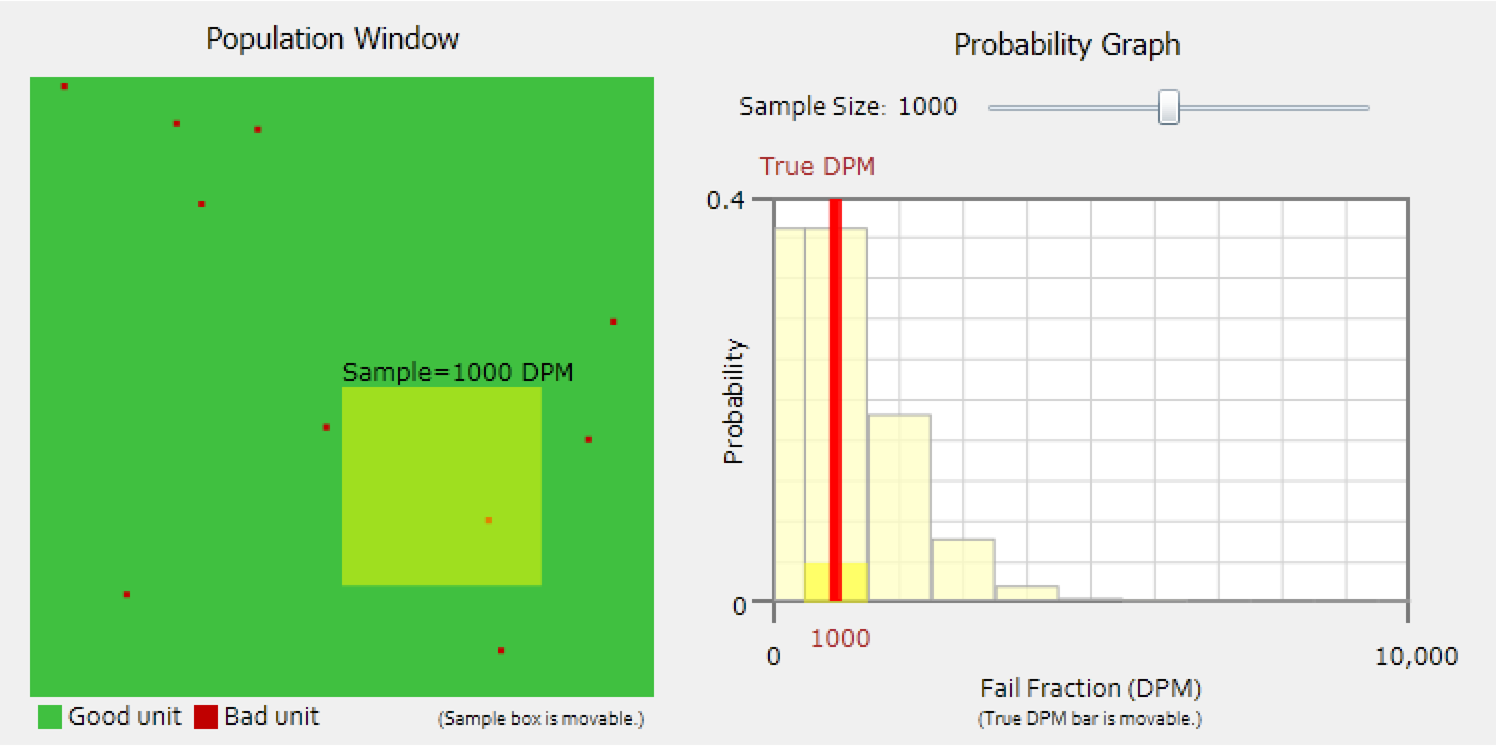
$$\text{binomdist}\left(f, N, p, \text{false}\right) = \binom{N}{f} p^{f} \left(1-p\right)^{N-f}$$

True DPM



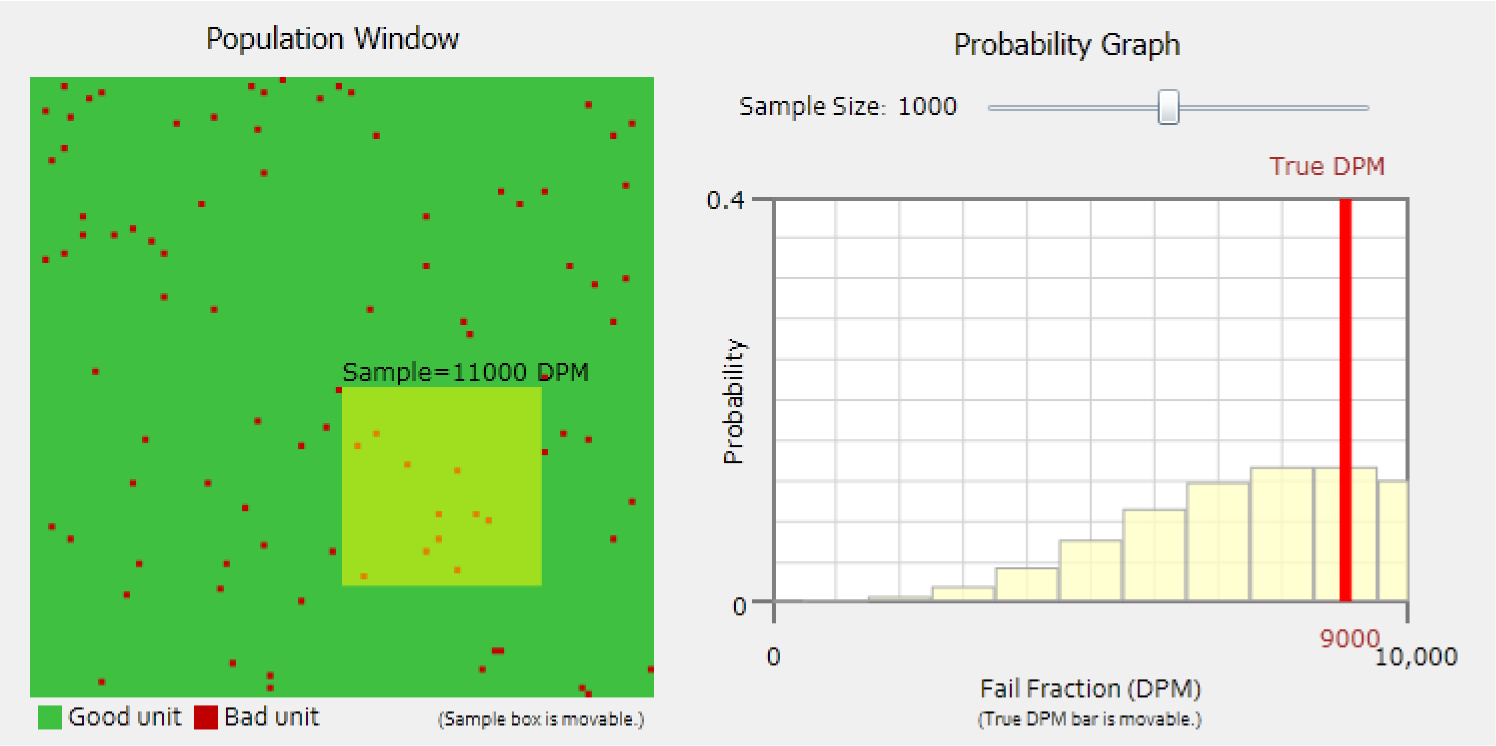
- True DPM is adjustable
  - Not in the real world, only the simulation!

Low DPM

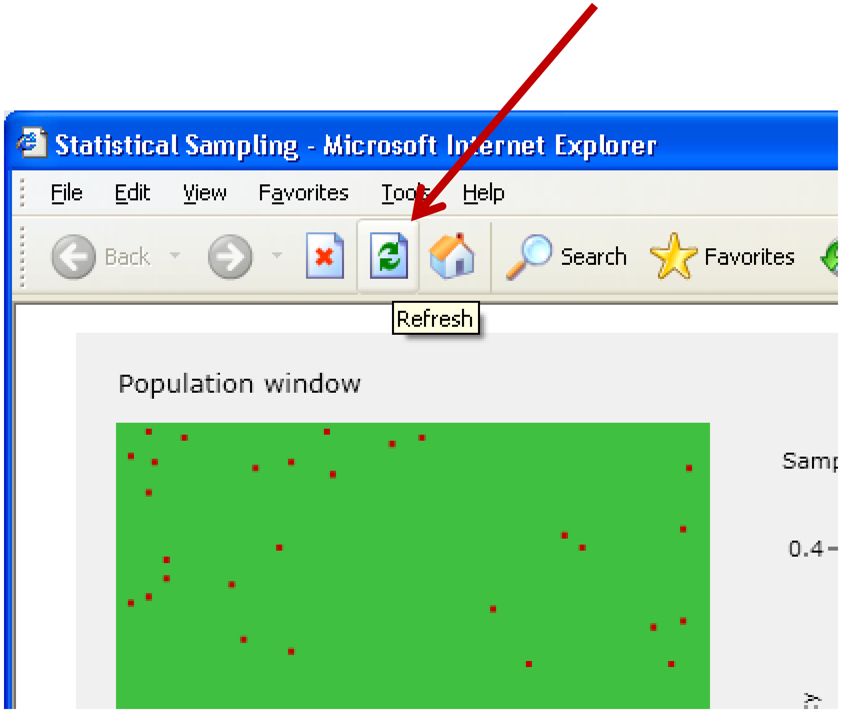


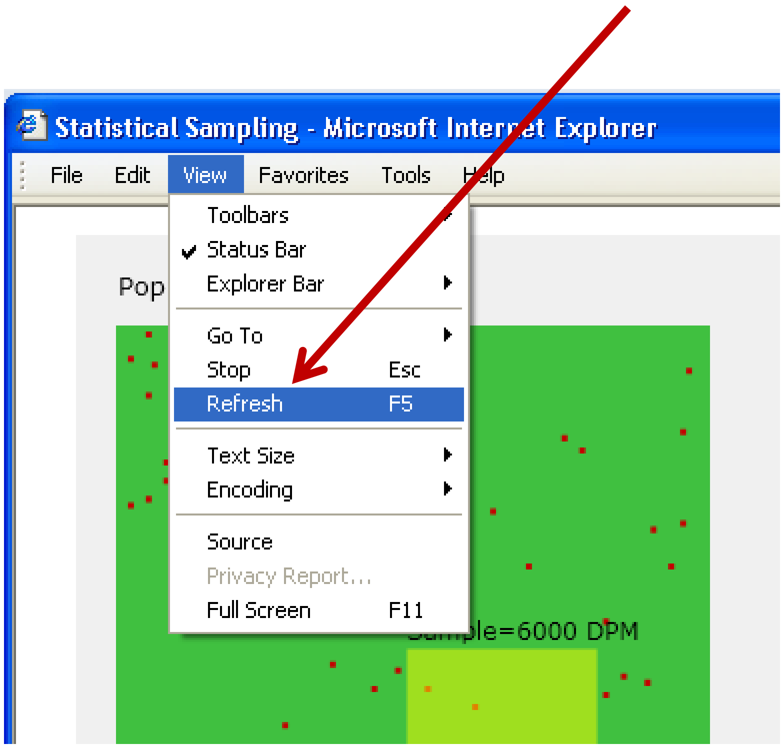
Population Window
Sample=1000 DPM
Good unit
Bad unit
(Sample box is movable.)
Probability Graph
Sample Size: 1000
True DPM
0.4
0
Probability
0
1000
10,000
Fail Fraction (DPM)
(True DPM bar is movable.)

High DPM

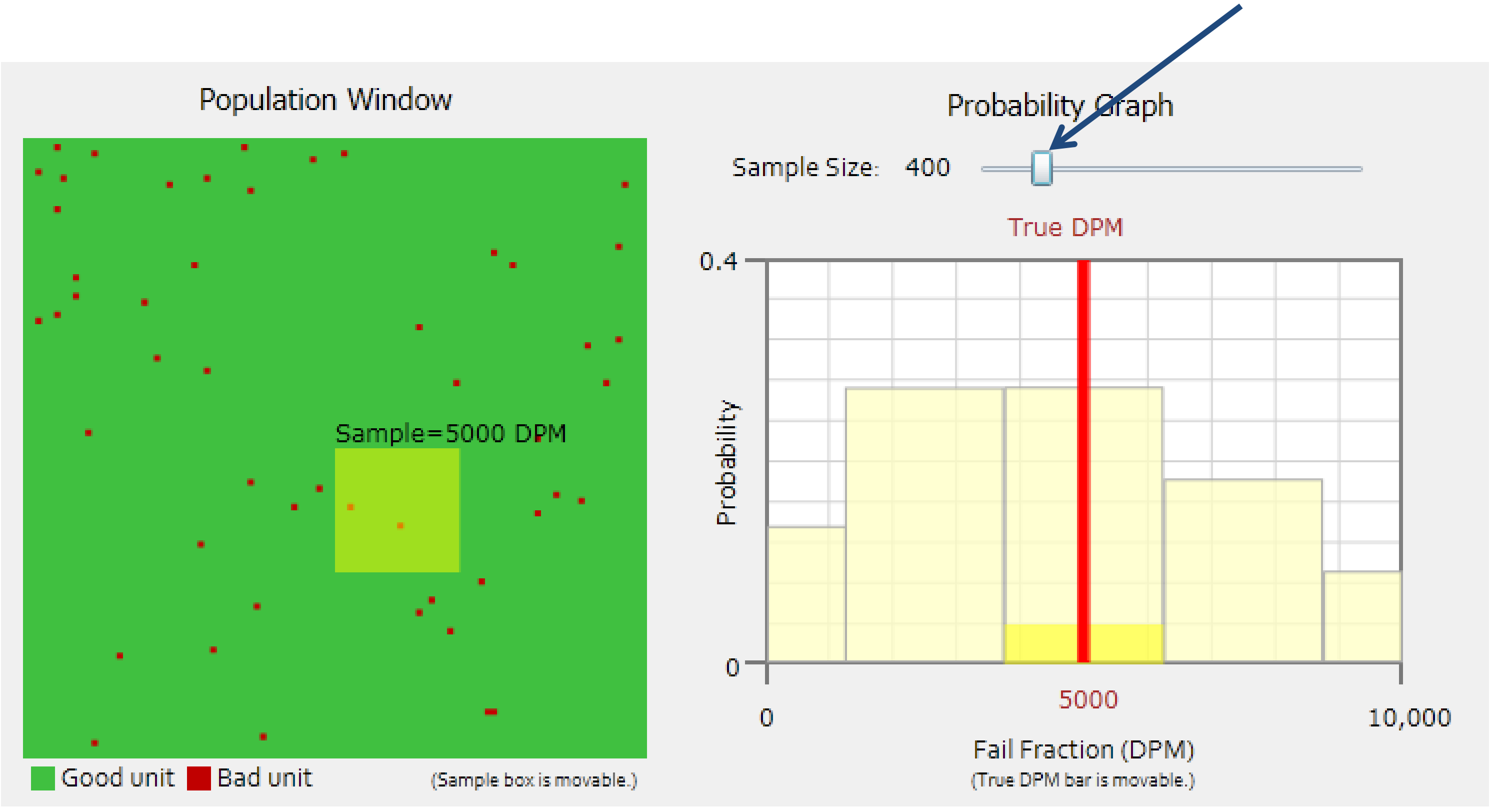

Population Window
Sample=11000 DPM
Good unit
Bad unit
(Sample box is movable.)
Probability Graph
Sample Size: 1000
True DPM
0.4
0
Probability
0
9000
10,000
Fail Fraction (DPM)
(True DPM bar is movable.)

Please put True DPM back to 5,000

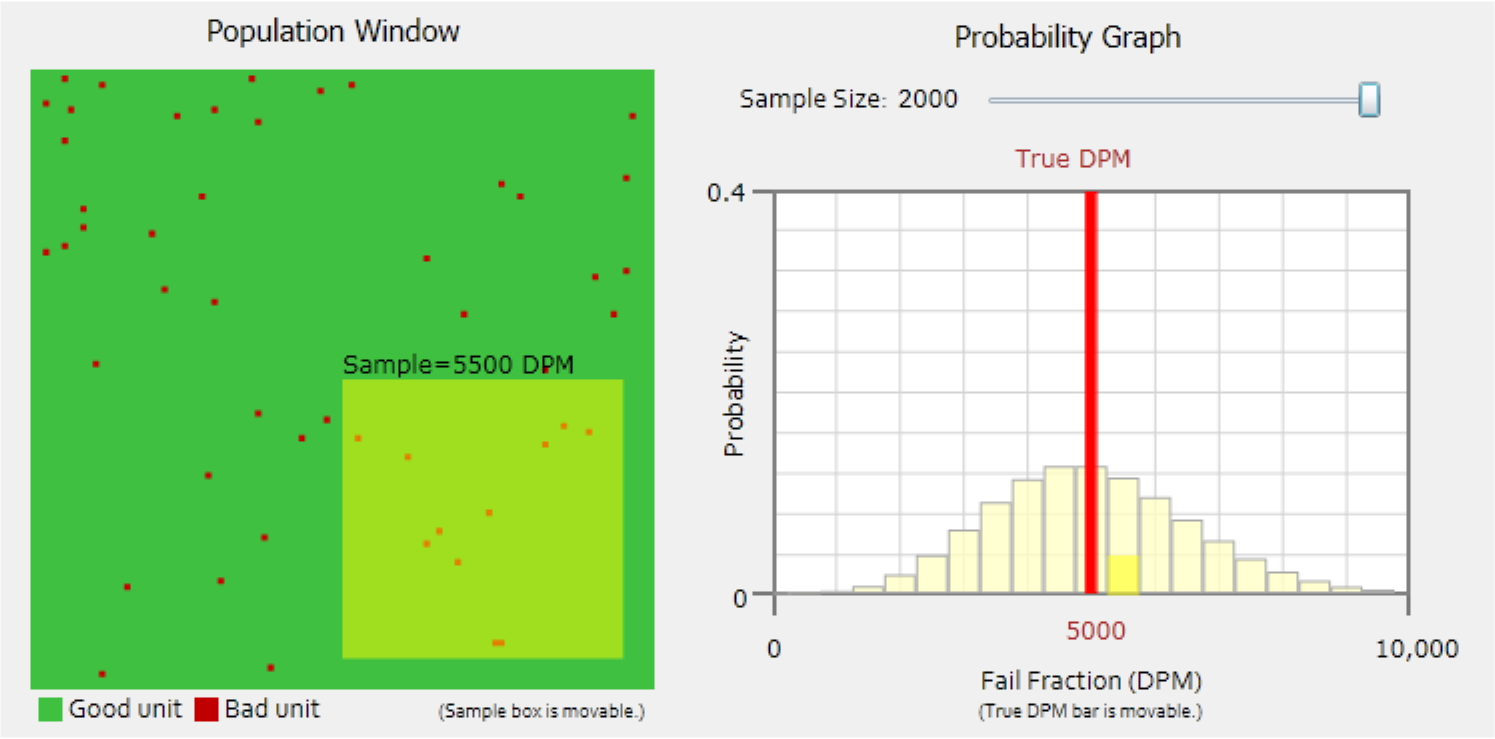




Small Sample Size

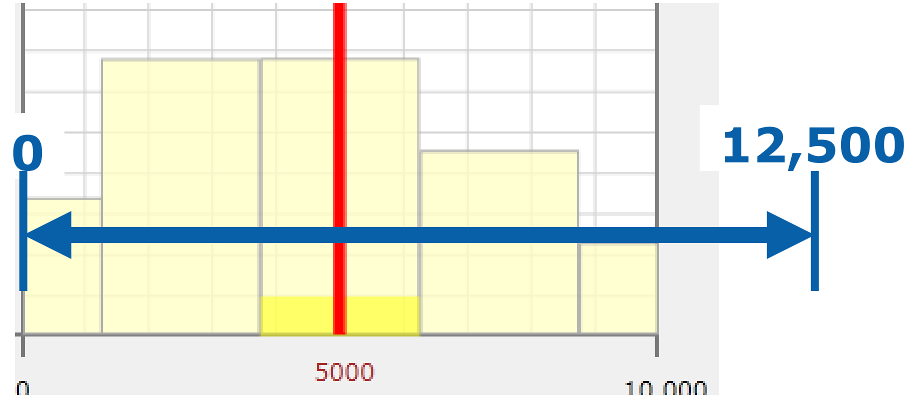


Large Sample Size

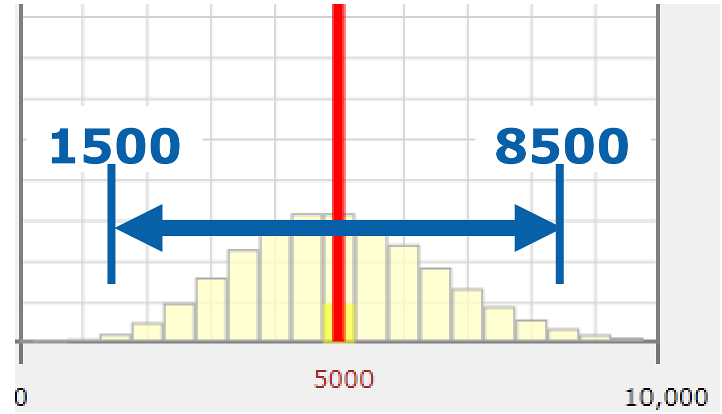


Statistical Measurement Uncertainty

Small sample (400)
= wide range



Large sample (2000)
= narrow range



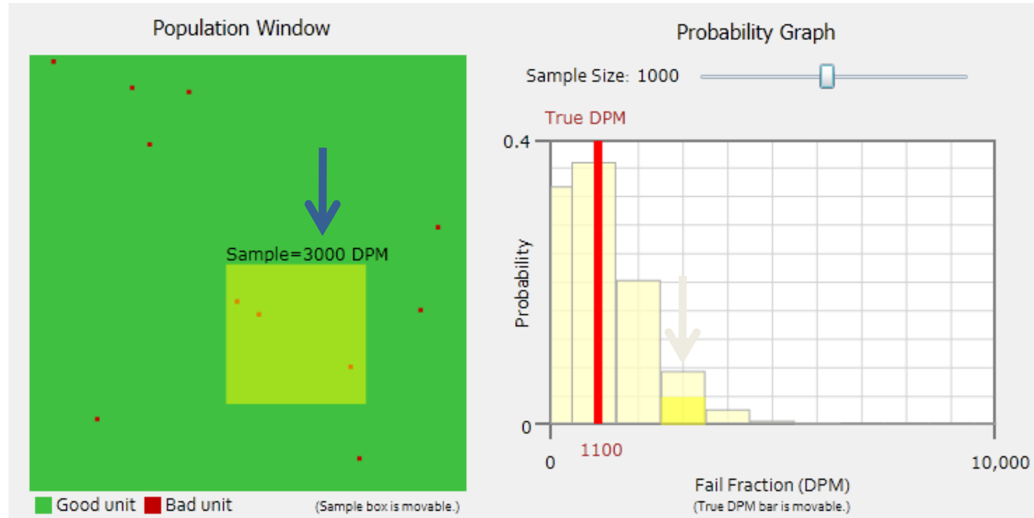
Exercise 6.1

- (A) Set sample size = 1000
- (B) Set True DPM = 1100 DPM and look for a sample with 3 fails – what DPM does that represent?
- (C) Set True DPM = 6700 DPM and look for a sample with 3 fails – what DPM does that represent?

Why We Need Confidence Limits

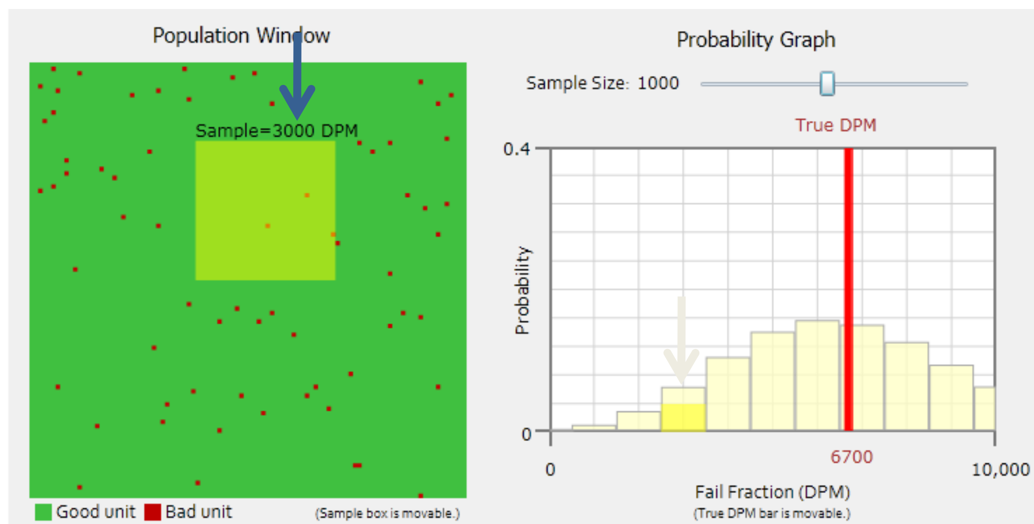
Did you get…

a *bad* sample from
a *good* population?



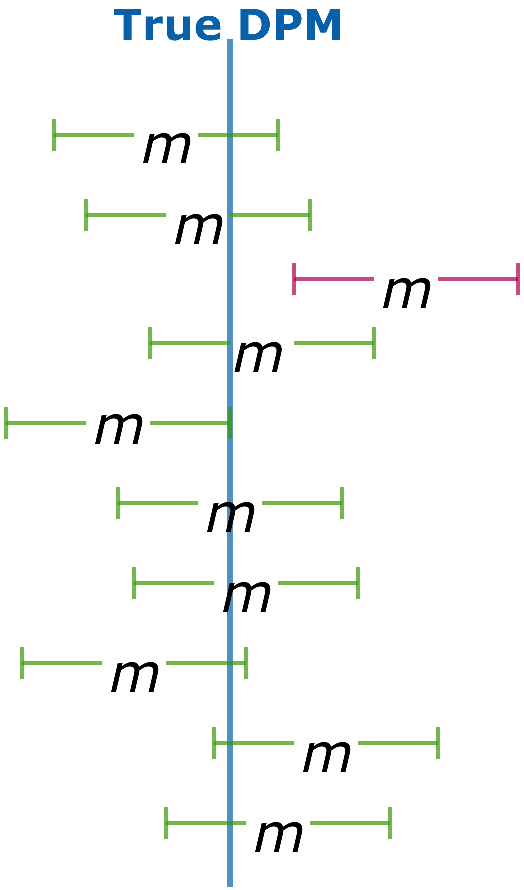
…or…

a *good* sample from
a *bad* population?



Confidence Limits

Confidence Interval Meaning



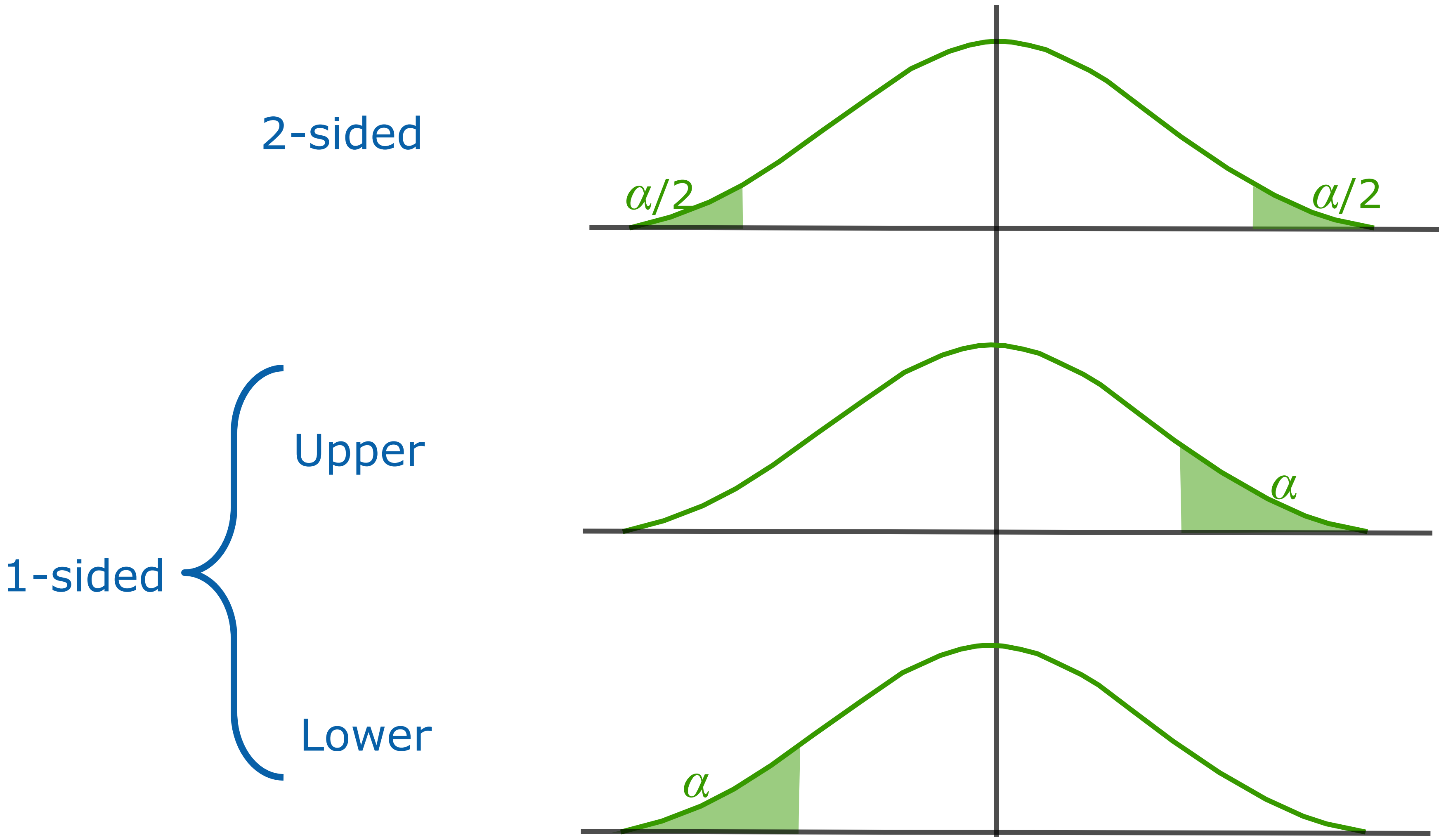
Confidence level = 90% = 0.9

Risk of being wrong
= 1 − confidence level

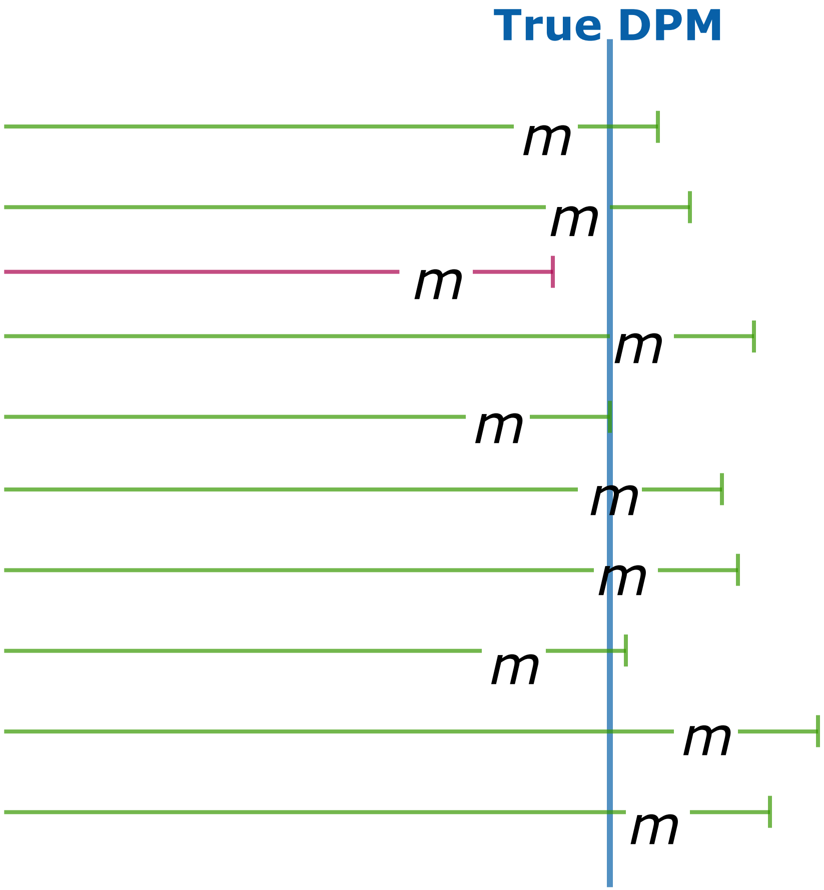
= α = 10% = 0.1

- 90% of random sample means with this confidence interval include the true population mean

1-Sided vs. 2-Sided



1-Sided UCL Meaning



- 90% of random sample means with this confidence interval include the true population mean

Calculating Confidence Limits

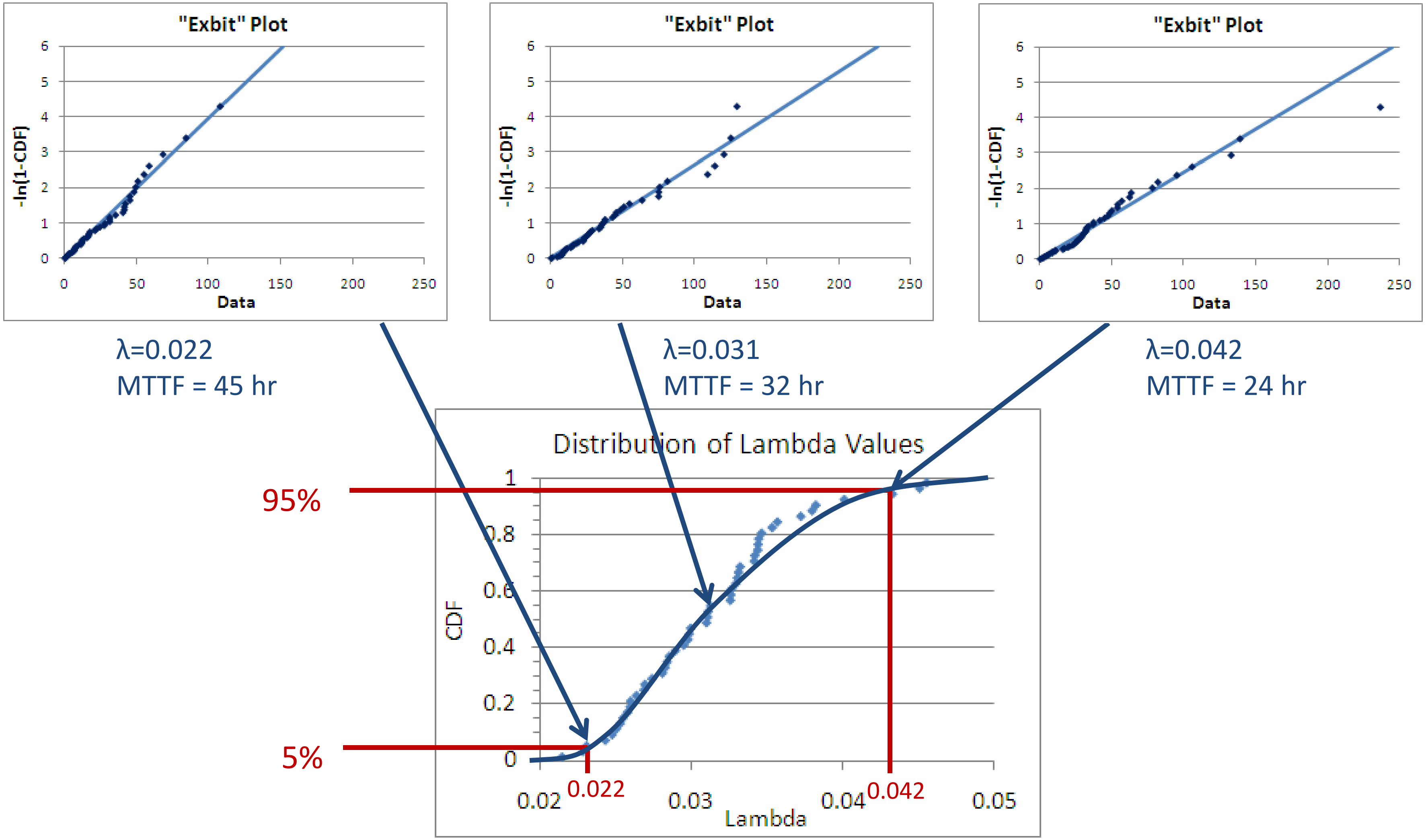
Exercise 6.2a

Monte Carlo determination of binomial CL

Exercise 6.2b

Compare Monte Carlo CL to analytic

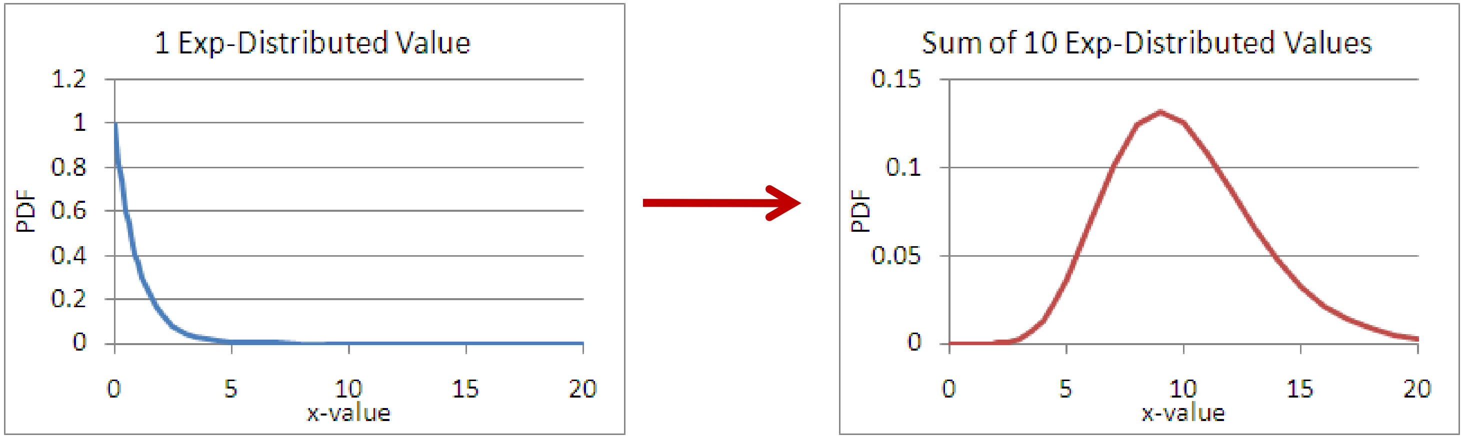
Monte Carlo Exponential CL



Exercise 6.3a

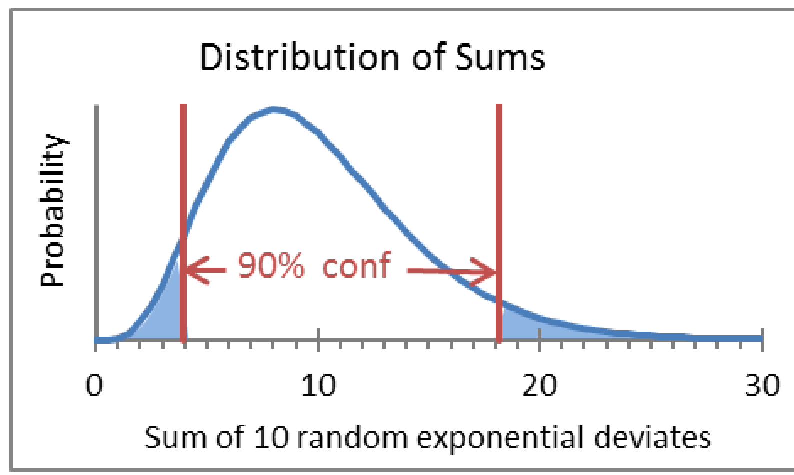
Monte Carlo determination of exponential CL

Analytic Exponential CL



- For $f(t) = \lambda\, e^{-\lambda t}$, best estimate for $1/\lambda$ is $\frac{1}{\mathrm{N}}\sum t_i$ where t_i are the data
- So, what is the *distribution* of $\sum t_i$ where t_i are distributed exponentially?

- Answer: a gamma or a chi-square distribution
- Confidence intervals taken from that



Exercise 6.3b

Compare analytic to Monte Carlo values

The End