Sample Size

Kim Iles

Sampling specialist

Nanaimo, BC, Canada

What do you want?

- You <u>want</u> to be <u>comfortable</u> with the "precision" (error %) of the result.

- You <u>want</u> to be <u>confident</u> enough to make decisions (if there are consequences … otherwise who cares?).

You do <u>not</u> want to put in plots after they no longer do you any good.

You do <u>not</u> want to waste time, money, or staff. (<u>or</u> to see them idle)

- PS – there is no minimum sample size

- Well, 1, I suppose

- 2 if you want to do statistics

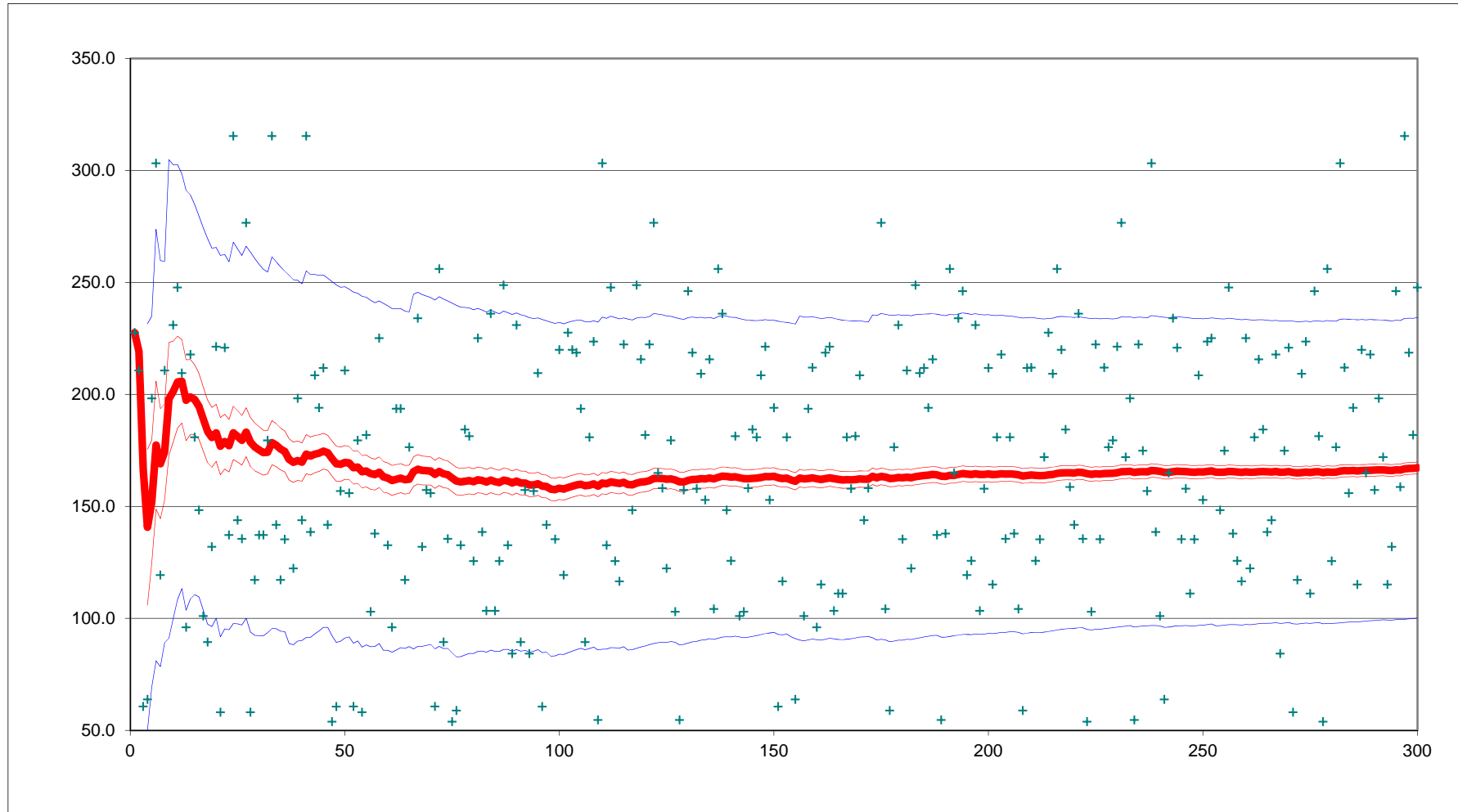- … but nobody *wants* to do statistics.

How does <u>Kim</u> decide?

What is traditional? Uggg
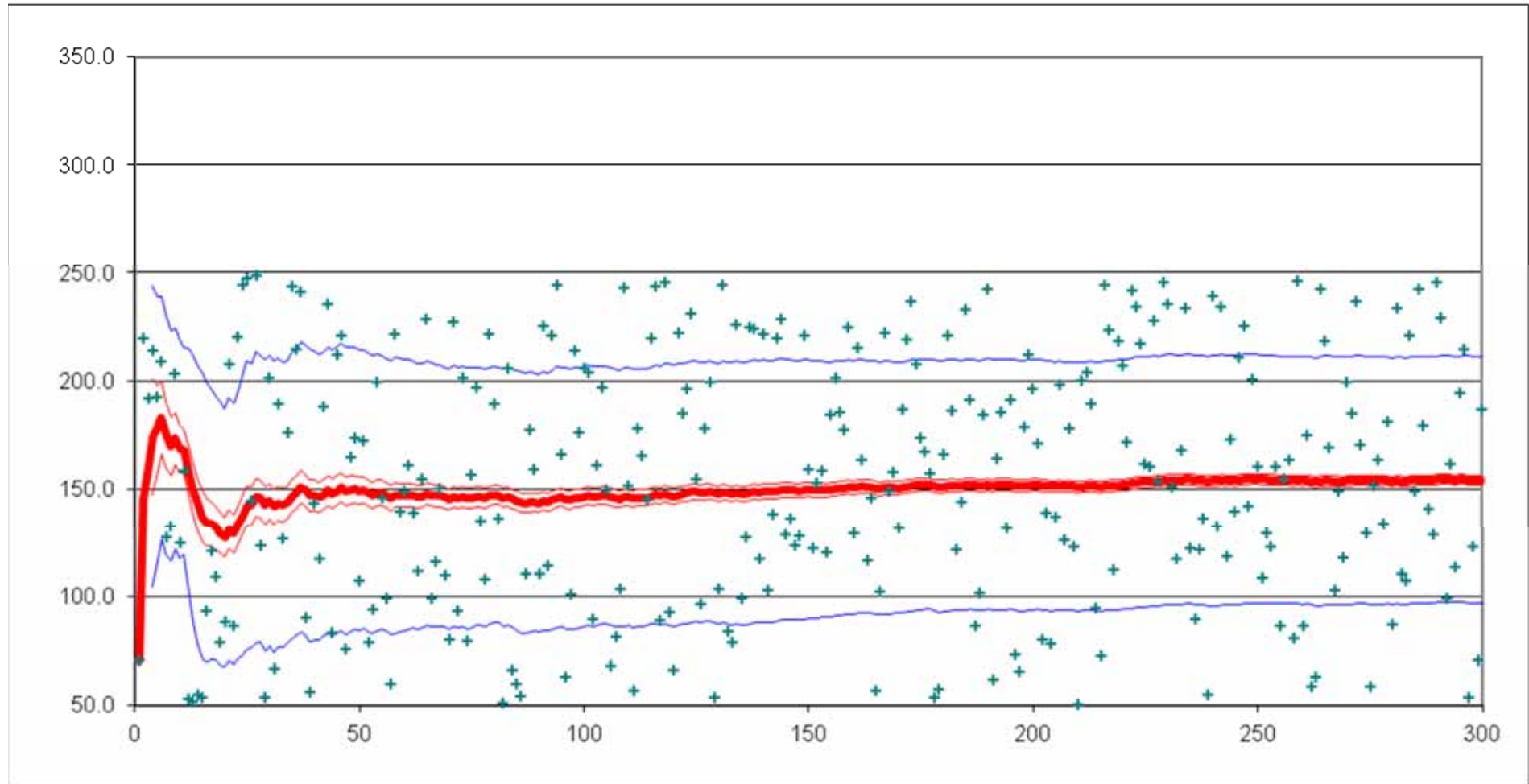
What is "comfortable"? hmmmm

When is the answer "stable enough" <u>to make the decision</u>?

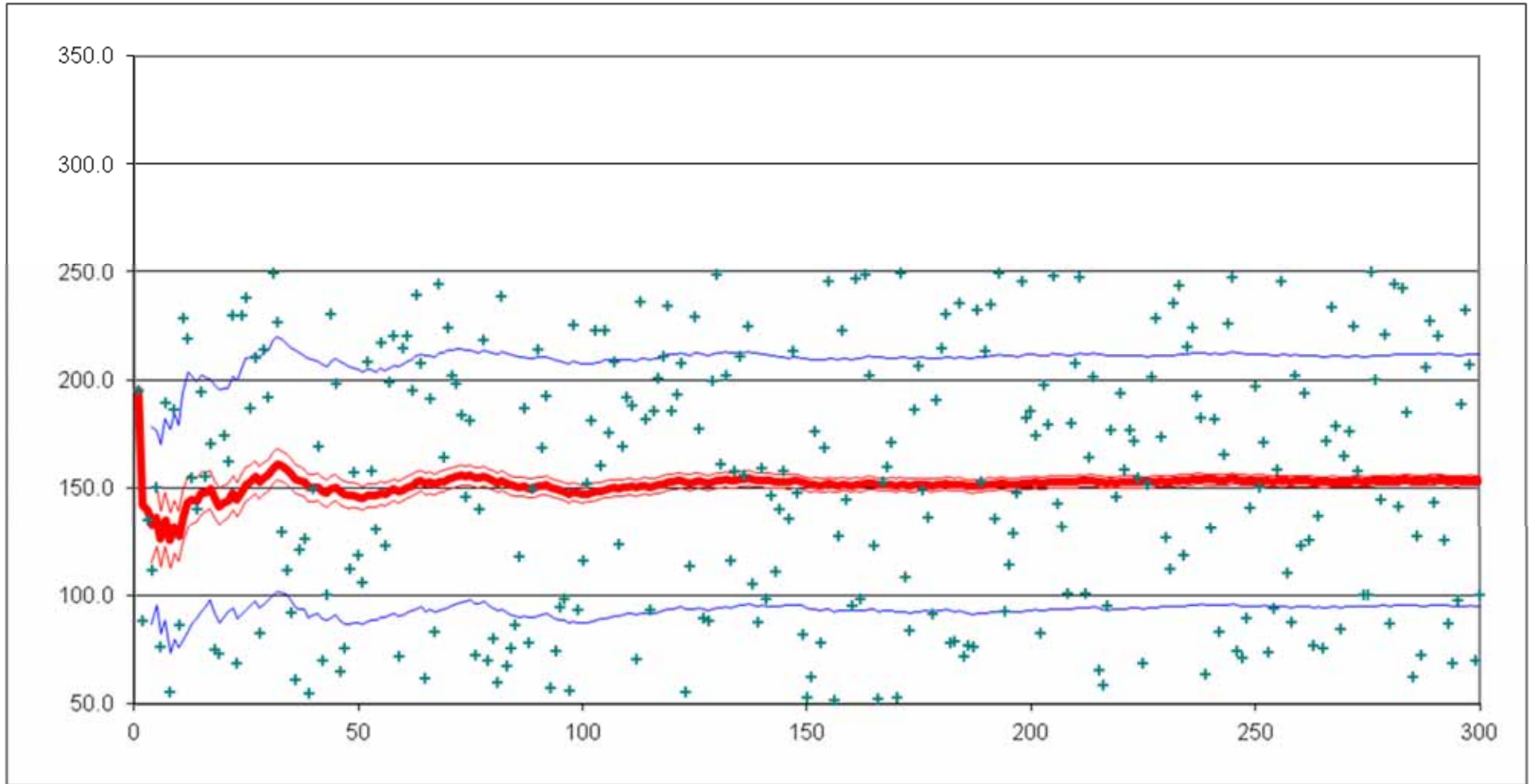<u>pretty good</u> criteria …

When does the answer stabilize?

2nd try, with real data

3d try

A typical (bad) question …

What sampling error do you "**NEED**"?

Nobody ever knows.

… EVER …

It might be in legislation, or a manual, but that is NOT what you "need".

The RIGHT question

- Here is the sampling error we <u>got.</u>
- Here is the sampling error we <u>will</u> get.
- Here is the sampling error we can <u>afford</u> to get.

- **CAN YOU <u>LIVE</u> WITH THAT ???**

What is the "**Probable** Error"

- It's like the average.
- "Here is a error, chances are about equal that the <u>actual</u> error is above or below that number".
- Statistics used that for centuries.
-     so simple, really ….

Enter the "95% Sampling Error"

- This is the sampling error that we do <u>NOT</u> have.
- It's <u>less</u> than that 95% of the time.
- The full description is …
- "This is the Sampling Error that we are <u>very sure</u> (95%) that we do <u>NOT</u> have – it's less than that".

Example of Actual Errors
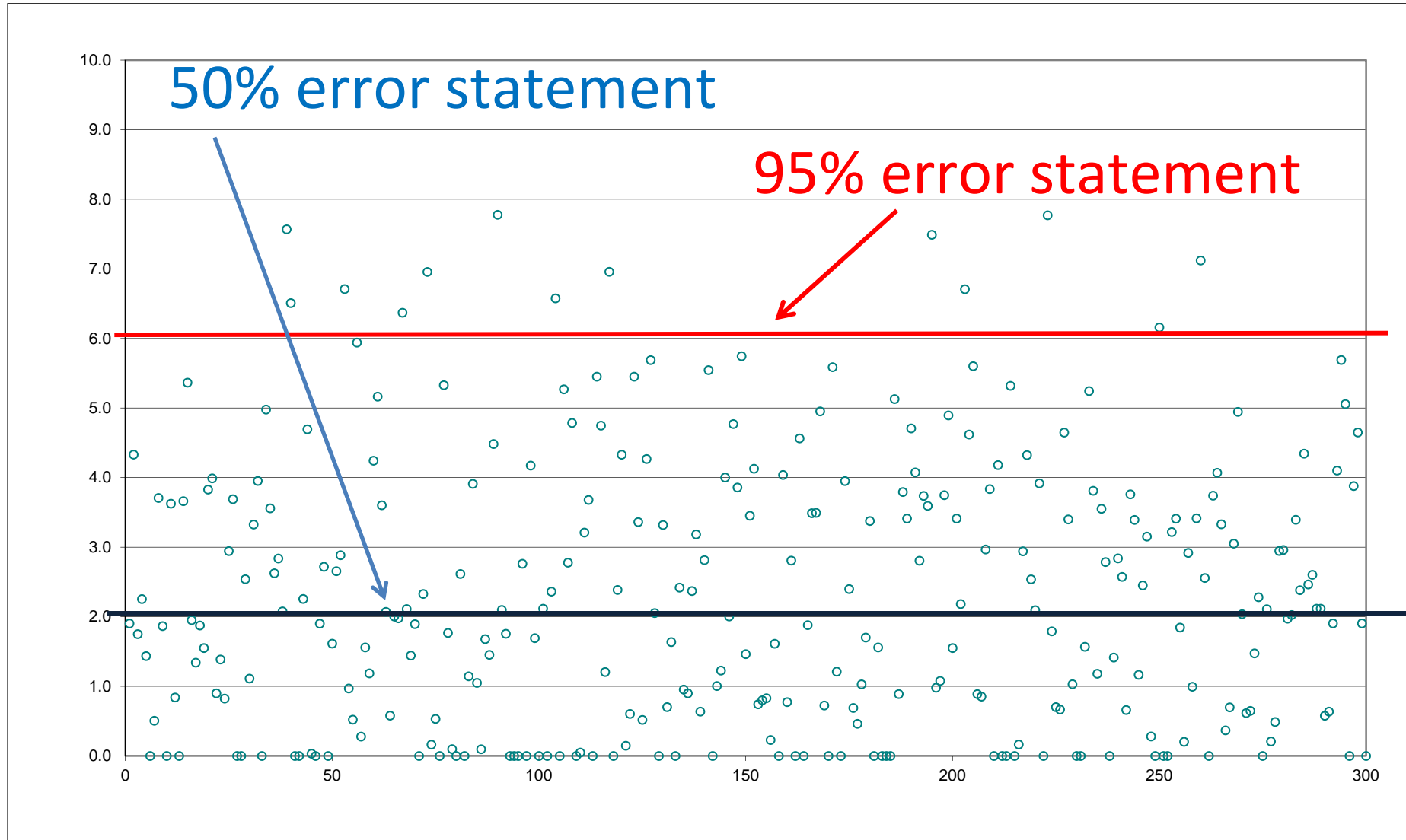
If you <u>must</u> use an equation

$$\sqrt{\frac{CV}{\sqrt{samplesize}}}$$ * .7 to get Probable error

$$\sqrt{\frac{CV}{\sqrt{samplesize}}}$$ * ~2 to get 95%

"Improbable" error

Problem $\longrightarrow$ Result

$$\frac{20\%}{\sqrt{25}} = 4\%$$
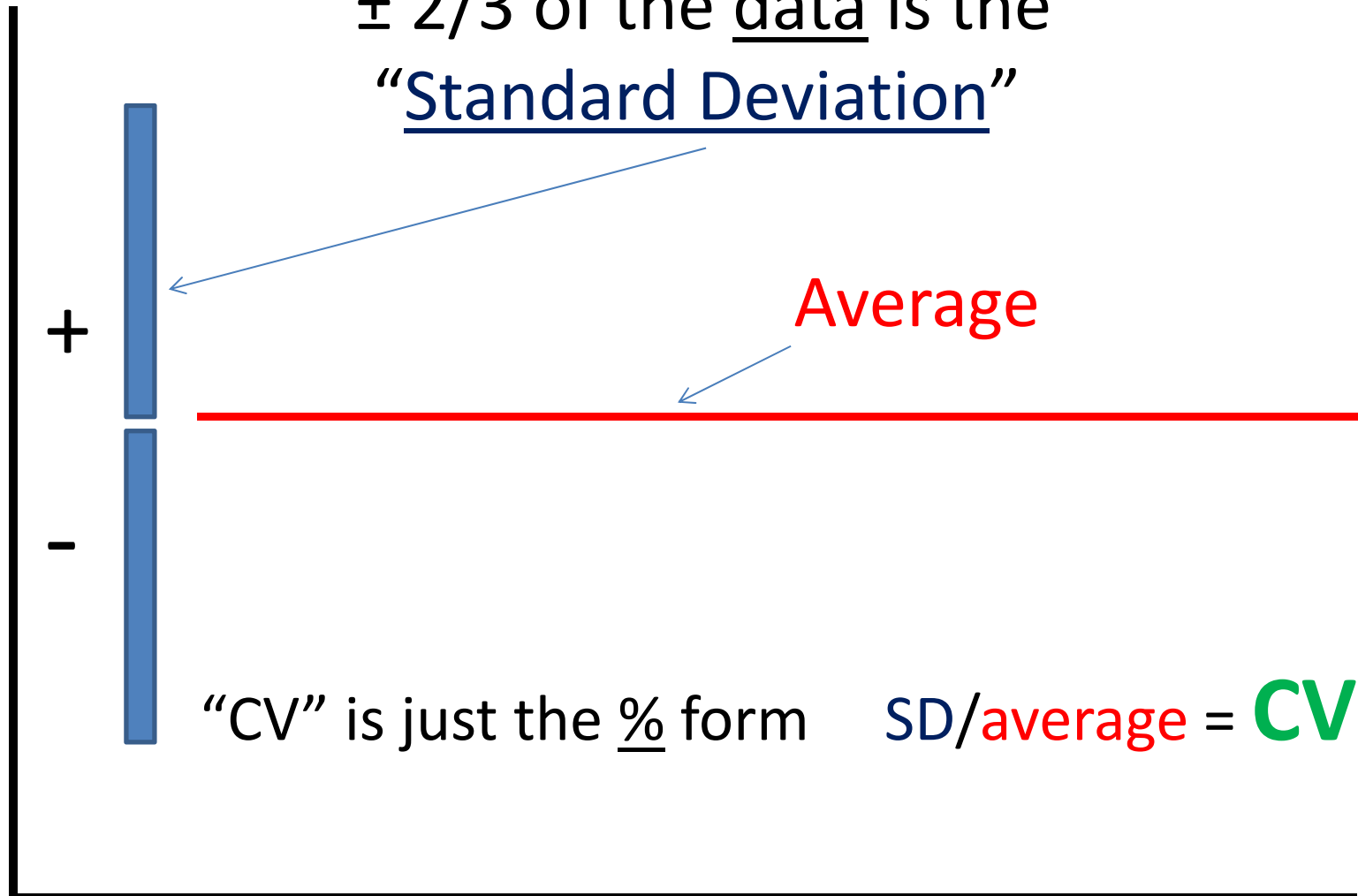
Effort $\longrightarrow$

* .7 = SE% of **2.8%** (50% level)

* 2 = SE% of **8%** (95% level)

What are the primary "terms" is stats?
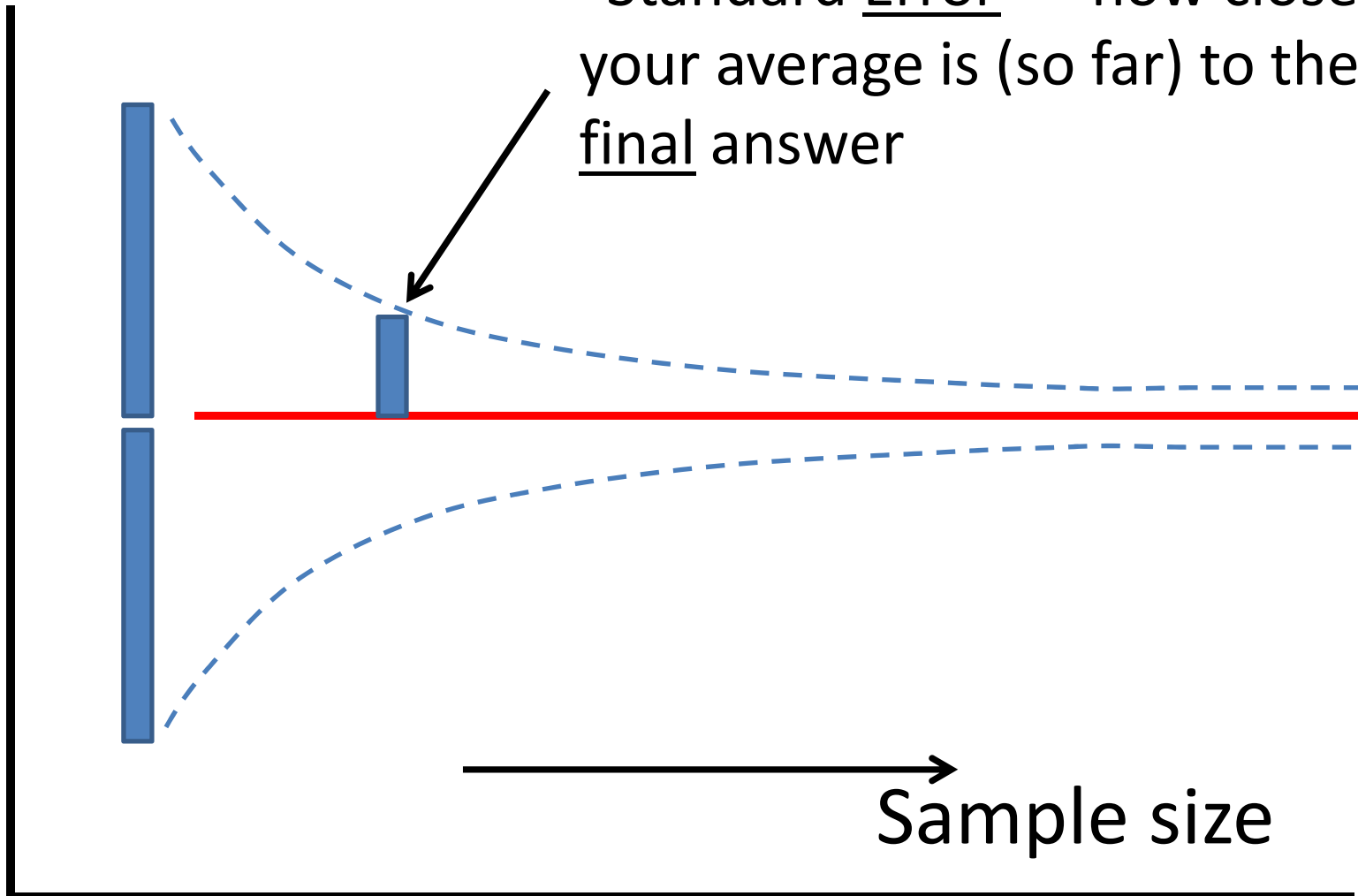
Only 2, really

± 2/3 of the data is the
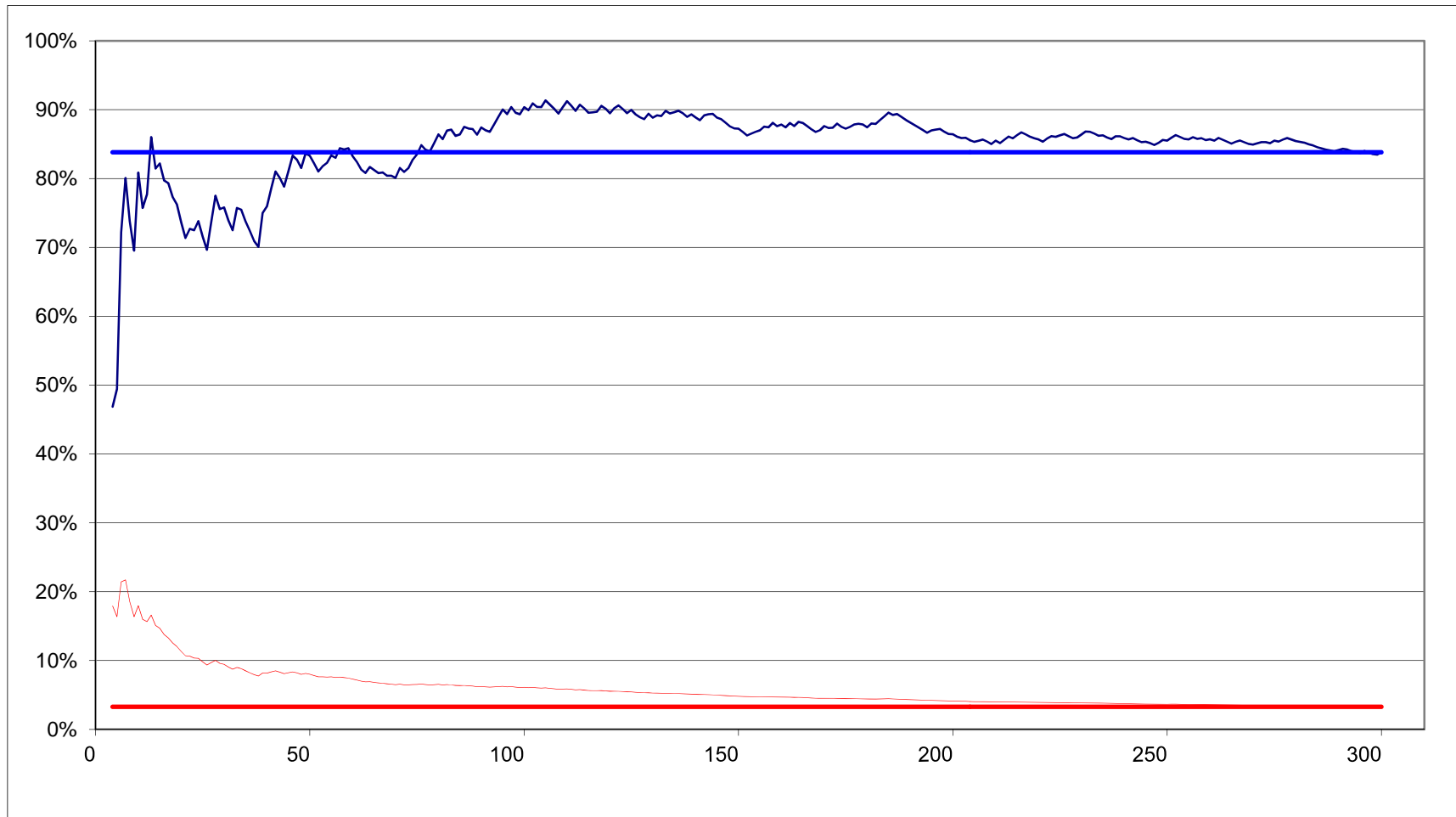"Standard Deviation"

Average

"CV" is just the % form   SD/average = CV

Suppose the data is variable by ± 6 firkins per square rod.

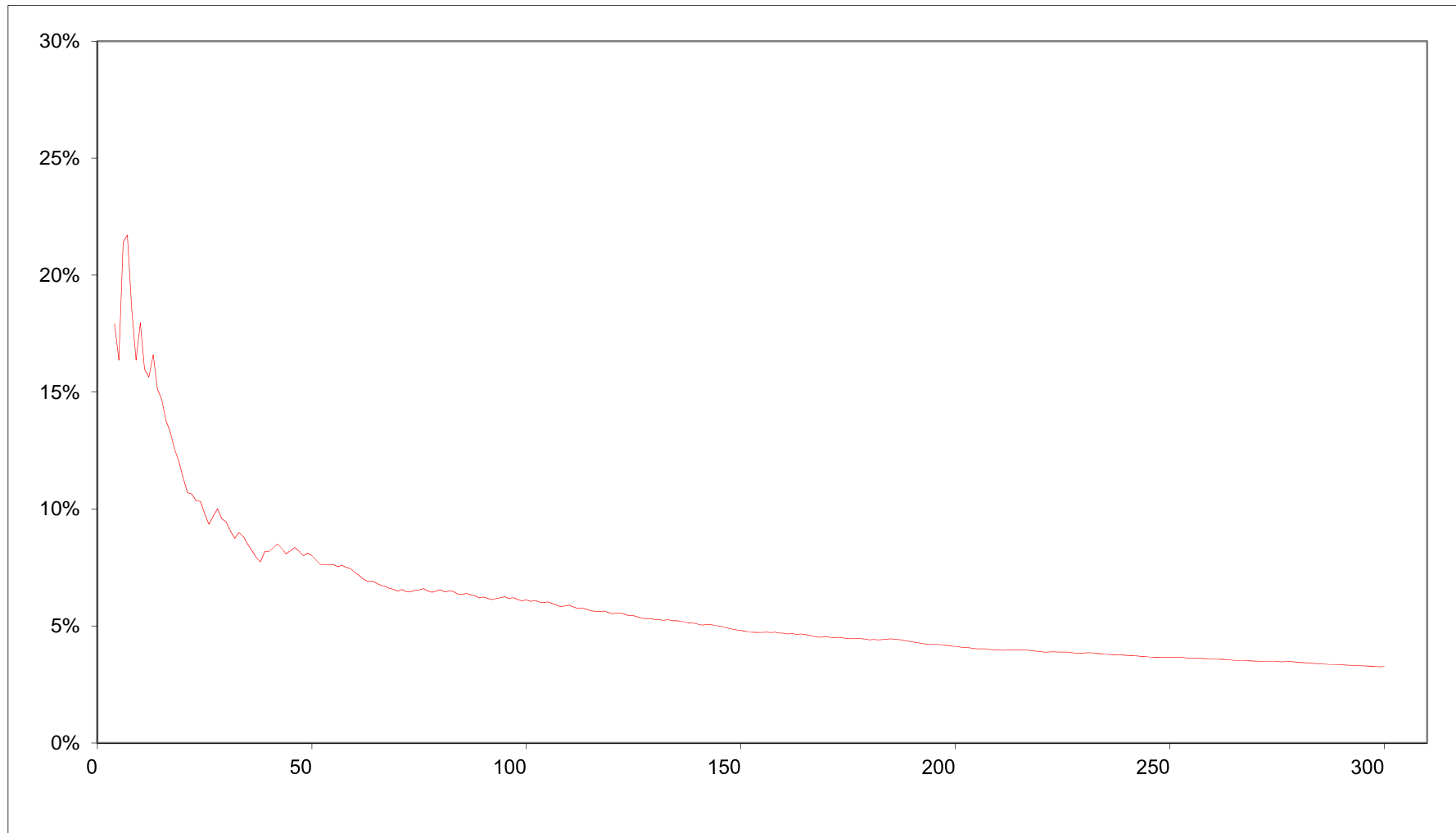"Standard Error" = how close your average is (so far) to the final answer

Sample size

The Population (CV) stabilizes
Standard Error (SE%) reduces to 0%

Diminishing Returns set in

PS ...

- The <u>actual</u> error is a <u>lower</u> with **Systematic Sampling**.

- Sometimes a LOT lower.
- So ... use Systematic Sampling (when you can).

Random Sampling

- **Is a dog**, and always has been.
- It is valued for having **simple math**, …. *nothing* else.

- GPS makes it more possible, in case there is an <u>emotional</u> reason to sample randomly.

Random Sampling

- Random Sampling is loved by Rookies, <u>nobody</u> else.

- Less efficient, harder to do, almost impossible to catch errors or missed plots, more expensive, less credible.

- Does that sound smart ??

What is the statement when using systematic sampling?

from

"We are <u>exactly</u> 95% sure that the error is less than 6%".

to

"We are <u>more than </u>95% sure that the error is less than 6%".

This is just about all you
need to know about statistics
and sampling.

The final issues are clear.

1) What do we need?" is the wrong question.

2) The "95% level" is misleading.

3) Systematic samples give better results than the equations show.

4) Most of our procedures are habit, not logic.

5) Most people oversample,
   often greatly.

6) Sample size is a <u>business</u> decision.
   Think of it that way.

7) Don't beat the problem to death
   by lots of measurements.
   THINK, and sample cleverly.

Thanks ……… Questions ??