Review Slides (for reference)
on
Sampling Distributions, Standard Errors, and Confidence Intervals

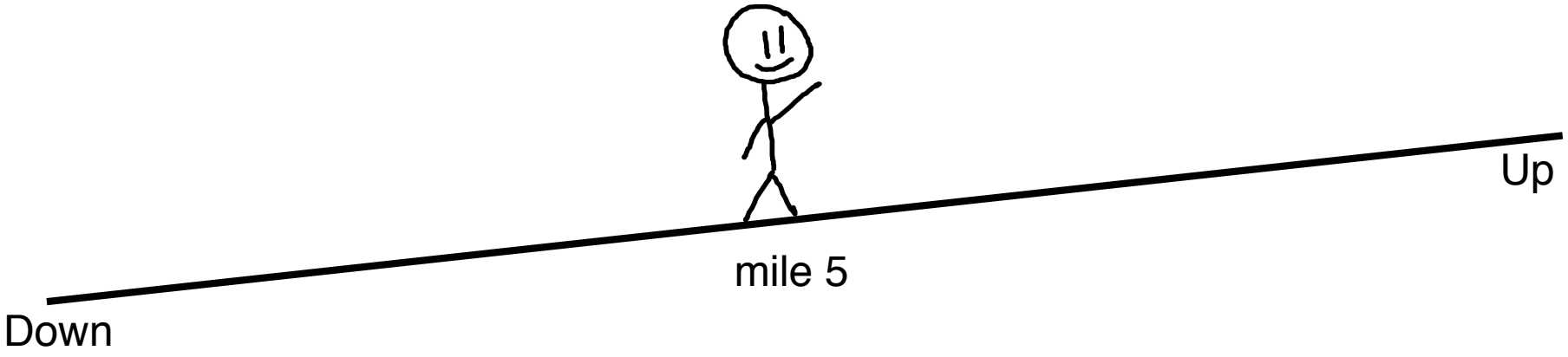
Confidence Interval

A confidence interval for a parameter is an interval computed from sample data by a method that will capture the parameter for a specified proportion of all samples

- The success rate (proportion of all samples whose intervals contain the parameter) is known as the confidence level
- A 95% confidence interval will contain the true parameter for 95% of all samples

Logic of Confidence Intervals

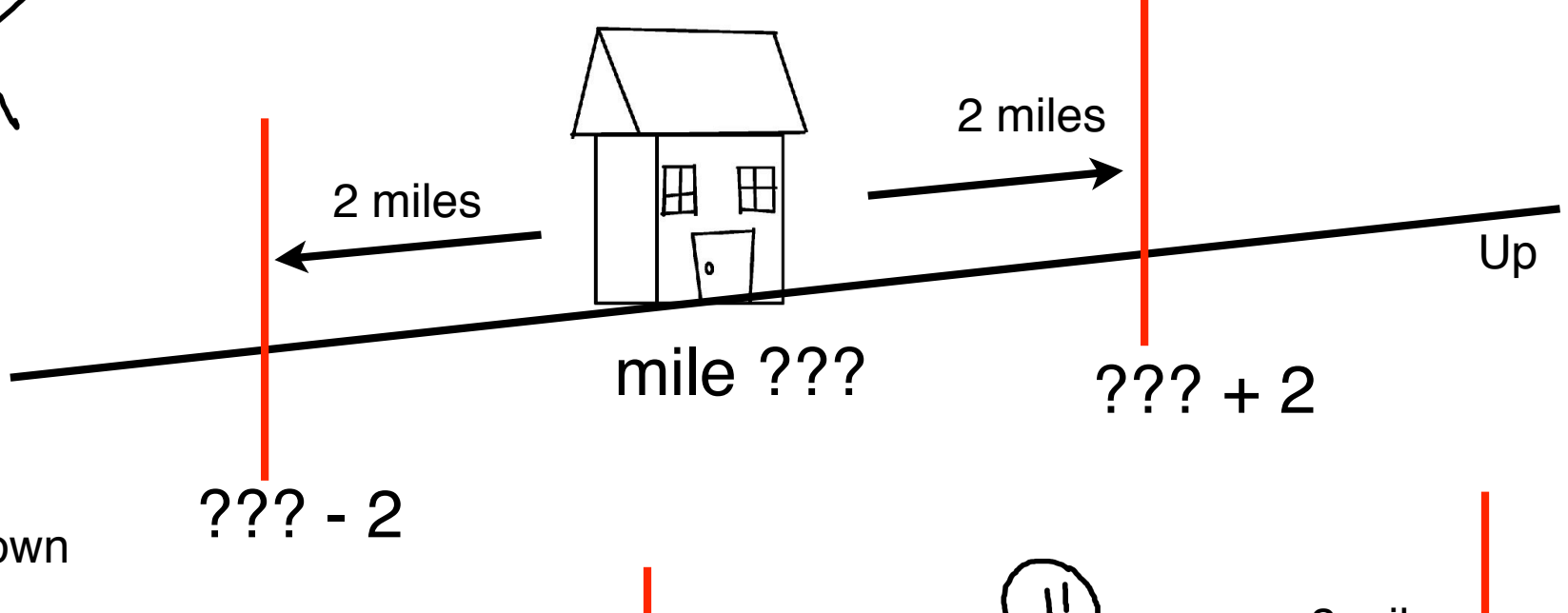
Say you are trying to find a house on a road. A friendly alien drops you on a road in the middle of the night at mile marker 5.



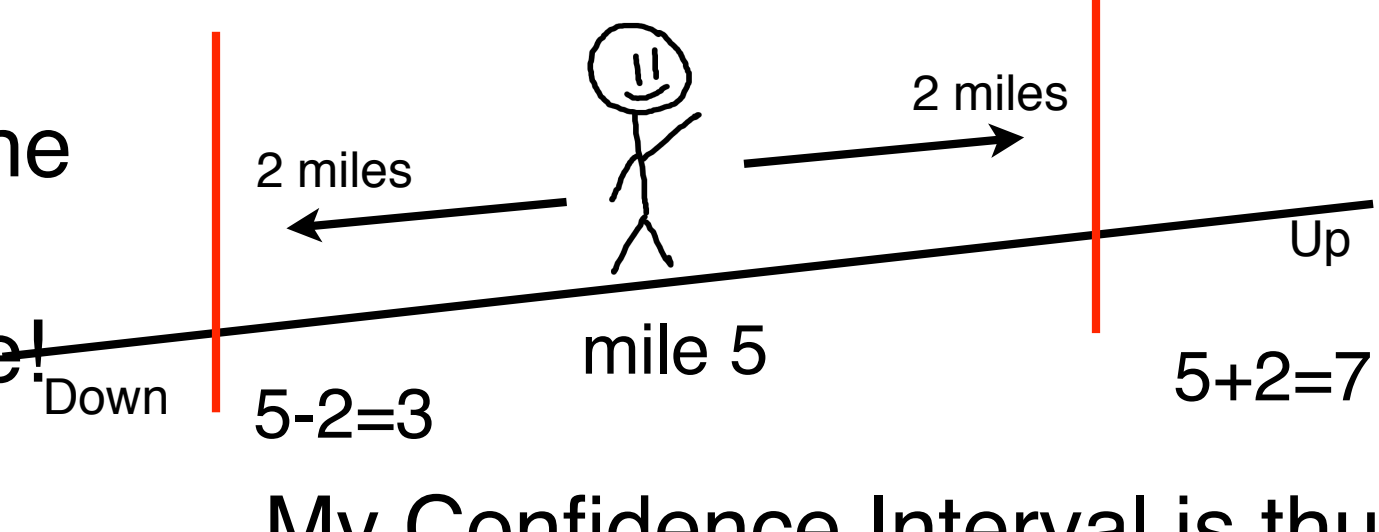
Now say the alien tells you that it isn't perfect, but does drop people no more than 2 miles from their destination 95% of the time.

Logic of Confidence Intervals, Cont.

So I'm fairly certain I was dropped within two miles from the house!



This means the house is 2 miles from me!



My Confidence Interval is thus (3,

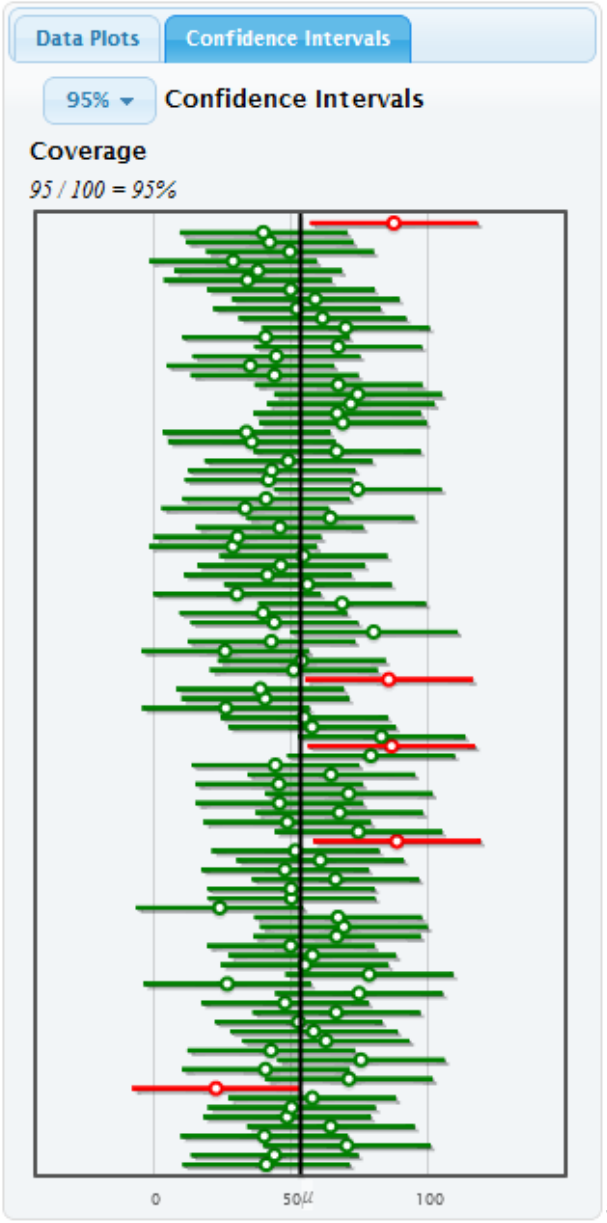
Confidence Intervals

Note:

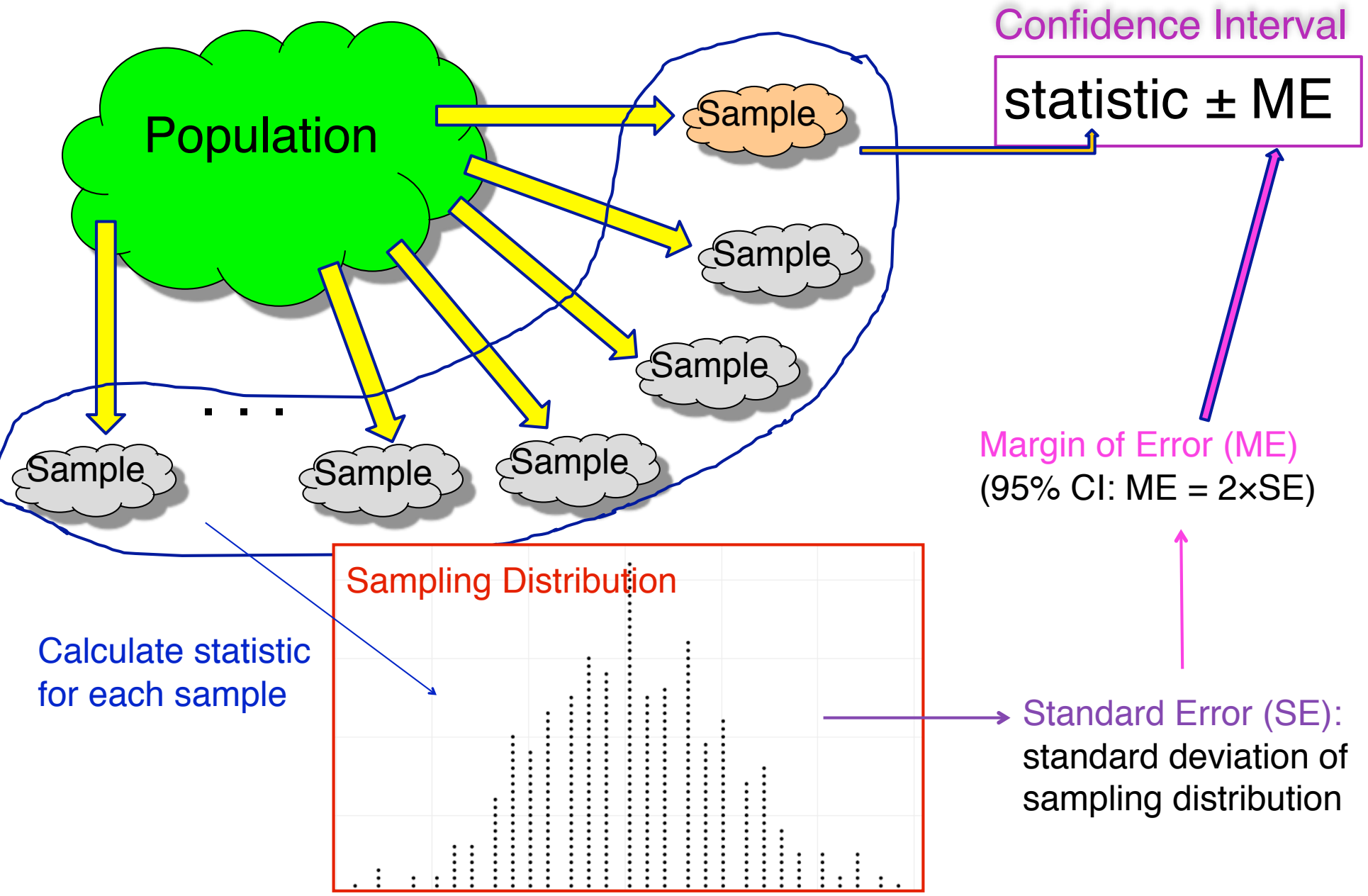
- ★ The parameter is fixed
- ★ The statistic is random (depends on the sample)
- ★ The interval is random (depends on the statistic)

To play around see:

- ★ www.lock5stat.com/StatKey



The Sampling Distribution: Many samples



Finding the Margin of Error

If you had access to the sampling distribution, how would you find the margin of error to ensure that intervals of the form

$$\text{statistic} \pm \text{margin of error}$$

would capture the parameter for 95% of all samples?

Standard Error

The standard error of a statistic, SE, is the standard deviation of the sample statistic

The standard error can be calculated as the standard deviation of the sampling distribution

95% Confidence Interval

If the sampling distribution is relatively symmetric and bell-shaped, a 95% confidence interval can be estimated using

$$\text{statistic} \pm 2 \times \text{SE}$$

Why is a confidence interval better than a significance test?

A significance test is asking only if i.e., if a parameter is equal to some particular number.

A confidence interval tells us ALL of the numbers which the parameter might be equal to.

Confidence intervals also give us significance tests for free (just check if the CI contains the parameter of interest.

The Logic of Confidence Intervals
If I think I am close to you, then I conclude you are close to me

If:

A random sample statistic has a 95% chance of being 1 MoE (Margin of Error) from the true parameter,

Then:

We are 95% confident the true parameter is within 1 MoE of our observed sample statistic!

Interpreting a Confidence Interval

95% of all samples would yield intervals that contain the true parameter, so we say we are "95% sure" or "95% confident" that one interval contains the truth.

"We are 95% confident that the true proportion of all Americans that considered the economy a 'top priority' in January 2012 is between 0.84 and 0.88"

Common Misinterpretations
(DON'T SAY THESE THINGS!)

- ★ "A 95% confidence interval contains 95% of the data in the population"
- ★ "I am 95% sure that the mean of a sample will fall within a 95% confidence interval for the mean"
- ★ "The probability that the population parameter is in this particular 95% confidence interval is 0.95"

Can you explain why these are wrong?

Conceptually How to Make a Confidence Interval

To create a plausible range of values for a parameter:

1. Take many random samples from the population, and compute the sample statistic for each sample
2. Compute the standard error as the standard deviation of all these statistics
3. Now take one more sample, compute its statistics and give:

$$\bar{x} \pm 2 \times SE$$

One small problem…

We only have one sample!

Given that we only have one sample, we have two options:

1) Use something called the bootstrap to simulate multiple samples by resampling from our original data many times.

2) Use the mathematical model for our data to get (asymptotic, approximate) formula for estimating the standard error directly.