Confidence Intervals:

- Confidence and variation are opposites
- Range of values surrounding the sample estimate that is likely to contain the population parameter
- Quantified Plausibility
- Useful as a first pass to see if a phenomenon is worth further study

Correct:

- **"95% of all 95% confidence intervals calculated from samples include the population mean"**

- "The true value of the parameter is likely to be within the 95% CI."

- "I can say with 95% confidence that the true population proportion, p, is between value 1 and value 2"

Not Correct:

- "There is a 95% probability that the population mean is within a particular 95% confidence interval"

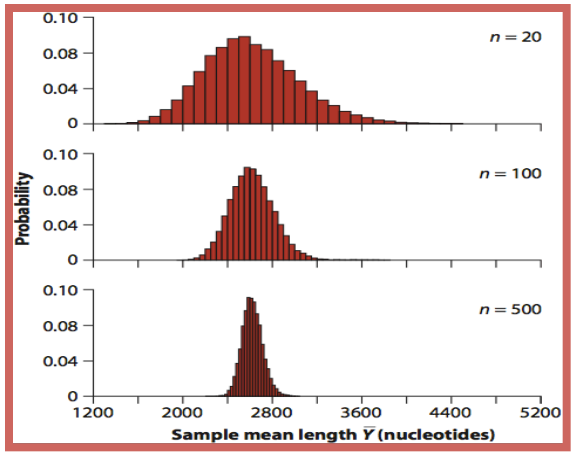
How to correctly talk about confidence intervals is a subtle art.

- Because **you don't know** the true population mean and standard deviation, you have to be a little careful when interpreting the results of a study. Imagine you did a study with 100 students and calculated a BMI of 26.5 and a standard deviation of 2.5. You could then calculate standard error to be 0.25 and from standard error calculate a 95% confidence interval of 26.5+/- 0.25x2 = 26.0 ---- 27.0.

- What you ***can't*** *then say is that "if I repeated my study a large number of times, 95% of the time, the mean body mass index would be between 26 and 27" Why can't you say this? Because you are not using the true mean or the true standard deviation but rather the estimates you obtained from your study.*

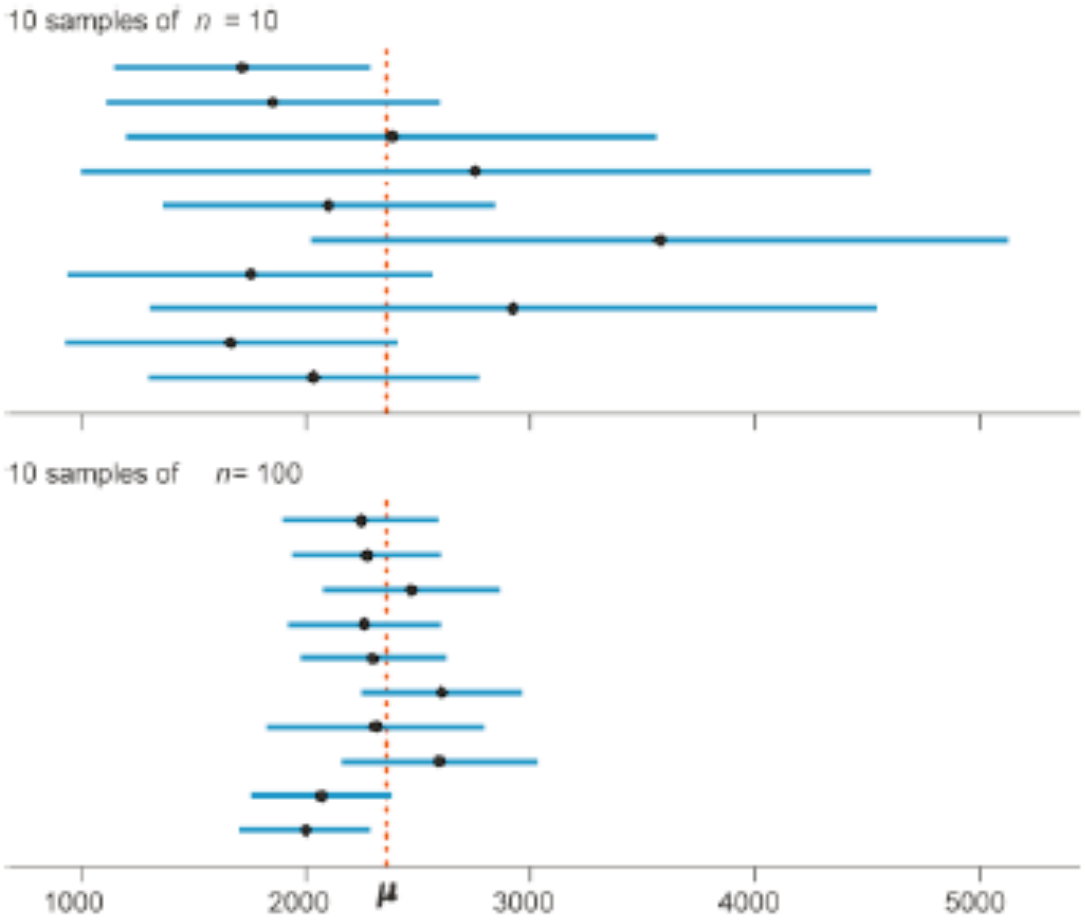
2SE rule of thumb:

The interval from $\overline{Y} - 2SE_{\overline{Y}}$ to $\overline{Y} + 2SE_{\overline{Y}}$ provides a rough estimate of the 95% confidence interval for the population mean

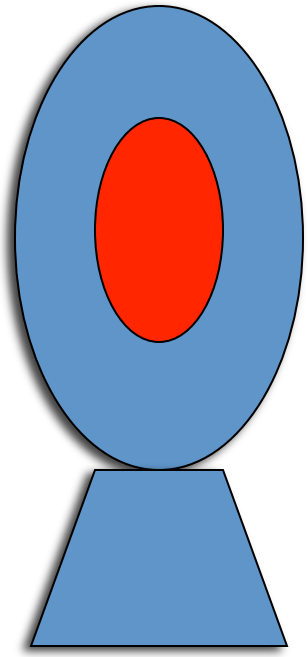


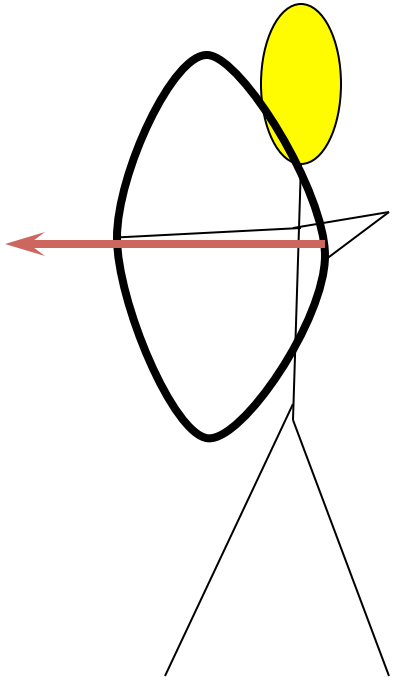
This is due to the fact that the sampling distribution of the mean is approx. normally distributed

Measuring Uncertainty

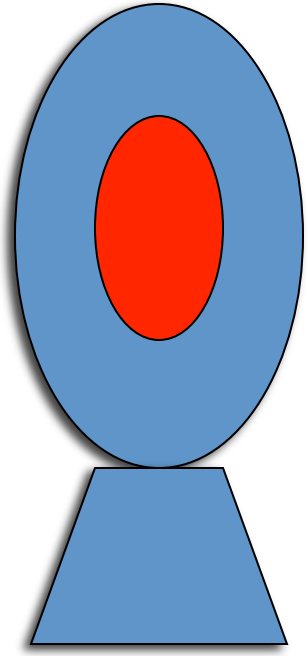


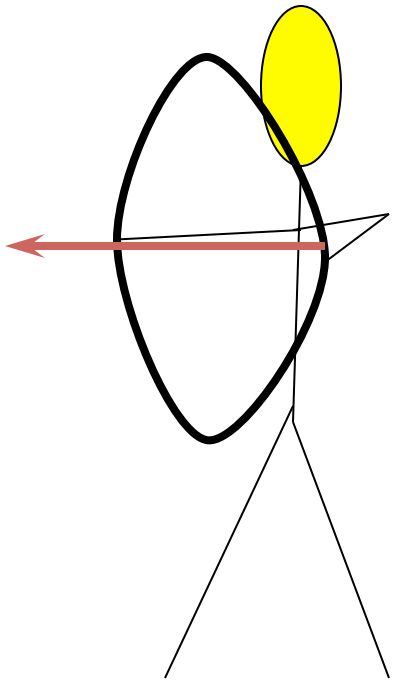
Measuring Uncertainty



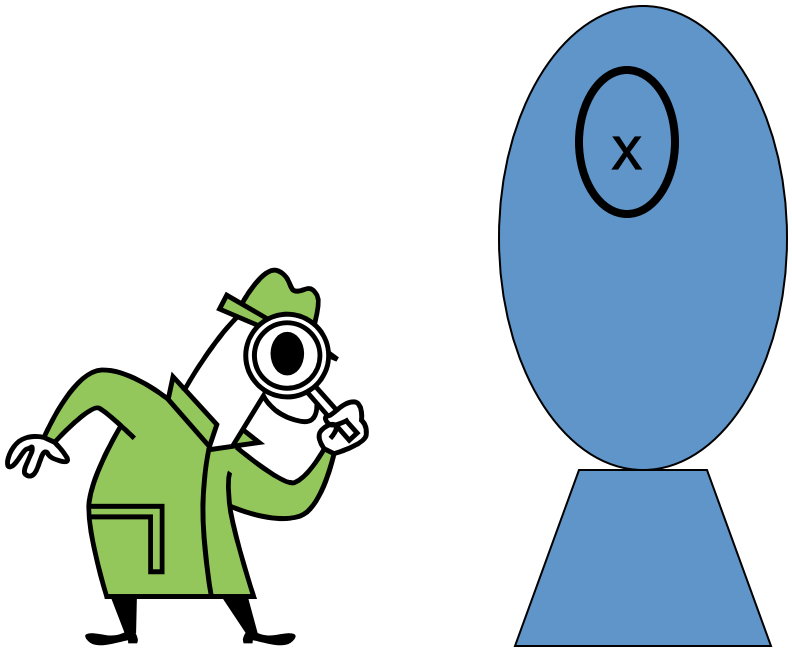


Measuring Uncertainty

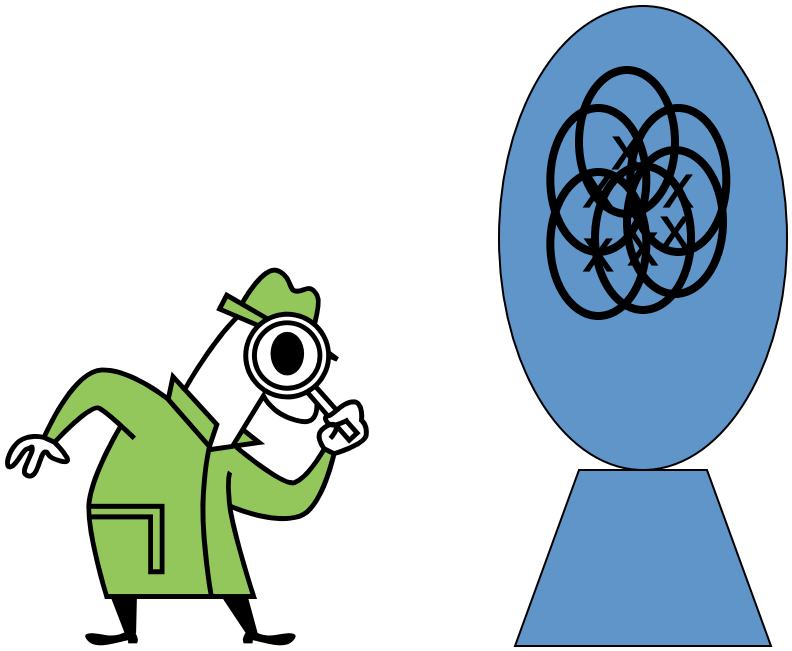




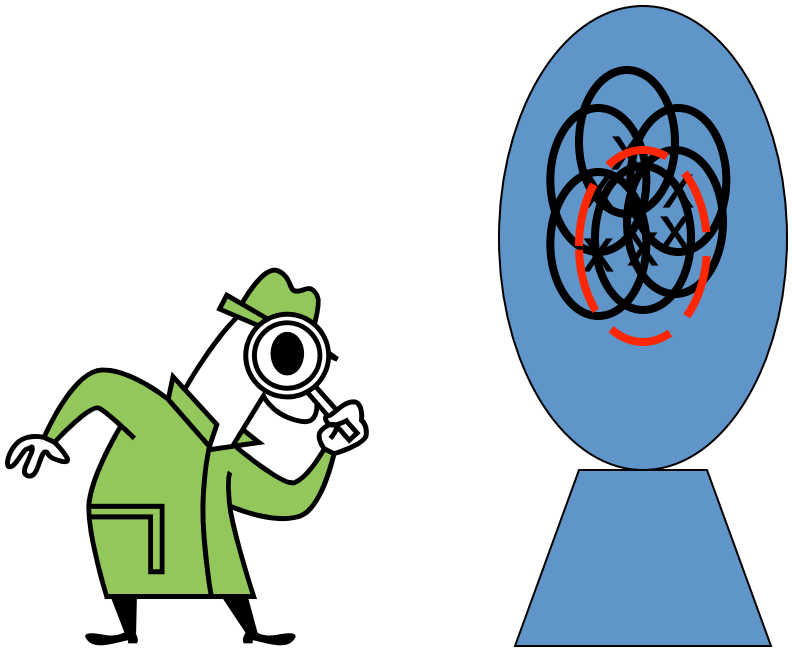
Measuring Uncertainty



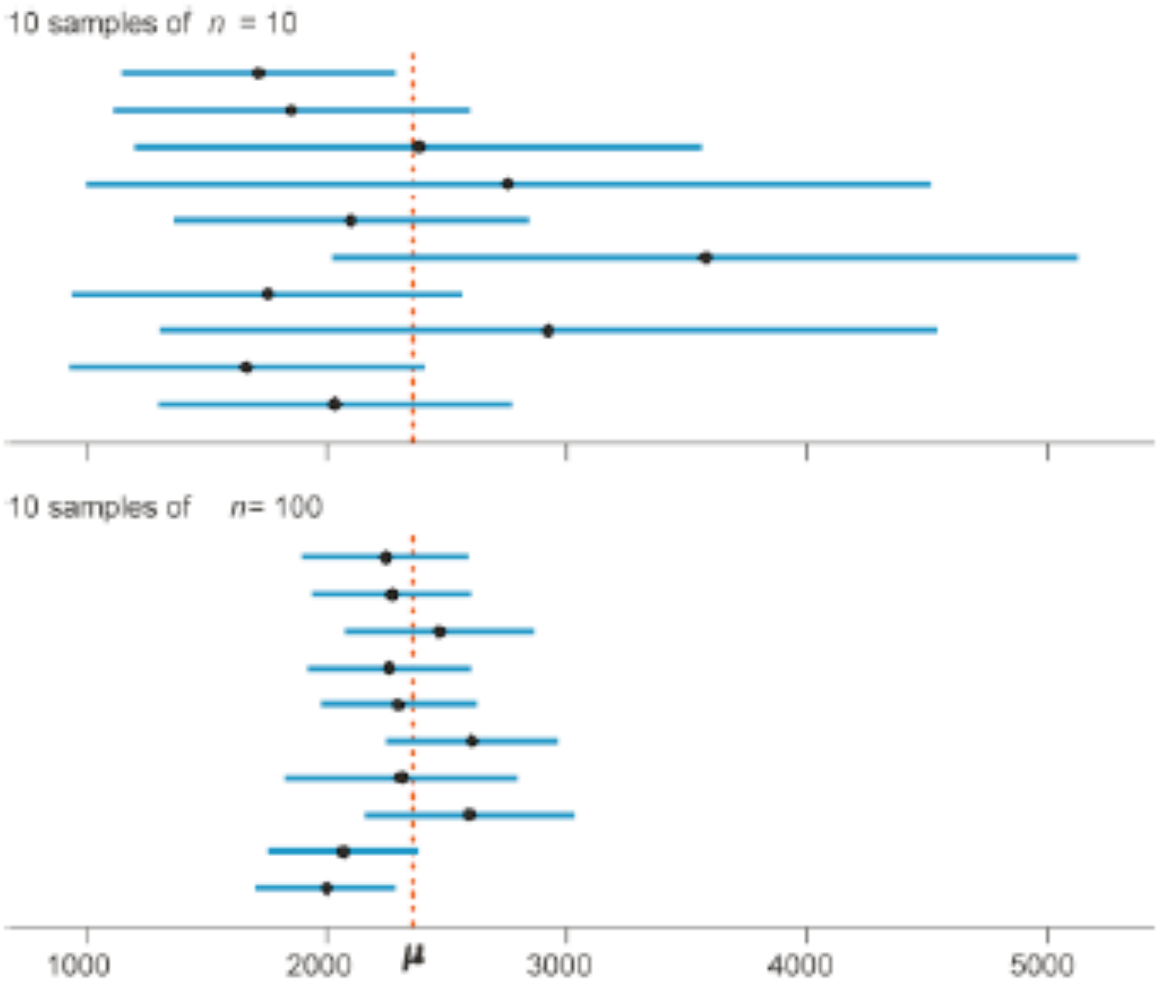
Measuring Uncertainty



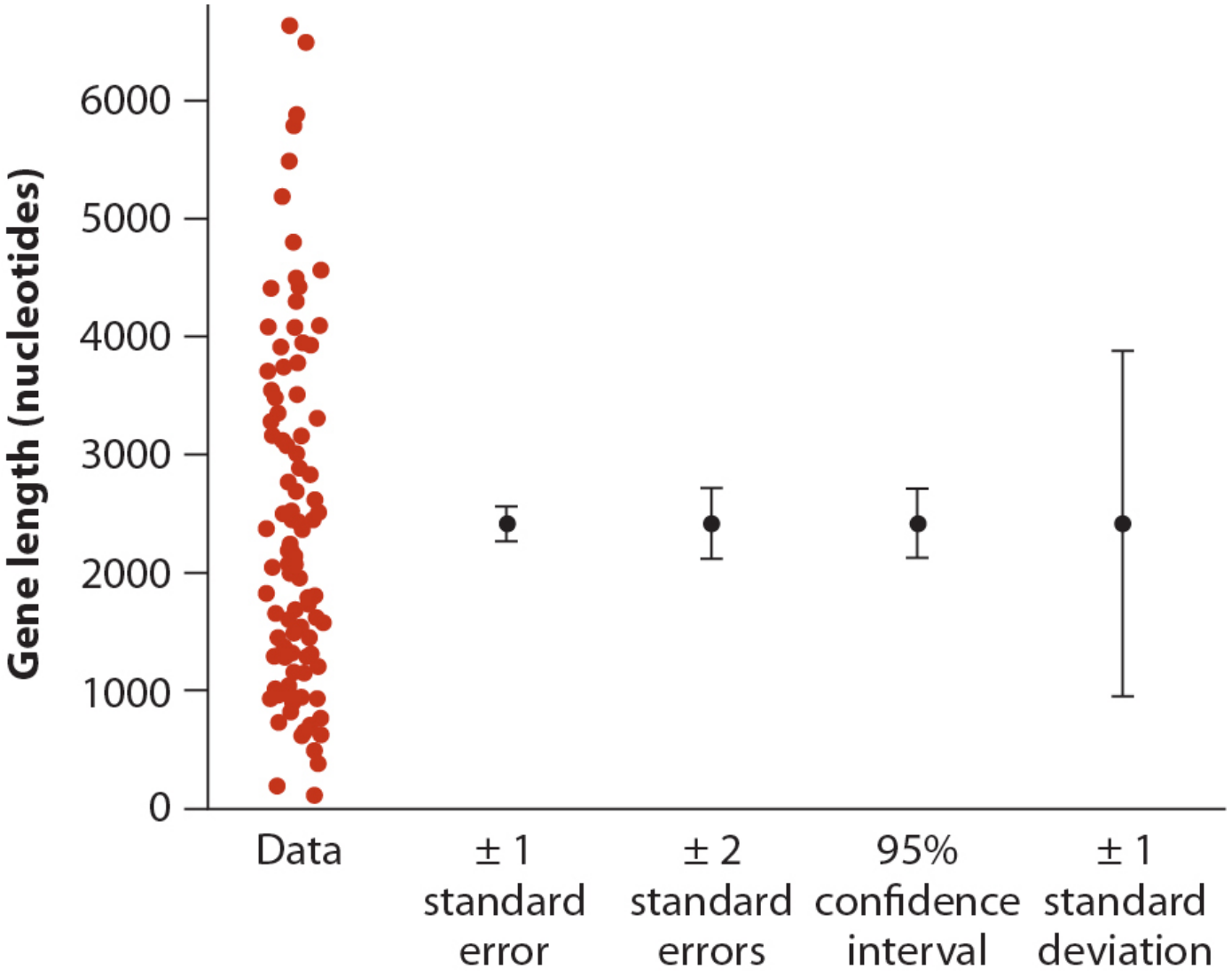
Measuring Uncertainty



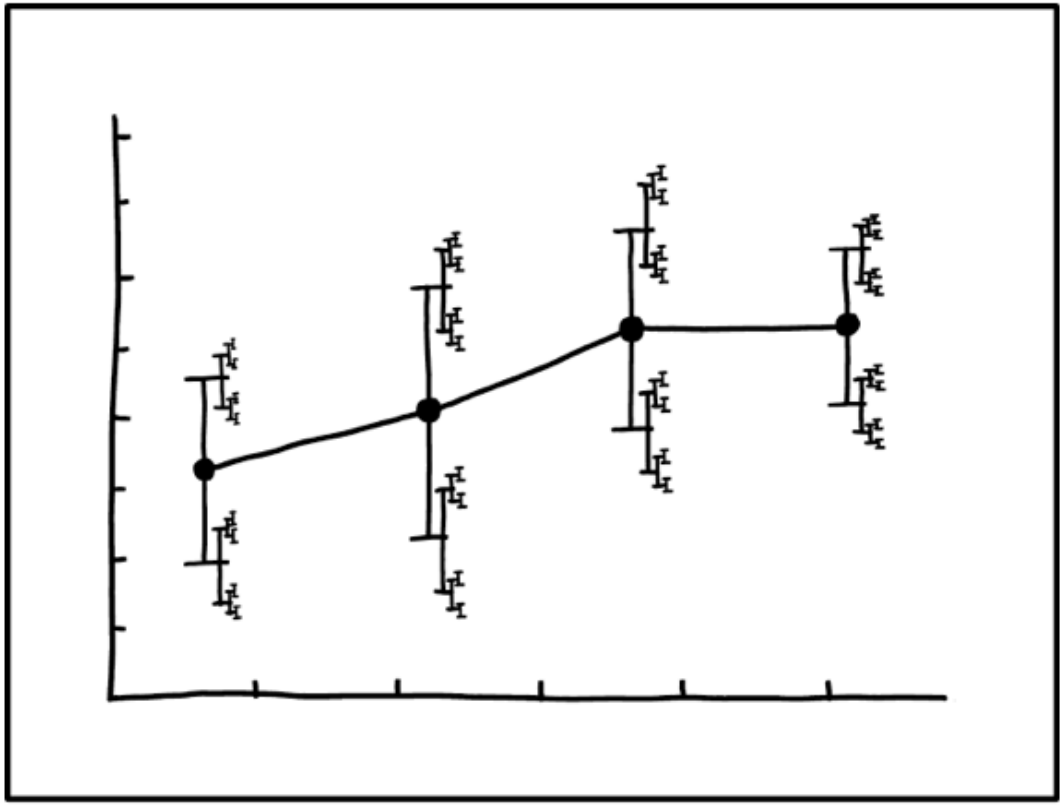
Measuring Uncertainty



Error Bars


Gene length (nucleotides)
6000
5000
4000
3000
2000
1000
0
Data
± 1 standard error
± 2 standard errors
95% confidence interval
± 1 standard deviation

https://xkcd.com/2110/



I DON'T KNOW HOW TO PROPAGATE ERROR CORRECTLY, SO I JUST PUT ERROR BARS ON ALL MY ERROR BARS.

Pseudoreplication:

- Treating non-independent data points as though they are independent so it seems as though we have more data then we actually do!
  - Artificially inflated ‘n’
  - Examples
    - multiple measurements on same individual
    - Surveys of clustered groups of individuals who might share the same characteristics that you are measuring (phylogenetic trees).