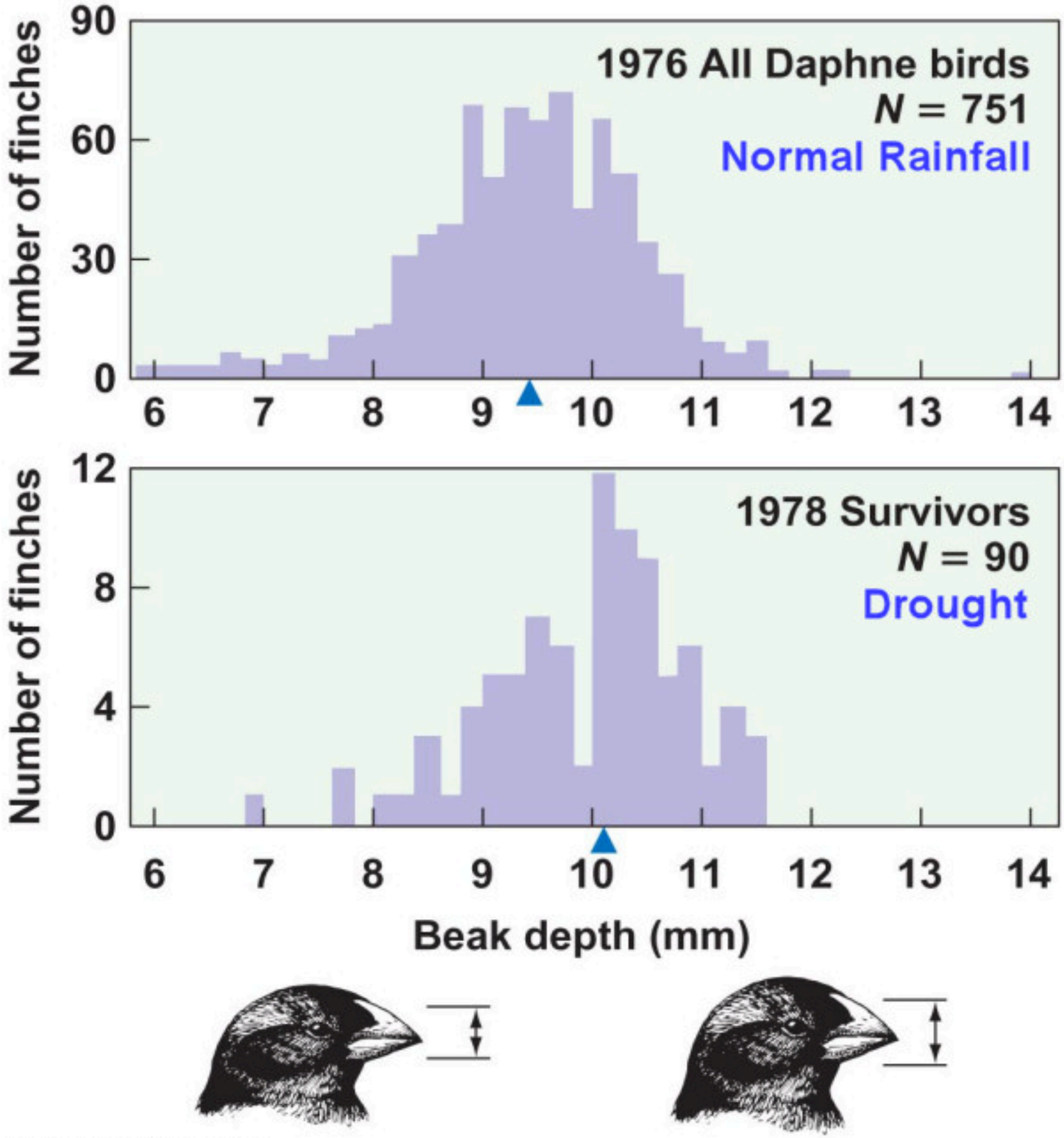
Introduction to Statistics

Sampling Populations

Peter and Rosemary Grant and the Ongoing Evolution of Galapagos Finches



Sampling Populations

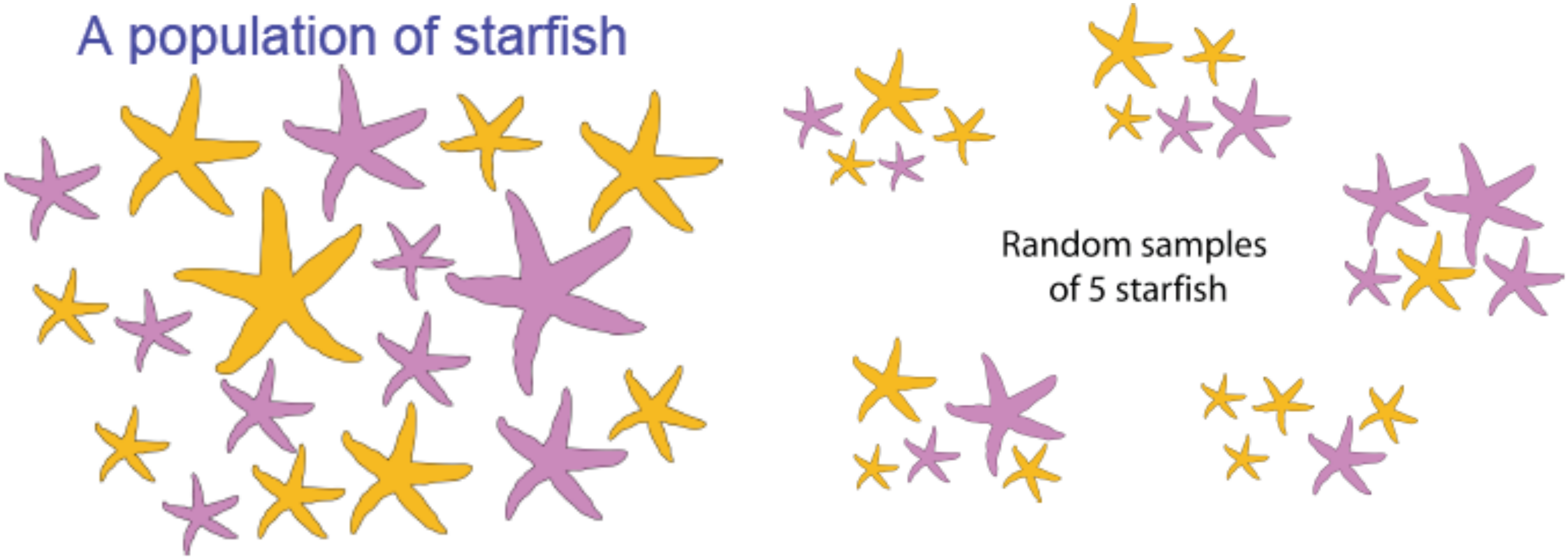
Populations
have
***P*ARAMETERS**

- Represented by Greek Letters
- **μ; σ**

Samples
have
E***S***TIMATES

- Represented by Roman Letters
- $\overline{X}$ **; s**

Sampling Populations



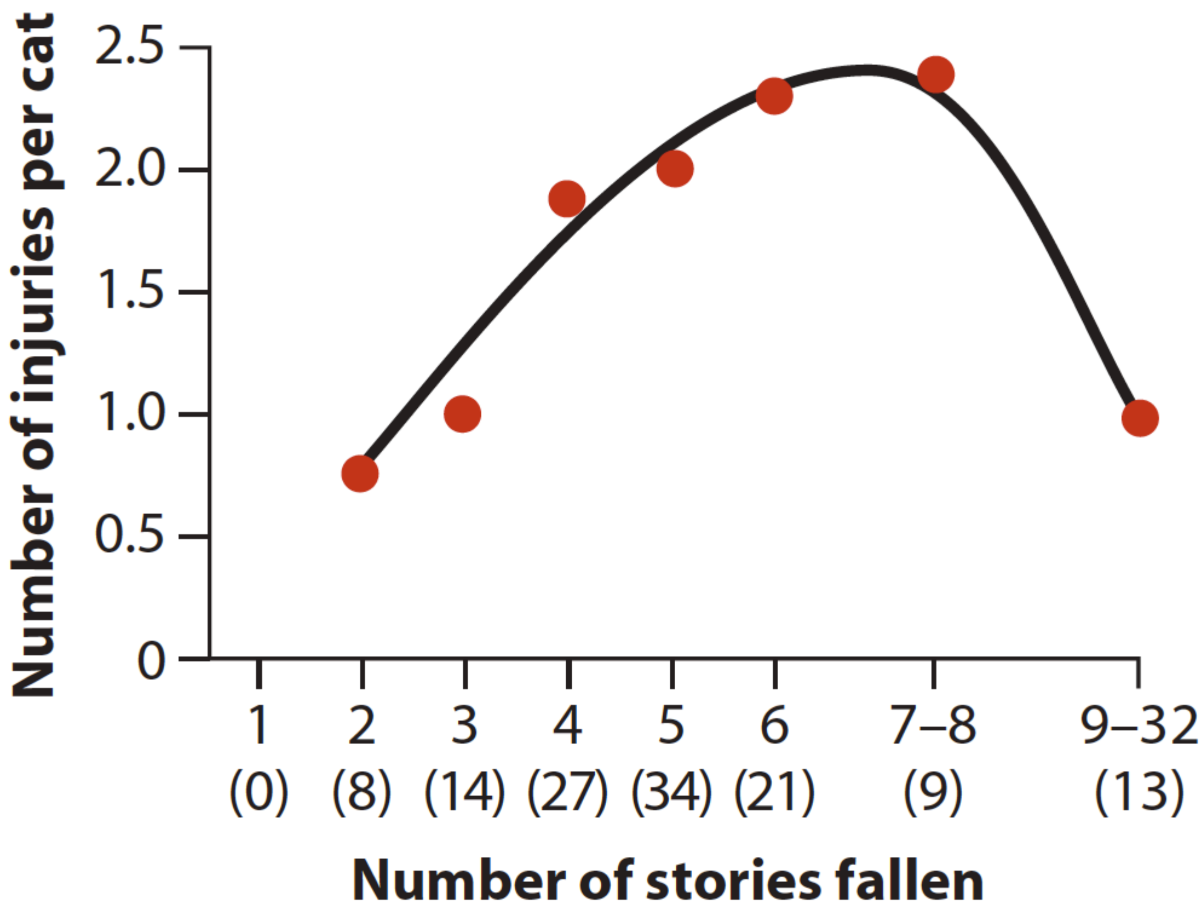
- Bags of colored marbles
- Election polling

Bias:

a systematic discrepancy between estimates and the true population characteristic

A few types of bias

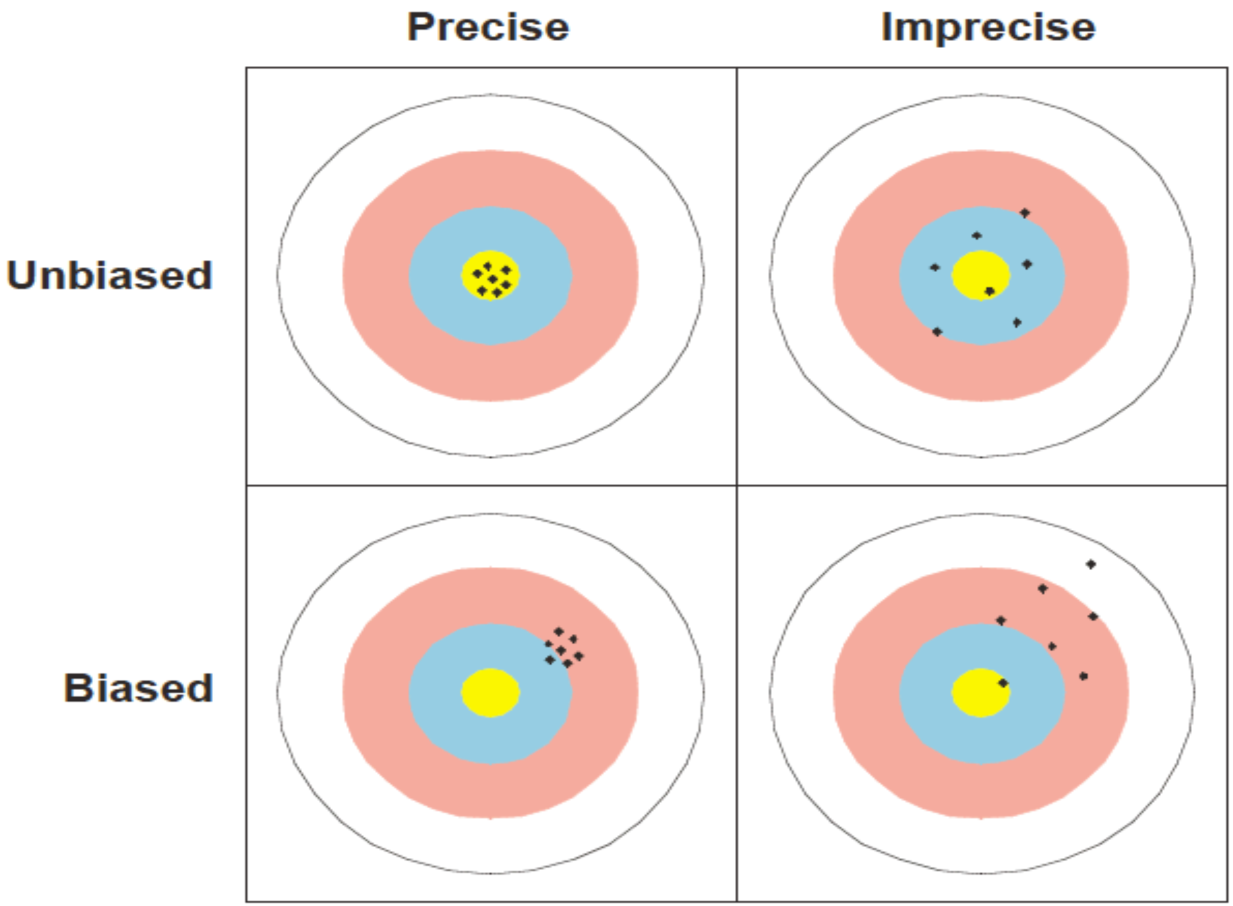
- The sample of convenience
  - Example: cats falling
    - Counter-intuitively higher falls seem to result in lower injury rates



Sampling Populations

What makes a ‘good’ sample?

What makes a good sample?



What makes a good sample?

1. Sufficiently large
   - Larger samples have a smaller ***sampling error***

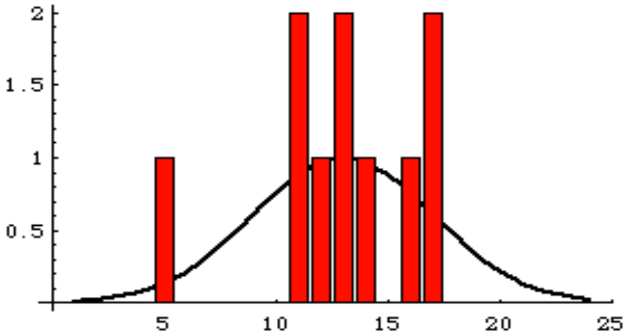
$$SE_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

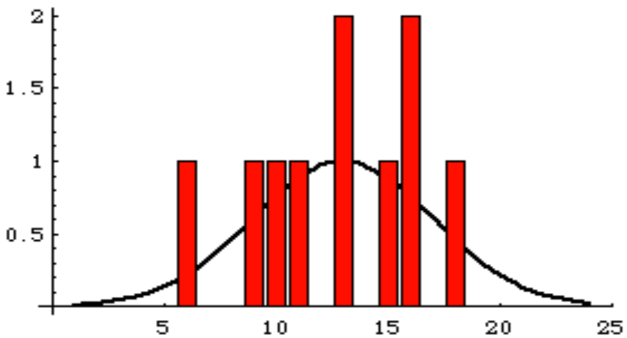
   - *The above equation makes clear the inverse relationship of **SE** and **N***

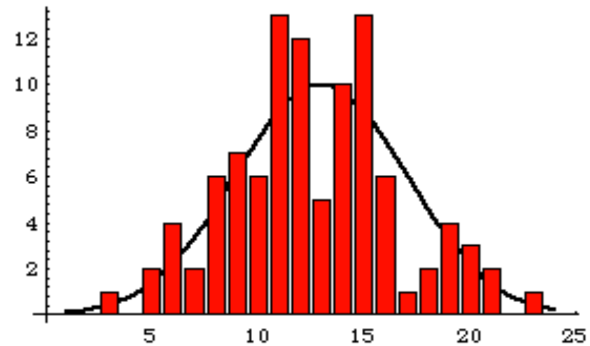
Sampling Populations

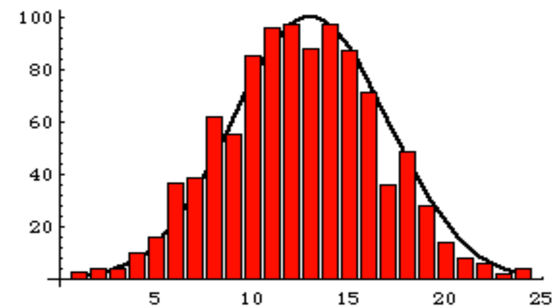
What makes a good sample?

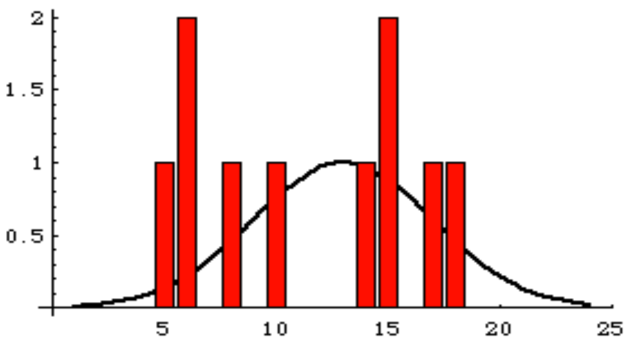
n = 10 | n = 100 | n = 1000











Sampling Populations

What makes a good sample?

1. Sufficiently large
    - Larger samples have a smaller ***sampling error***

2. Randomly sampled
    - Equal chance of selection
        - unbiased
    - Independent
        - The selection of one unit does not influence the selection of any other unit
        - Sample size is smaller than we think if the individuals measured are not independent

    - Random number generator is one way to pick a sample

Random Variable

- The numerical outcome of a random experiment
- Differs between individuals
- Population parameters are constants versus estimates which are also random variables which means that they change from one random sample to the next
- Categorical or numerical

- Ex: Draw a student from the student body
    - Weight, height, grade point average, SAT score are all numerical random variables

- Ex: Toss two dice and allow ***Y*** to represent the sum of sum of the dots on the two dice. ***Y*** can then be any value between 2 and 12.