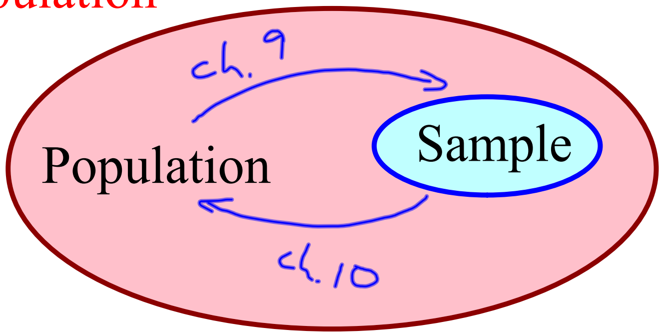
9.1 Sampling distributions

One day some papers catch fire in a wastebasket in the Dean’s office. Luckily, a physicist, a chemist, and a statistician happen to be nearby. Naturally, they rush in to help. The physicist whips out a notebook and starts to work on how much energy would have to be removed from the fire in order to stop the combustion. The chemist works on determining which reagent would have to be added to the fire to prevent oxidation. While they are doing this, the statistician is setting fires to all the other wastebaskets in the adjacent offices. "What are you doing?" the Dean demands. To which the statistician replies, "To solve a problem of this magnitude, you need a large sample size."

Recall these terms:

parameter: a fixed, often unknown # describing a population



statistic: a known # that describes a particular sample, but the # likely changes from sample to sample

We estimate parameters 2 ways:
point estimates (ch. 9) and
interval estimates (ch. 10).

		mean	variance	standard deviation	proportion
Use these:	statistics	$\bar{x}$	s^2	s	$\hat{p}$
to estimate these:	parameters	μ	σ^2	σ	p π

P = proportion of all students w/ blue eyes

#1 Random samples are good
should represent the population well
statistic should reasonably estimate parameter

#2 Statistics have error
rarely = parameters exactly

#3 Statistics have distributions
they vary
there's a mean and standard deviation

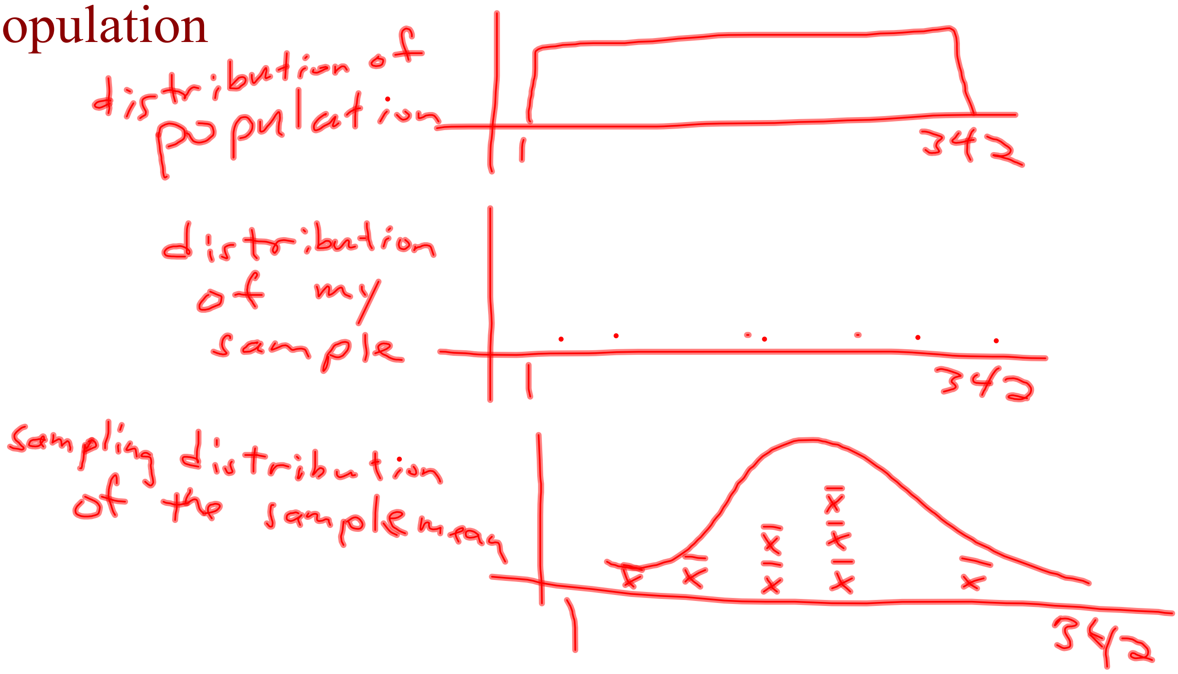
#4 Larger sample size is better
more info about population
less error

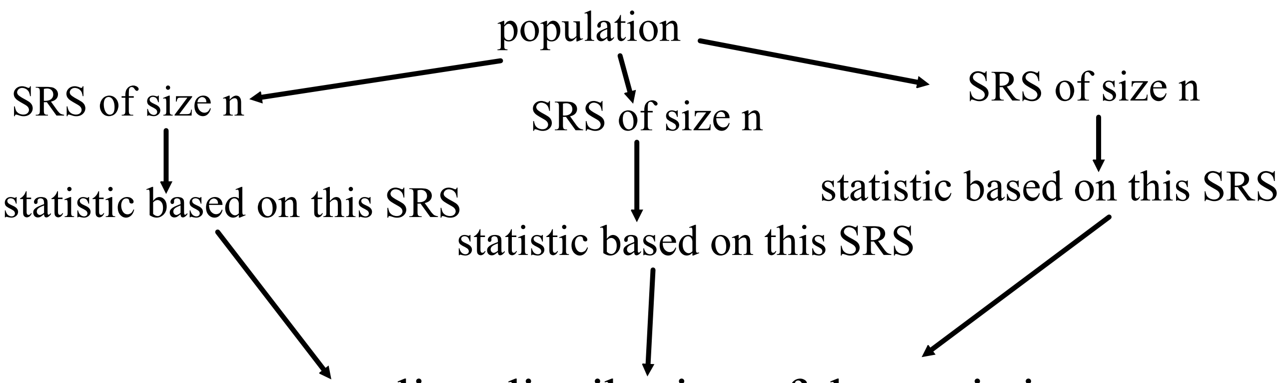
sampling variability:
different samples yield different values of statistics

sample mean

sampling distribution of a statistic:

- probability distribution of the statistic
- distribution of values taken by the statistic in all possible samples of the same size from the same population





sampling distribution of the statistic:

S specific shape

O few, if any

C close to the parameter

S smaller when n is bigger

A statistic is **unbiased** if the
mean of its sampling distribution = the parameter estimated

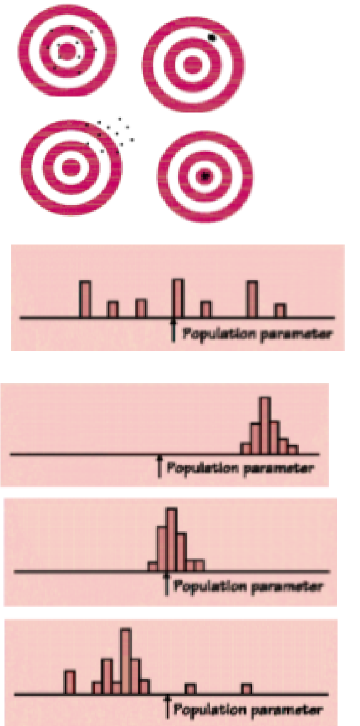
The **variability of a statistic** is described by the spread of its sampling distribution.

Larger samples give smaller spread, but our formulas are for samples under 10% of the population (N>10n).

Drag these displays to the appropriate places!

	Low Variability	High Variability
Low Bias	The ideal situation is here!	
High Bias		

Drag from here.



Suppose we expect a parameter to be some #.
SRS $\Rightarrow$ statistic
Why would statistic $\neq$ parameter?

- sampling procedure was biased

    and/or
- chance error

    and/or
- the parameter $\neq$ what we assumed