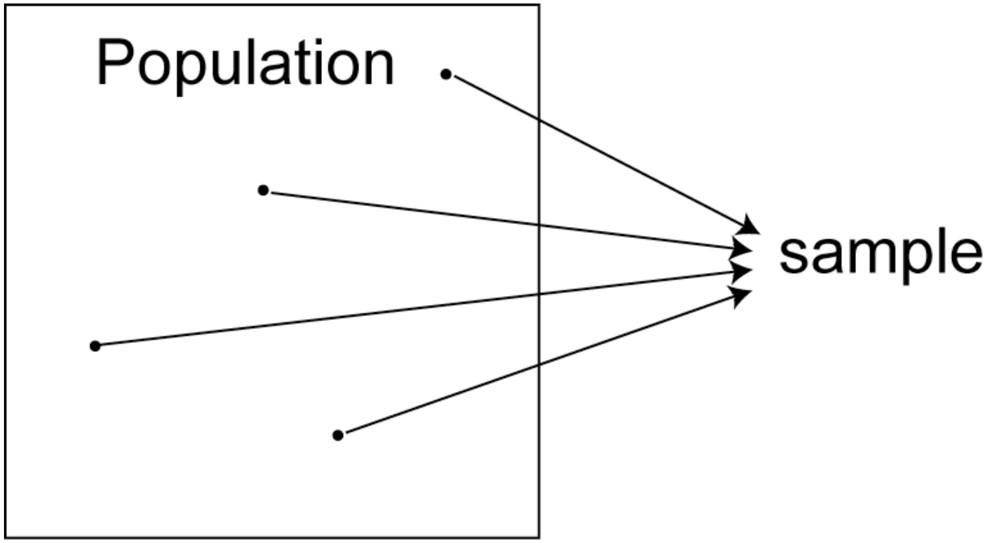
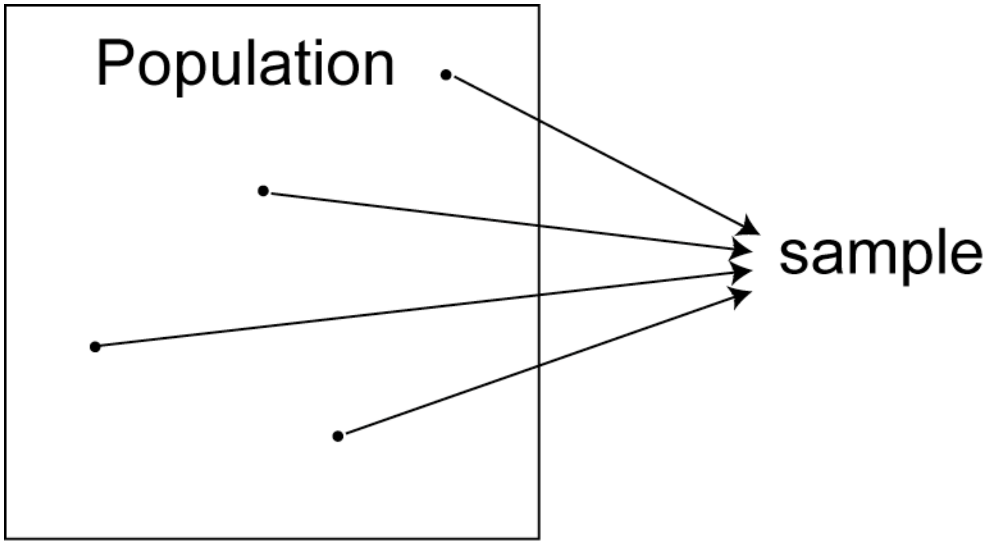
Simple Random Sample (SRS):



Why do we study Probability?

- Random samples eliminate bias (that’s good)
- Random samples will vary from sample to sample (that’s bad)

We will use probability to tell us what will happen if we take very many samples.

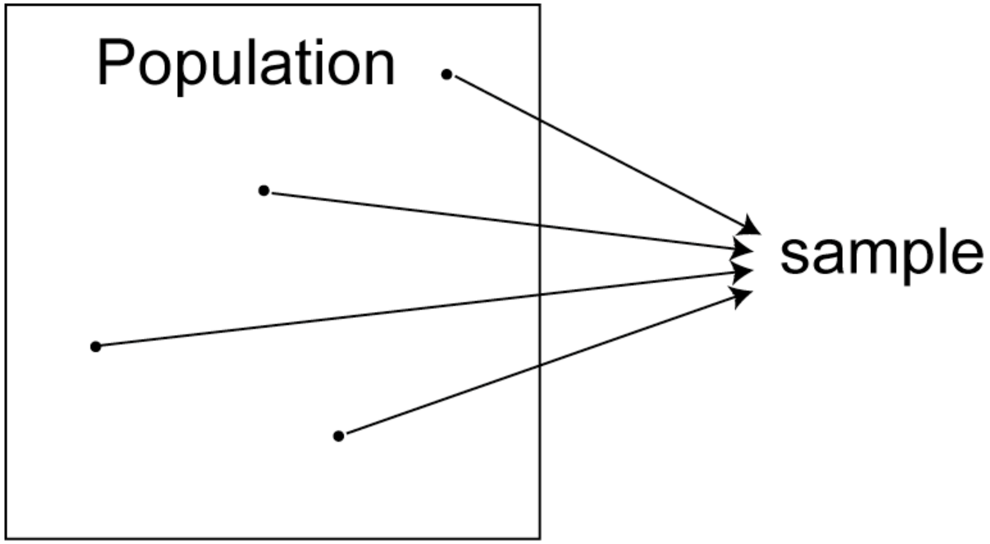


Population

sample

A number associated with the population is called a parameter.
(Typically unknown.)
A number associated with a sample is called a statistic.
(Use to estimate parameter.)

Example:
Parameter: mean number μ of study hours, all college students
Statistic: sample of 100 students, mean = 11.6 hours per week



Population

sample

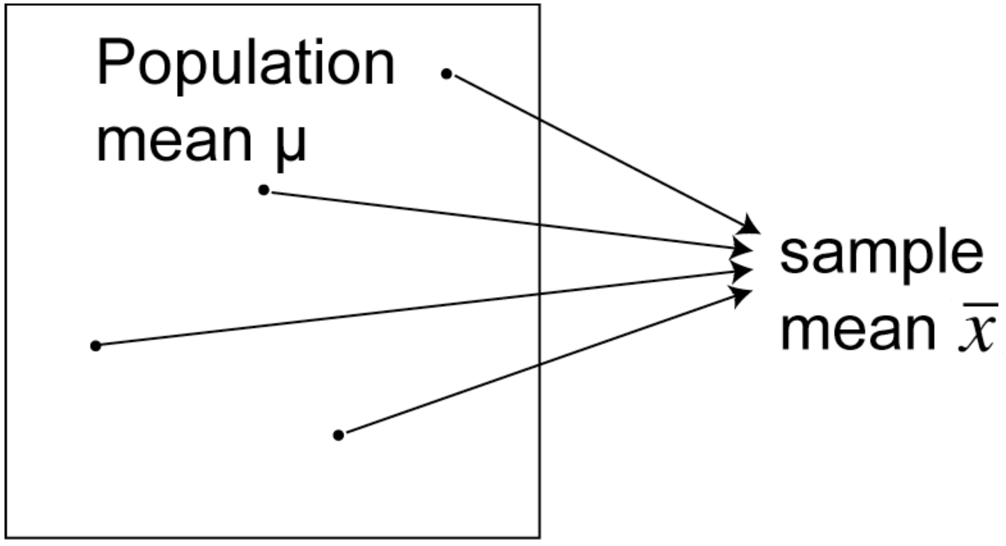
Example:

Parameter: mean level μ of dimethyl sulfide (DMS) in wine that is detectable by humans

Statistic: sample of 10

28 40 28 33 20 31 29 27 17 21

$\bar{x} = 27.4$ mg/dl



Law of Large Numbers:

As the sample size *n* of a SRS increases, the mean $\bar{x}$ of the sample gets closer and closer to the mean μ of the population.

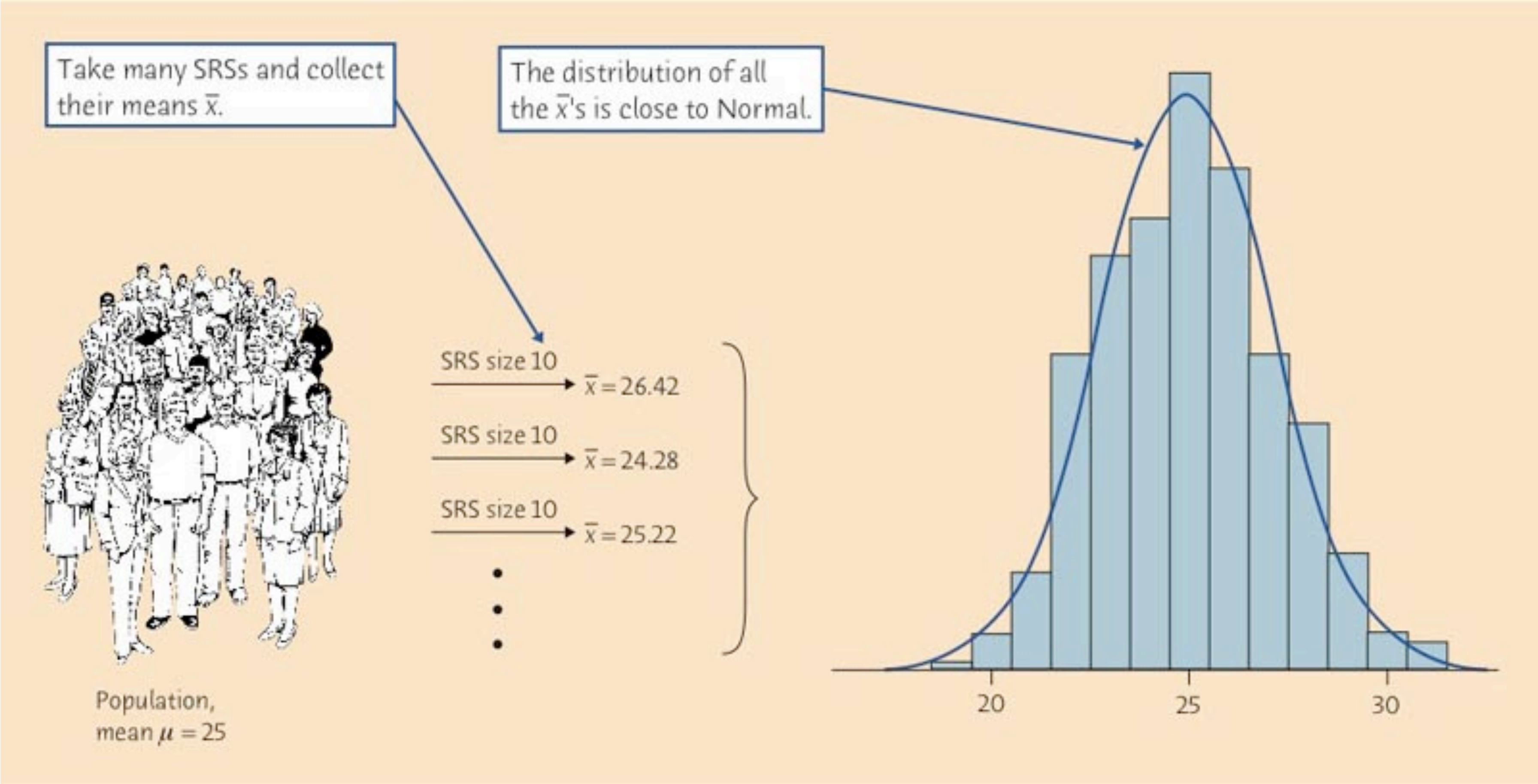
Sampling Distributions:

Example:
Parameter: mean level μ of dimethyl sulfide (DMS) in wine that is detectable by humans

- Imagine a large number of *different* samples of size 10

- Calculate the sample mean $\bar{x}$ for each.

- Make a histogram of the values of $\bar{x}$.


Take many SRSs and collect their means $\bar{x}$.
The distribution of all the $\bar{x}$'s is close to Normal.
SRS size 10
$\bar{x} = 26.42$
SRS size 10
$\bar{x} = 24.28$
SRS size 10
$\bar{x} = 25.22$
Population, mean $\mu = 25$
20
25
30

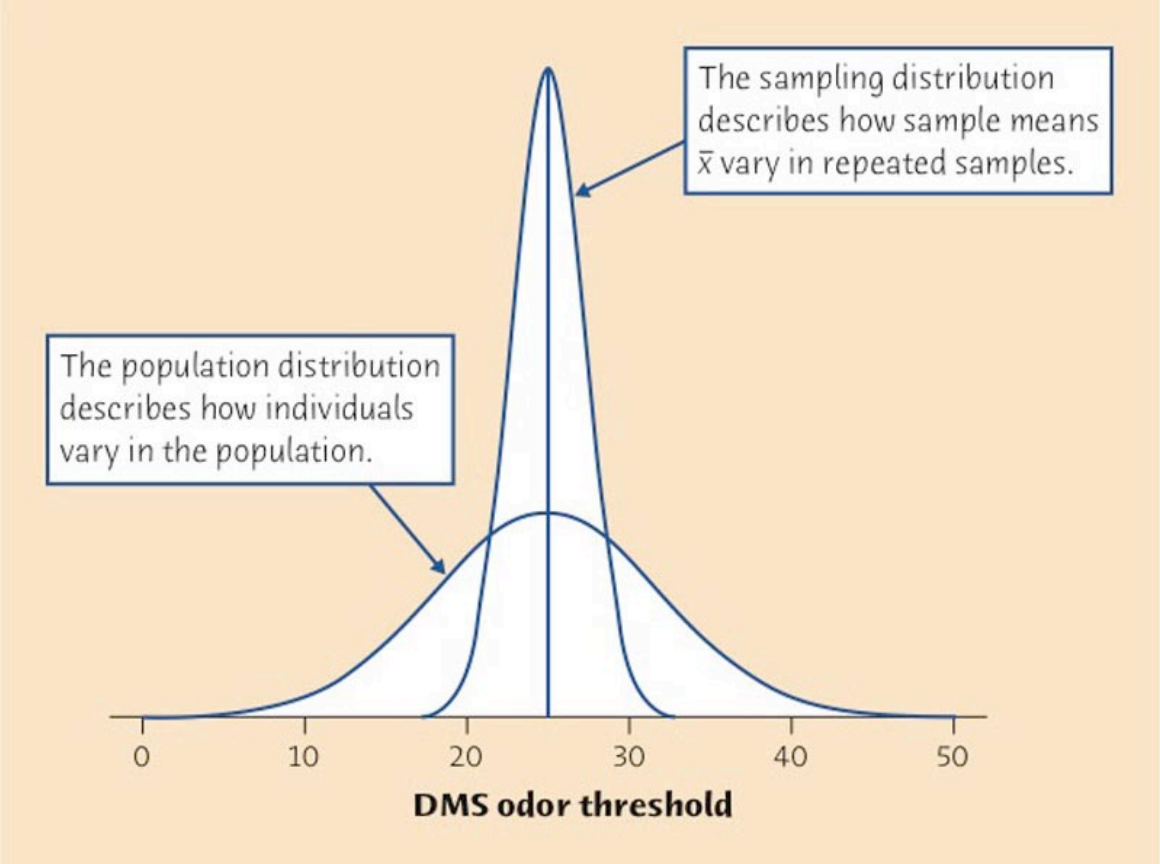
Sampling Distribution of a Sample Mean:

If individual observations have normal distribution N(μ,σ), then the sample mean $\bar{x}$ of a SRS of size n has a normal distribution N(μ , $\sigma/\sqrt{n}$).

This says, in general, that averages have less variation than individual observations.

Example: DMS odor threshold known to be $N(25, 7)$ mg/dl.

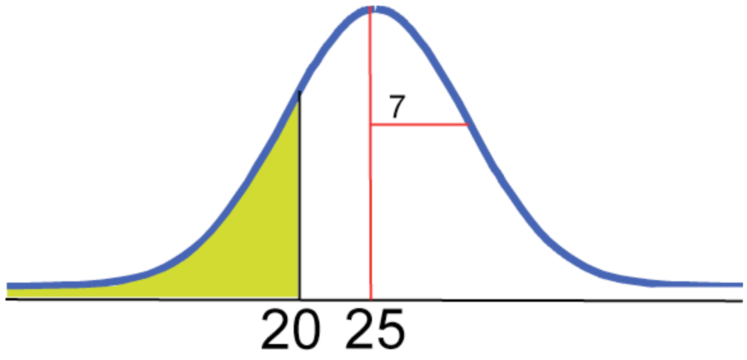
So sampling distribution for $\bar{x}$ for samples of size 10 is $N(25\ ,\ 7/\sqrt{10}) = N(25, 2.2136)$.



- What is the probability a randomly chosen person has a DMS threshold below 20 mg/dl?

$z = (20–25)/7 = –0.71$

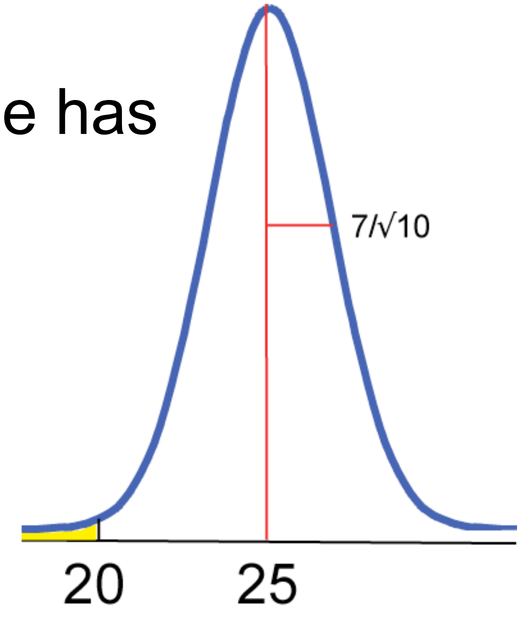
$P(X < 20) = P(Z < –0.71) = 0.2389$



- What is the probability a sample of 10 people has a mean DMS threshold below 20 mg/dl?

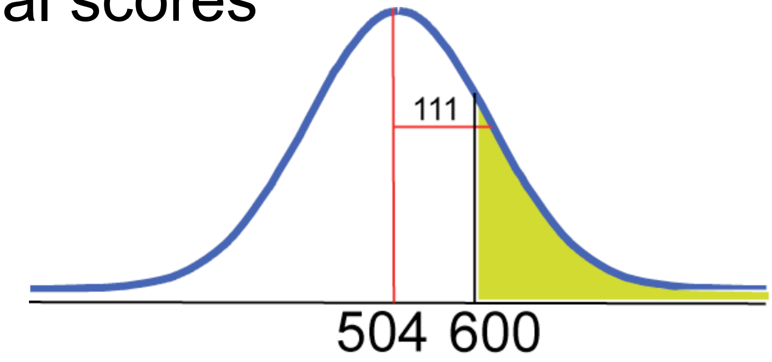
$$z = (20 - 25)/(7/\sqrt{10}) = -2.26$$

$$P(\overline{X} < 20) = P(Z < -2.26) = 0.0119$$



Example: SAT verbal exam, individuals ~ N(504, 111)

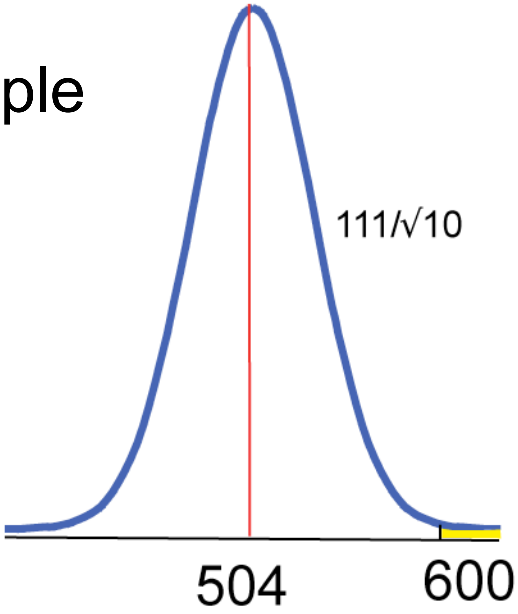
- What is the probability an individual scores above 600?



$$z = (600 - 504)/111 = 0.86$$

$$P(X > 600) = P(Z > 0.86) = 0.1949$$

- What is the probability a sample of 10 people has a mean score above 600?



$$z = (600 - 504)/(111/\sqrt{10}) = 2.73$$

$$P(\overline{X} > 600) = P(Z > 2.73) = 0.0032$$

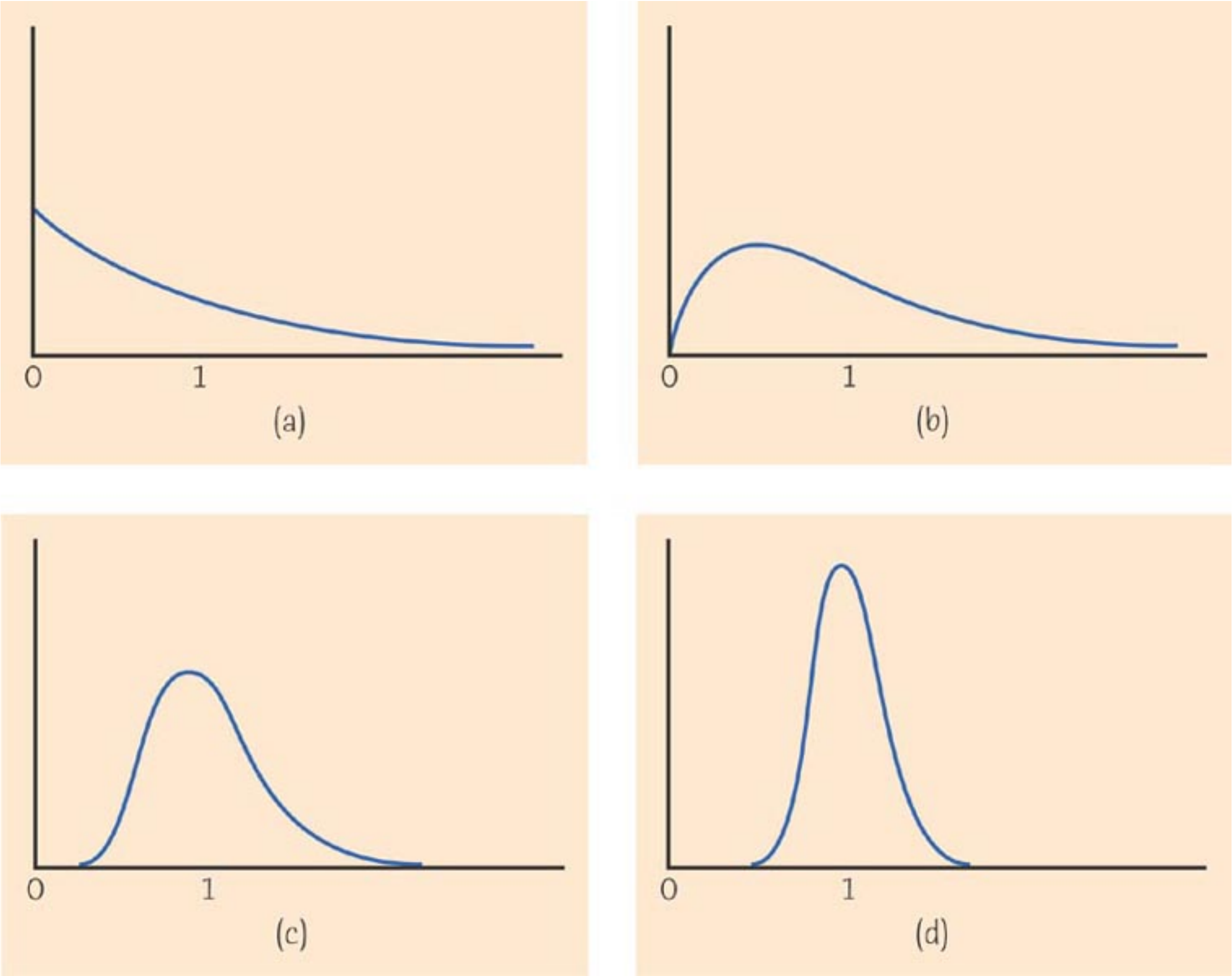
Central Limit Theorem:
SRS of size n from a population with mean μ and standard deviation σ, *not necessarily normal*.

Then for large n, the sample mean is approximately normal:

$$\bar{x} \text{ is approximately } N(\mu\ ,\ \sigma/\sqrt{n})$$

The central limit theorem in action for a skewed distribution:

(a) n=1.
(b) n=2.
(c) n=10.
(d) n=25.



Example: Home values in a community have mean μ = \$400,000, standard deviation σ = \$95,000.

Distributions of home values are generally skewed.

However, for samples of 50 homes,

$$\bar{x} \text{ is approximately } N(400000,\ 95000/\sqrt{50})$$

What is the probability a sample of 50 homes has average price below \$350,000?

$$z = (350{,}000 - 400{,}000)/(95{,}000/\sqrt{50}) = -3.72$$

$$P(\bar{X} < 350{,}000) = P(z < -3.72) < 0.0002$$

(z is out of the range of Table A.)