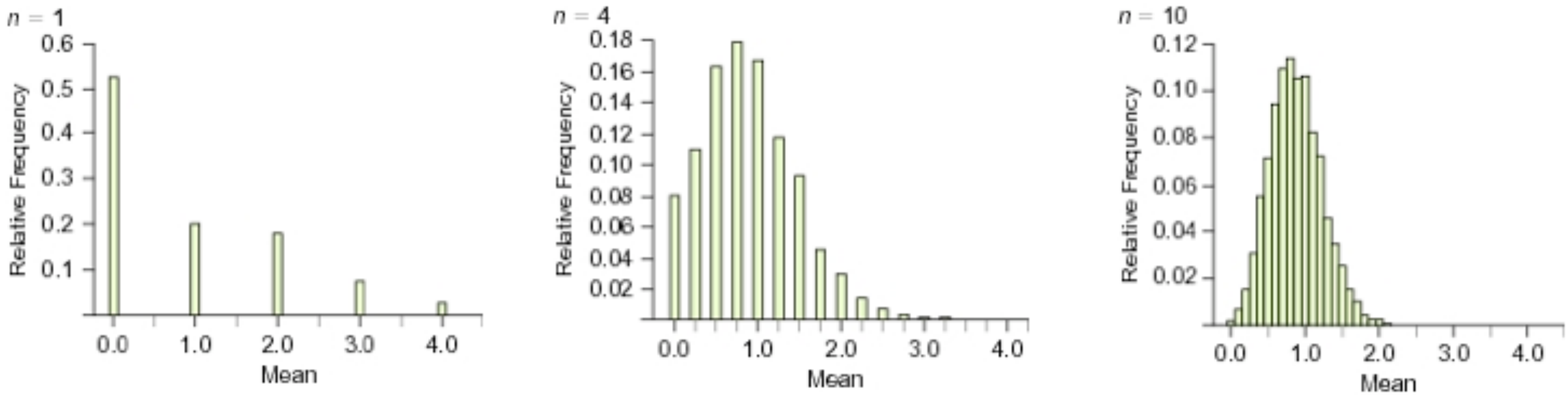
Math 140
Introductory Statistics

Next test November 18th

Cheating is forbidden

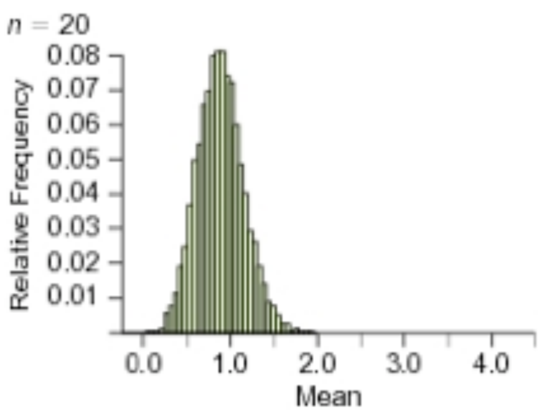
Using different sampling sizes n

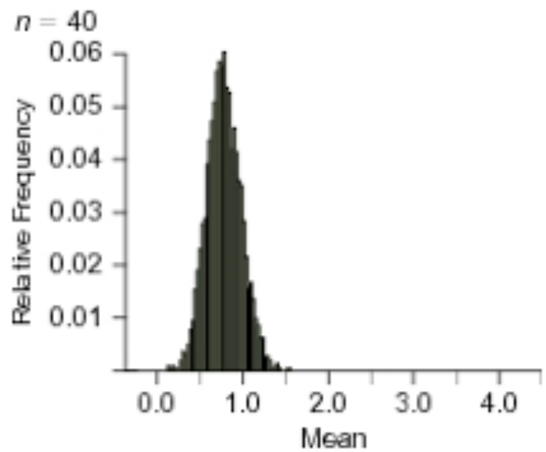


n=1, n=4, n=10

To find the average number of children in US households

Try with different sampling sizes n



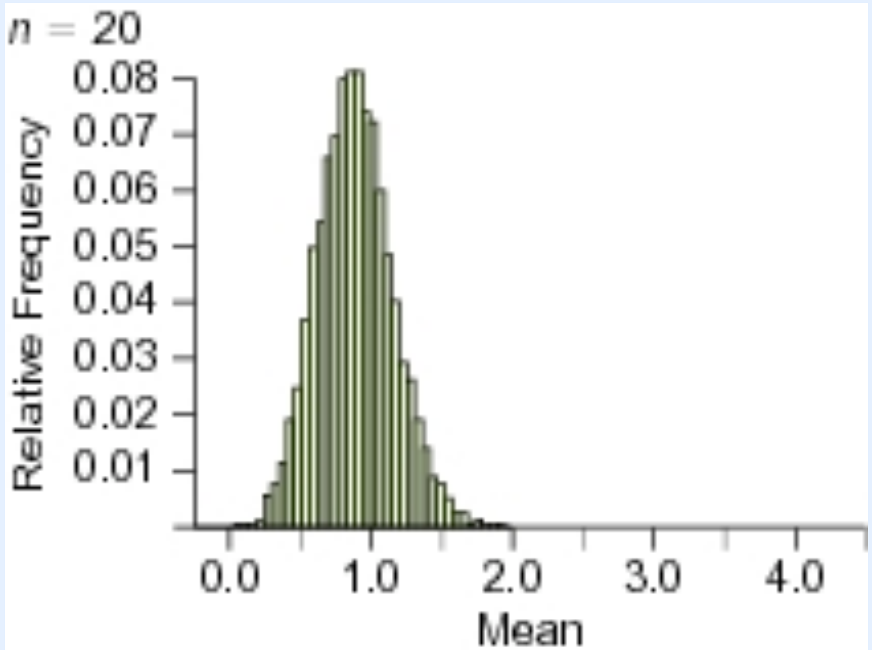


Sample Size, n	Mean	Standard Error, $\sigma_{\bar{x}}$
1	0.9	1.1
4	0.9	0.55
10	0.9	0.35
20	0.9	0.25
40	0.9	0.17
Population	0.9	1.1

n=20, n=10

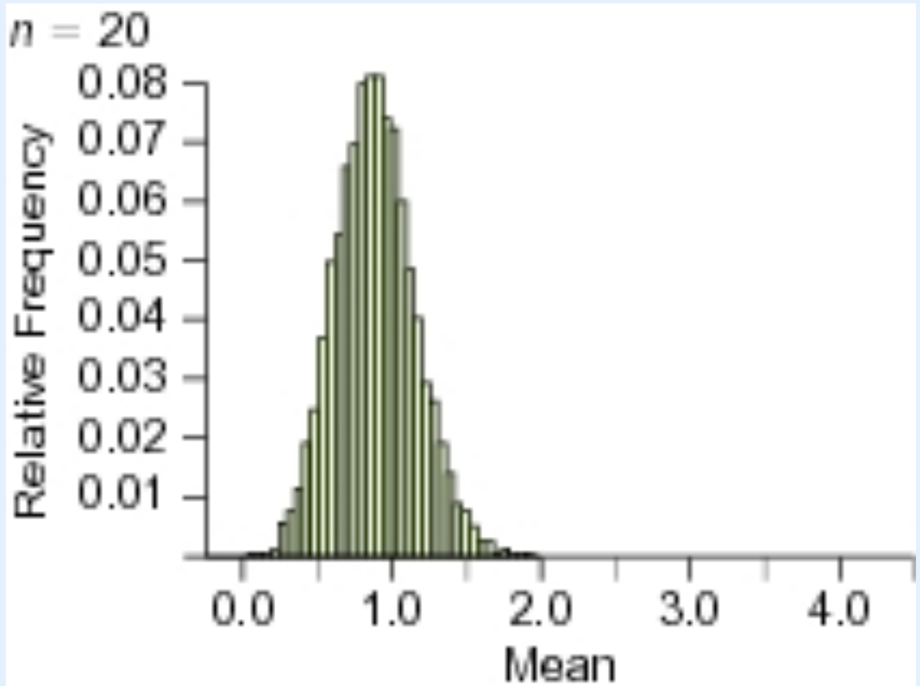
The mean is the same and
the standard error becomes smaller as n increases

In more detail, n=20



What is the probability that a random sample of 20 families will have an average of 1.5 children or less?

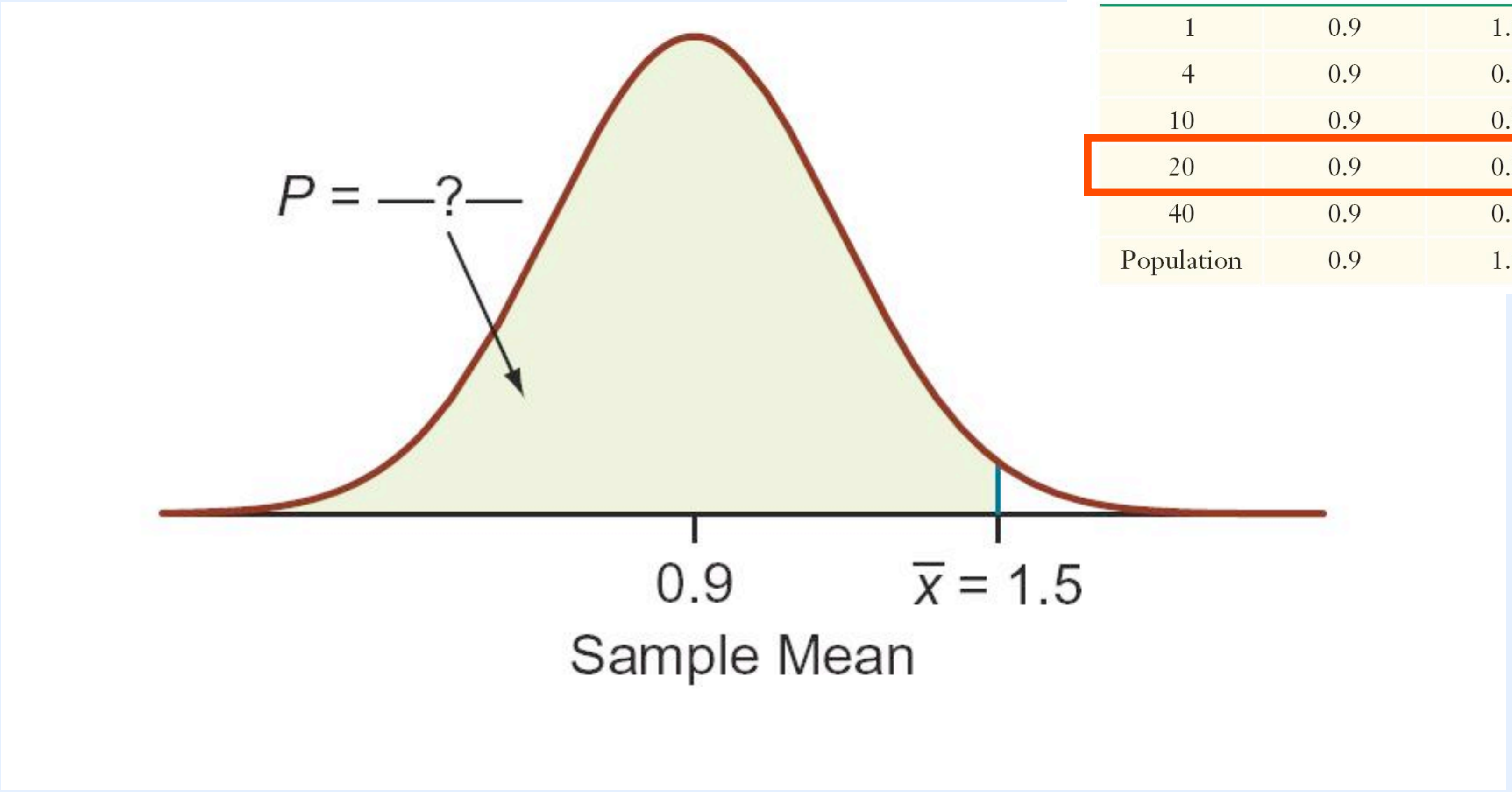
Children in US households



Well, it looks normal, so let's use what we know about normal distributions!

Center = 0.9, standard error = 0.25

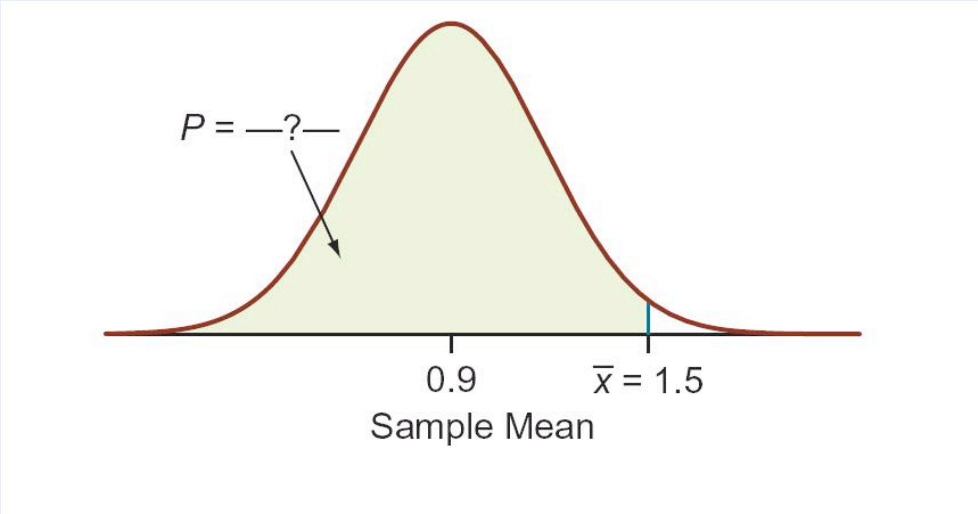
Children in US households



Sample Size, n	Mean	Standard Error, $\sigma_{\bar{x}}$
1	0.9	1.1
4	0.9	0.55
10	0.9	0.35
20	0.9	0.25
40	0.9	0.17
Population	0.9	1.1

What is the probability that a random sample of 20 families will have an average of 1.5 kids or less?

Recall rescaling and recentering?

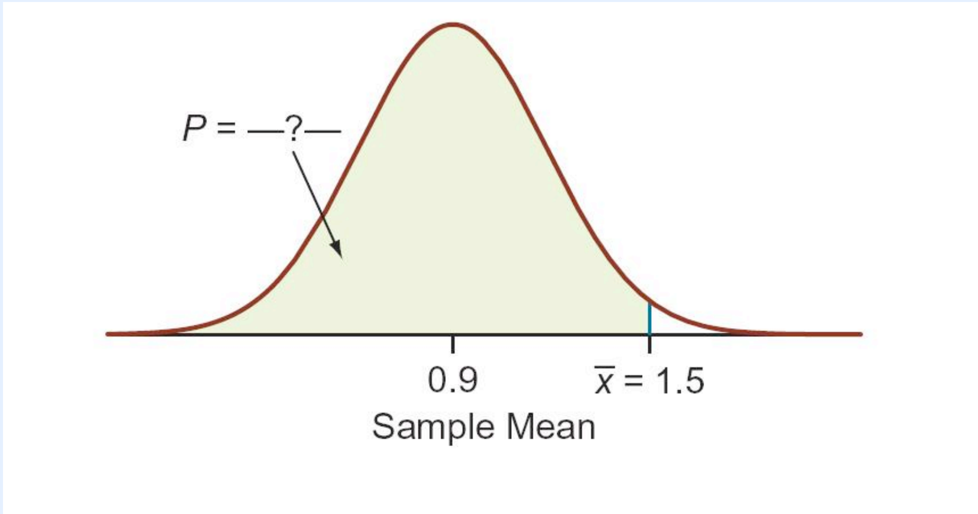


We need to calculate the corresponding z-score!

$$z = \frac{x - mean}{std\ deviation} = \frac{1.5 - 0.9}{0.25} = 2.4$$

The x we care for is 1.5. We need to find the corresponding z so we can check on page 759

Recall rescaling and recentering?



From Table A, page 759

The probability is 0.9918

For you

What is the probability that we find 30 children total
From a random sample of 20 families in the US?

For you

What is the probability that we find 30 kids total or less
From a random sample of 20 families in the US?

A little bit of a trick question.
This means that PER FAMILY
we find 30/20 kids=1.5

The answer is the problem we just did!

Corn yields in Indiana

Let's select random plots of the size 1/1000 of an acre.
This is about 18 feet of 1 row of corn.

On average each plot has a yield of about 15,000 kernels of corn with SD 2,000. This is the POPULATION DATA

Suppose we decide to randomly sample n=25 plots

What is the probability that the mean number of kernels will exceed 16,000?

What are reasonably likely outcomes for the mean yield of these n=25 sample plots (95% interval)?

Corn yields in Indiana

We expect the mean of the sample to be …

We expect the standard error of the sample to be …

Then take this info to Table A

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Corn yields in Indiana

We expect the mean of the sample to be 15,000

We expect the standard error of the sample to be

$$\sigma = \frac{2000}{\sqrt{25}} = \frac{2000}{5} = 400$$

The z-score is?

Corn yields in Indiana

Mean = 15,000

$$\sigma = \frac{2000}{\sqrt{25}} = \frac{2000}{5} = 400$$

The z-score is

$$z = \frac{x - 15,000}{400} = \frac{16,000 - 15,000}{400} = 2.5$$

From table A the probability is 1-0.9938 = 0.0062

Corn yields in Indiana

Reasonably likely values are between

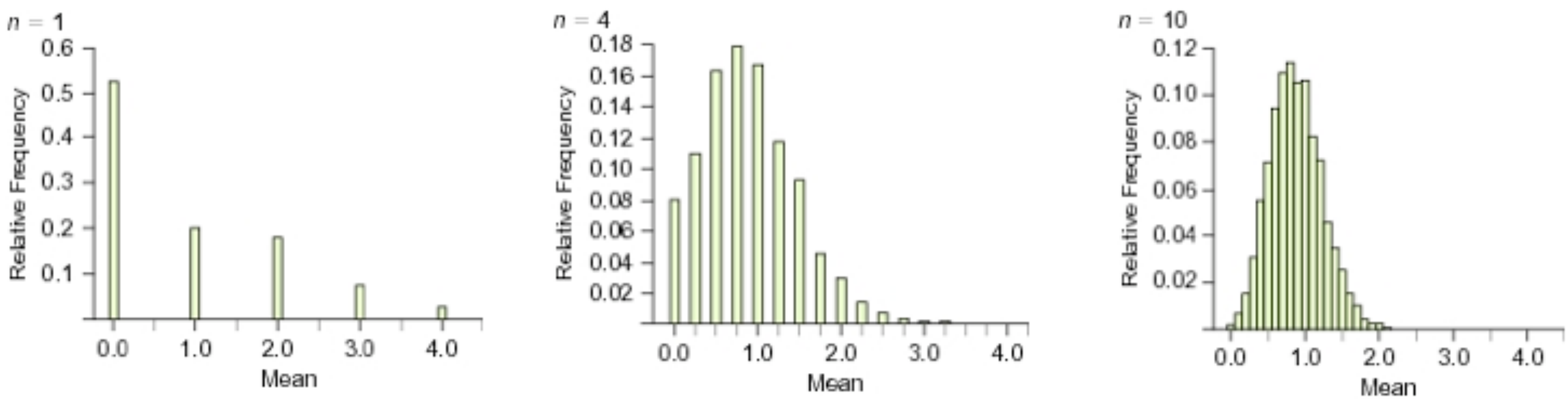
15,000 - 1.96*400 = 15,000 - 794 = 14,216

And

15,000 + 1.96*400 = 15,000 + 794 = 15,784

Can this always be done?

We need to make sure the sampling distribution is normally distributed



For example, n=1, n=4 are not really normal
n=10 begins to look normal

1. We can use z-scores only if the random sample distribution is normal

So for the case of the NBA
players we should be careful,
because the distribution was skewed

If the underlying population is itself normally distributed then we can use z-scores.

In general

When Can the Sampling Distribution of the Mean Be Considered Approximately Normal?

- If the population is approximately normally distributed, you can assume that the sampling distribution of $\bar{x}$ is approximately normal too, no matter what the sample size.
- If the sample size is 40 or more, it's pretty safe to assume that the sampling distribution of $\bar{x}$ is approximately normal.
- Using the normal approximation may be reasonably accurate with smaller sample sizes if the population isn't too badly skewed.

Also

2. The mean of the random sampling distribution is the same as the mean of the population

ALWAYS!

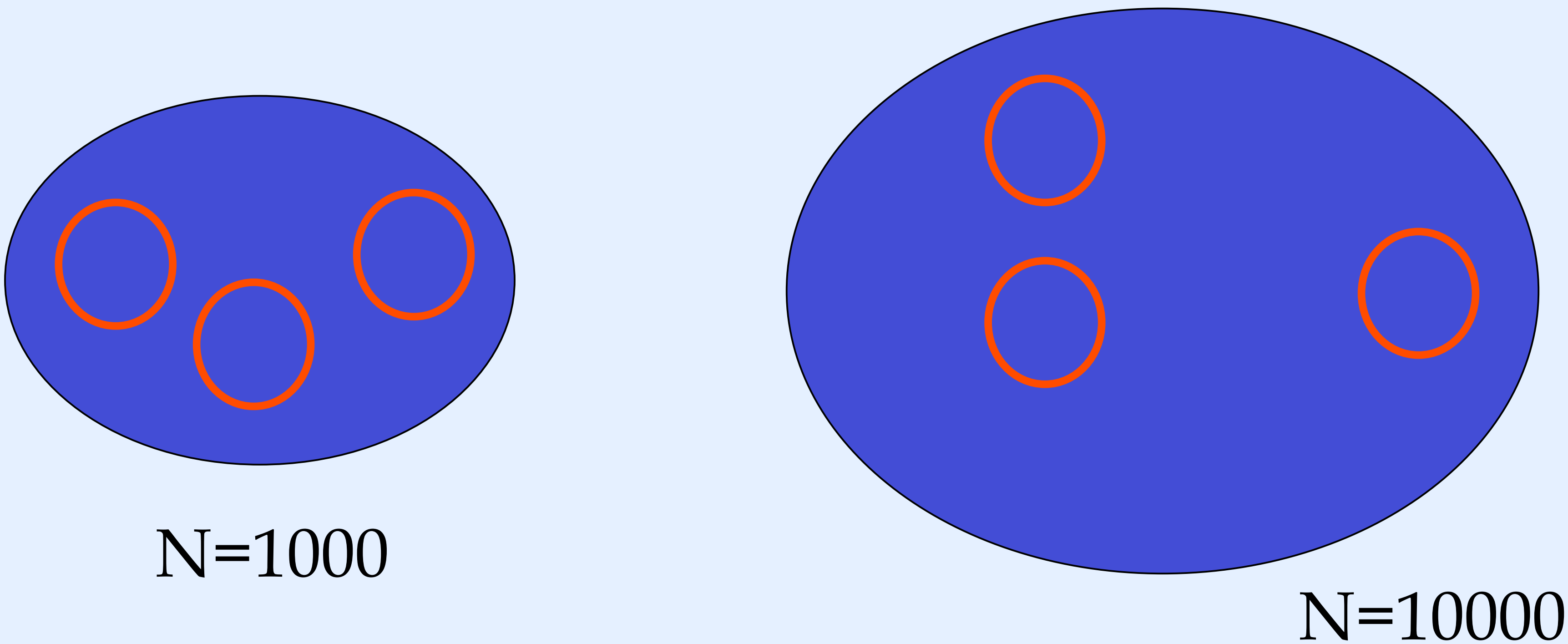
3. This too we can use almost always, only very rare exceptional cases when not appropriate

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Questions

Using the Properties of the Sampling Distribution of the Mean

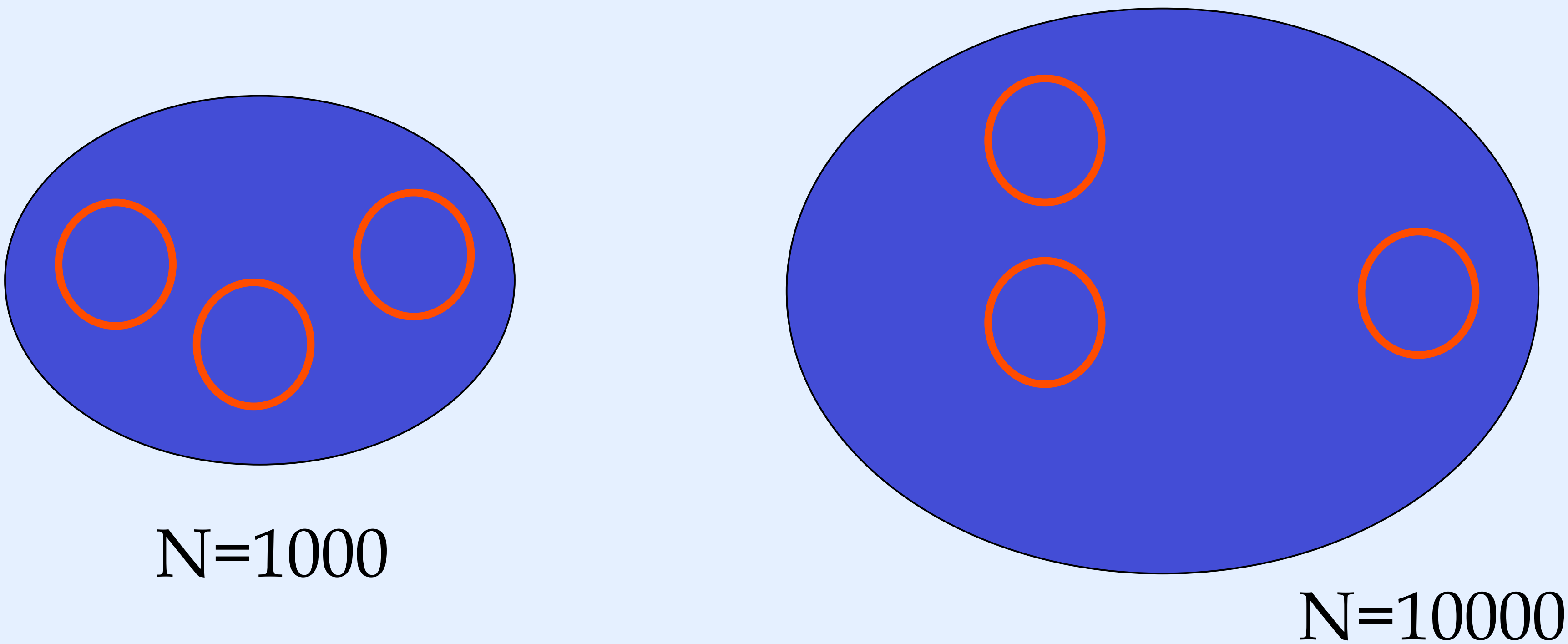
If you select 100 households at random, would the standard error of the sampling distribution of the mean number of children be larger if the population size, *N*, is 1000 or if it is 10,000?



Questions

Using the Properties of the Sampling Distribution of the Mean

No! Results will be approximately the same



Election time

What is the proportion of people that approve of Mr. Obama's job?

What is the proportion of CSUN students that voted?

What is the proportion of US households that have a goldfish?

Sample PROPORTIONS
Sample MEANS

Cellphones - according to the UN

4.1 billion subscribe to cell phone coverage

About 60% of people on the planet have a cell phone

$p = 60\%$ or $p = 0.60$ is the
population proportion of success

If we take a random sample of 40 people,
What is the probability that 75% or more of them have a cell phone?

Introduce a new variable

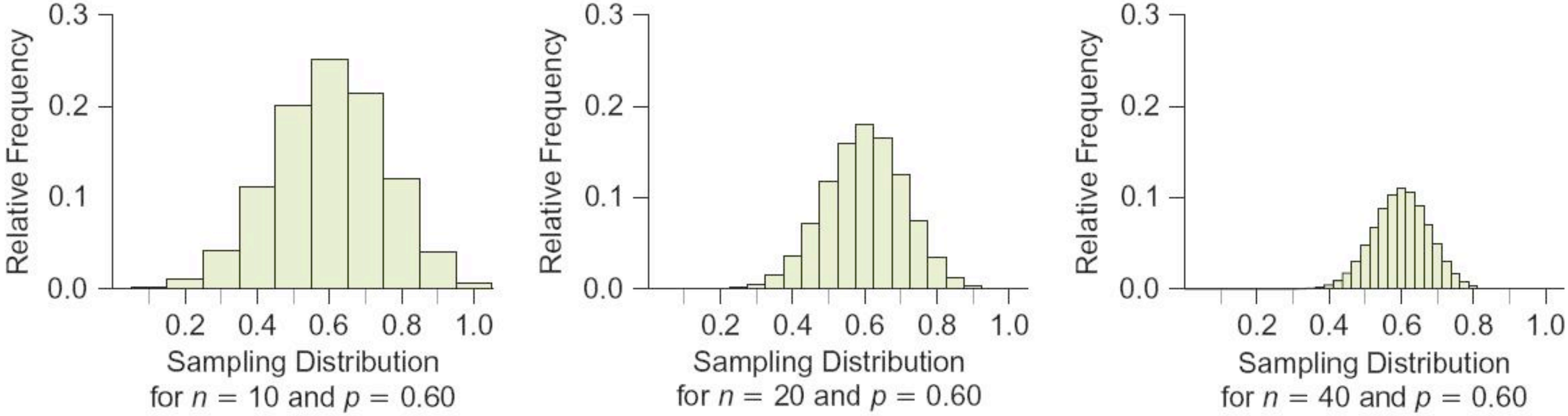
$$\hat{p} = \frac{number\ of\ successes}{sample\ size}$$

This is the sample proportion of success
For example, we poll 40 people at random
and 26 of them have a cell phone

In this case $\hat{p} = 26/40 = 0.65$

Repeat many times and create a
SAMPLE DISTRIBUTION FOR $\hat{p}$

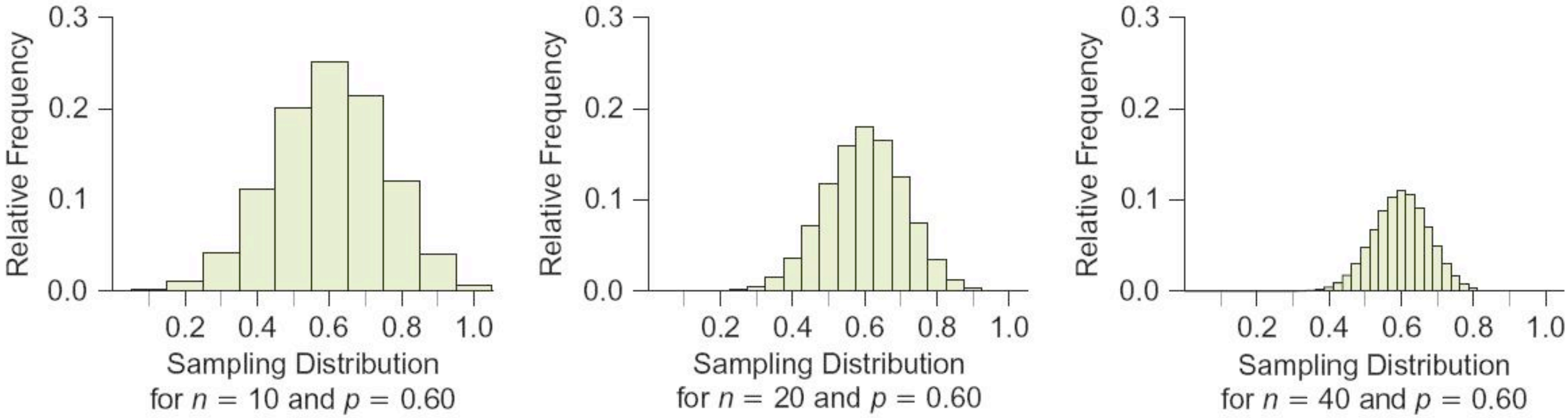
Sampling distributions for $\hat{p}$



Display 7.42 Exact sampling distributions of $\hat{p}$ for samples of size 10, 20, and 40 when $p = 0.60$.

Every time we pick our group of n=40,
we calculate $\hat{p}$
and create a sampling distribution

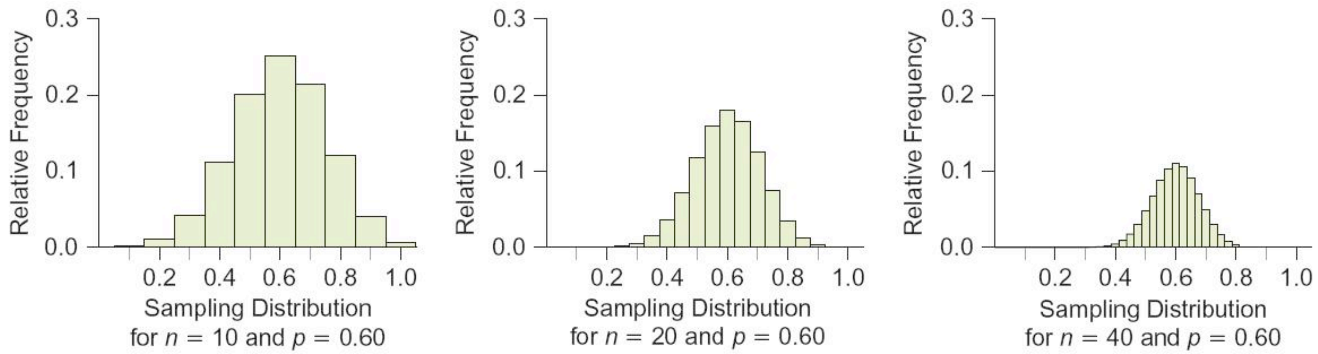
Sampling distributions for $\hat{p}$



Display 7.42 Exact sampling distributions of $\hat{p}$ for samples of size 10, 20, and 40 when $p = 0.60$.

What do you observe?

Sampling distributions for $\hat{p}$



Display 7.42 Exact sampling distributions of $\hat{p}$ for samples of size 10, 20, and 40 when $p = 0.60$.

As n becomes bigger the shape becomes more approximately normal

The mean does not change, and is centered about 0.60

The spread becomes smaller

VERY SIMILAR TO WHAT WE DID FOR THE MEAN!

Let us think small

N is my population (N=100),
p is the success rate (p = 60%)

People that have cell phones = ?
People that don't have cell phones = ?

Let us think small

N is my population (N=100),
p is the success rate (p = 60%)

People that have cell phones = 60 = p*N

People that don’t have cell phones = 40 =
N - p*N = (1-p)*N

Let us think small

If I have a cell phone my success rate is 1
If I don't my success is 0

On average the success is 0.60

What is the SD of this success rate?

Let us think small

Recall, SD

take values
subract from mean and square
Add over all values
Divide by data points
Take root

Here I have N data points (my population)
N*p have a success of 1
N*(1-p) have a success of 0
The mean is p

Let us think small

Here I have N data points (my population)
N*p have a success of 1
N*(1-p) have a success of 0
The mean is p

So, for N*p people (60 in the cell phone case)
the contribution to the SD is
$(1-p)^2$

So, for N*(1-p) people (40 in the cell phone case)
the contribution to the SD is
$(0-p)^2$

When we add them all together:

N*p (1-p)^{2} + N*(1-p) (0-p)^{2}

N*p (1-p)^{2} + N*(1-p) p^{2}

Factor out N*p (1-p) *(1-p+p)

N*p (1-p)

Divide by data N and take square root

$$\sigma = \sqrt{\frac{Np(1-p)}{N}} = \sqrt{p(1-p)}$$

So, the SD of the population is

$$\sigma = \sqrt{p(1-p)}$$

The SD of the n sample (Std error) is

$$\sigma_n = \frac{\sqrt{p(1-p)}}{\sqrt{n}} = \sqrt{\frac{p(1-p)}{n}}$$

Properties of the Sampling Distribution of the Sample Proportion, $\hat{p}$

If a random sample of size n is selected from a population with proportion of successes p, then the sampling distribution of $\hat{p}$ has these properties:

- The mean, $\mu_{\hat{p}}$, is equal to the proportion, p, of successes in the population, or $\mu_{\hat{p}} = p$.
- The standard error is equal to the standard deviation of the population divided by the square root of the sample size:

$$\sigma_{\hat{p}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{p(1-p)}}{\sqrt{n}} = \sqrt{\frac{p(1-p)}{n}}$$

- As the sample size gets larger, the shape becomes more normal and will be approximately normal if n is large enough.

As a guideline, if both np and $n(1-p)$ are at least 10, then using the normal distribution as an approximation of the shape of the sampling distribution will give reasonably accurate results.

East Coast (bad) drivers

20% failed on their tests

Study the sampling distribution of drivers that would fail if we take n=60 samples

The p of failure is 0.2

$$\sigma = \sqrt{p(1-p)} = \sqrt{0.2 * 0.8} = \sqrt{0.16} = 0.4$$

$$\sigma_p = \frac{\sqrt{p(1-p)}}{\sqrt{n}} = \frac{0.4}{\sqrt{60}} = 0.05$$

East Coast (bad) drivers

Reasonably likely proportions are between

0.2 - 1.96*0.05 and 0.2 + 1.96*0.05

About 0.1 and 0.3

Back to cell phones

60% of people have a cell phone.

1. What is the chance that 75% or more people selected from an n=40 sample have a cell phone?

2. Would it be unusual to find that among those 40 less than 10 people have a cell phone?

Back to cell phones - part 1

Here p = 0.6, 1-p = 0.4

$$\sigma_p = \frac{\sqrt{p(1-p)}}{\sqrt{n}} = \sqrt{\frac{0.6 * 0.4}{40}} = 0.0775$$

$$z = \frac{\hat{p} - p}{\sigma_p} = \frac{0.75 - 0.60}{0.0775} = 1.935$$

From table A the probability that
75% people or more will have a cell phone
is 1-0.9735 = 0.0265

Back to cell phones - part 2

Let’s calculate the z score for the number of success to be 10

$$\hat{p} = \frac{10}{40} = 0.25$$

Most likely values
0.6 - 1.96*0.0775 and 0.6 + 1.96*0.0775 =
0.441 and 0.752

The above value of 0.25 is extremely unlikely

Homework

Page 349 P16, P17, P18
E35, E36, E40, E41, E42, E43, E45,