

# DATA 500

## Lecture 9

### Estimating the population mean

### The Central Limit Theorem

# Parameters and Statistics

As we begin to use sample data to draw conclusions about a wider population, we must be clear about whether a number describes a sample or a population.

A **parameter** is a number that describes some **characteristic of the population**. In statistical practice, **the value of a parameter is not known** because we cannot examine the entire population.

A **statistic** is a number that describes some **characteristic of a sample**. The value of a statistic can be computed directly from the sample data. **We often use a statistic to estimate an unknown parameter.**

Remember **s** and **p**: **s**tatistics come from **s**amples and **p**arameters come from **p**opulations.

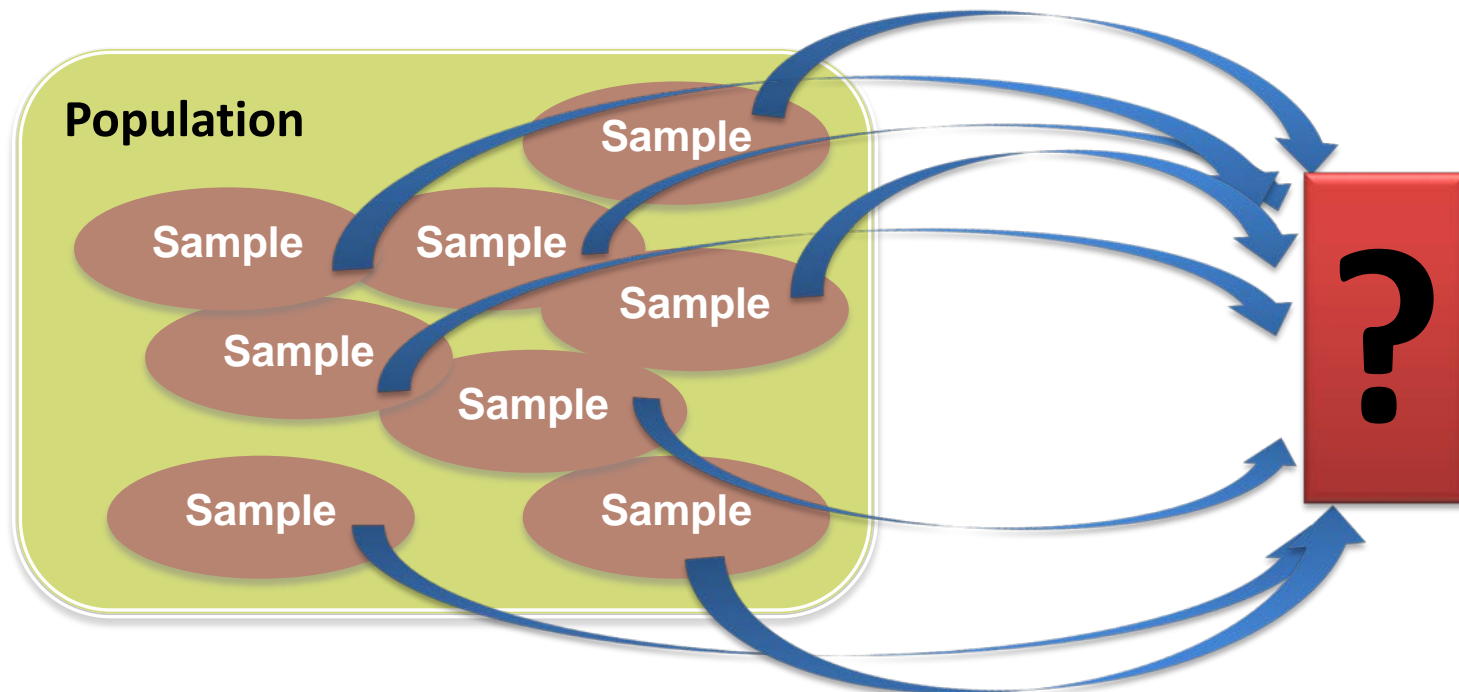
We write  $\mu$  for the population mean and  $\sigma$  for the population standard deviation.

<sup>2</sup> We write  $\bar{x}$  for the sample mean and  $s$  for the sample standard deviation.

# Sampling Variability

Different random samples yield different statistics. This basic fact is called **sampling variability**: **The value of a statistic varies in repeated random sampling.**

To make sense of sampling variability, we ask, “What would happen if we took many samples?”



# The Law of Large Numbers



How can  $\bar{x}$  be an accurate estimate of  $\mu$ ? After all, different random samples would produce different values of  $\bar{x}$ .

If we keep on taking larger and larger samples, the statistic  $\bar{x}$  is guaranteed to get closer and closer to the parameter  $\mu$ .

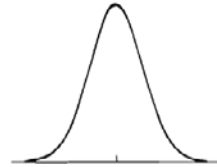
Draw independent observations at random from any population with finite mean  $\mu$ . The **law of large numbers** says that, **as sample size increases**, the sample mean of the observed values gets closer and closer to the mean  $\mu$  of the population.

**Once again, larger samples are better!**

# Sampling Distributions

The **law of large numbers** assures us that if we measure enough subjects,  $\bar{x}$  will eventually get very close to the unknown parameter  $\mu$ .

If we took every one of the possible samples of a certain size, calculated the sample mean for each ( $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \bar{x}_5, \bar{x}_6, \bar{x}_7, \dots$ ), and plotted the distribution of those values, we'd have a **sampling distribution**.



The **population distribution** of a variable is the distribution of **individual values** in the population. (**standard deviation**)

The **sampling distribution** of a statistic is the distribution of values taken by the statistic in all possible samples of the same size from the same population. ( $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \bar{x}_5, \bar{x}_6, \bar{x}_7, \dots$ )  
(**standard error of the mean**)

# Mean and Standard Error of the Sampling Distribution

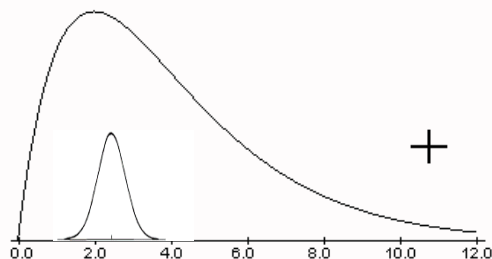
## The Mean of the sampling distribution (sample means)

There is no tendency for a sample mean to fall systematically above or below  $\mu$ , *even if the distribution of the raw data are skewed*. Thus, **the mean of the sampling distribution is an unbiased estimate of the population mean  $\mu$** .

## Standard error of sample means (standard deviation of means)

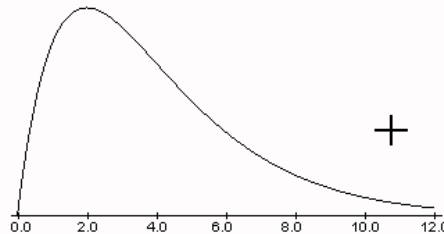
The **standard error of the sampling mean** measures how much the sample statistic varies from sample to sample. It is smaller than the **standard deviation of the population** by a factor of the square root of  $n$ .

➔ **Sample means are less variable than individual observations.**



# The Central Limit Theorem

What is the shape of the sampling distribution of sample means when the population distribution isn't Normal?

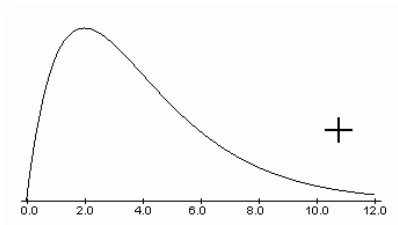


**Guess what?** *As the sample size increases, the distribution of sample means begins to look more and more like a Normal distribution!*

When the sample is large enough, **the distribution of sample means is very close to Normal, no matter what shape the population distribution has.**

# The Central Limit Theorem

The CLT involves an SRS (simple random sample)



Draw an SRS of size  $n$  from any population with mean  $\mu$  and finite standard deviation  $\sigma$ . The **central limit theorem (CLT)** says that when  $n$  is large, the sampling distribution of the sample mean  $\bar{x}$  is approximately Normal:

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



# Repeated samples and the SEM

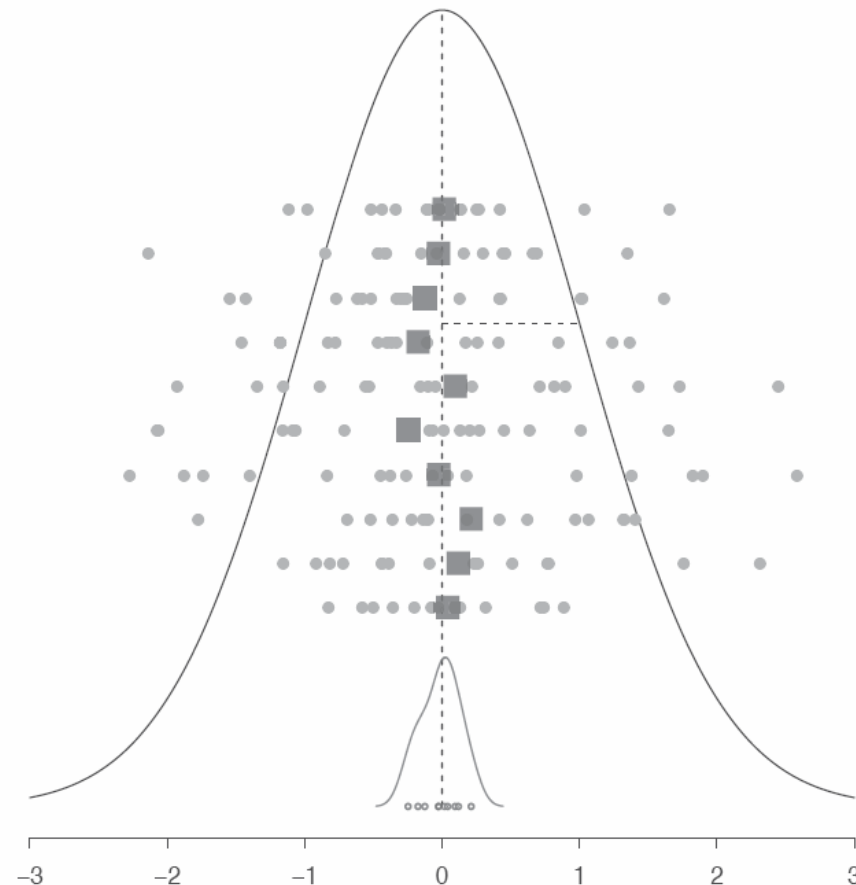
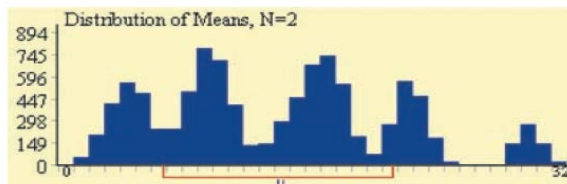
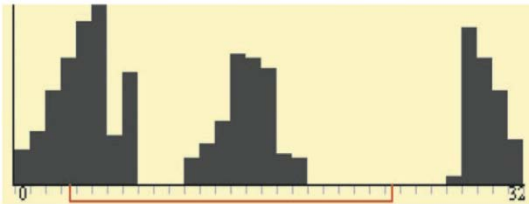
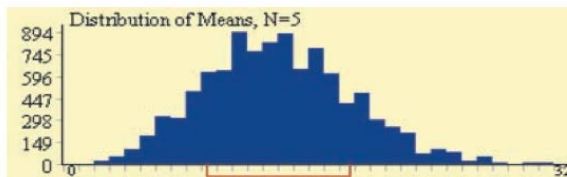


Figure 7.1: A population from which 10 samples of size 16 are chosen. Each sample is summarized by its mean (the boxes). These 10 means are summarized with the scaled, inset density estimate. This density is bell shaped, centered on the population mean, but has smaller spread.

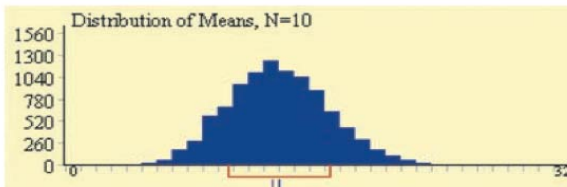
# A few more facts about the CLT



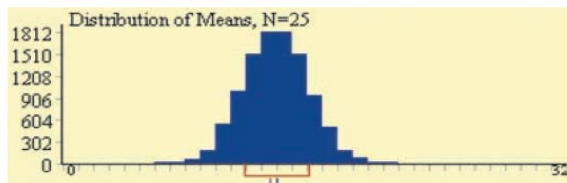
(a)



(b)



(c)



(d)

The central limit theorem notes that the distribution of a sum or mean of many small random quantities is close to Normal.

The central limit theorem also applies to discrete random variables. (heads/tails)

# The Central Limit Theorem

Sampling distributions of the estimated mean approach a normal distribution when the sample size is large, regardless of the shape of the parent population.

In other words, if you take multiple samples from the same population, the estimated means will show a normal distribution

So we can estimate the population mean from a sample ( $s$  is our **point estimate** of  $\mu$ ) and can quantify our uncertainty about it using the standard error

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

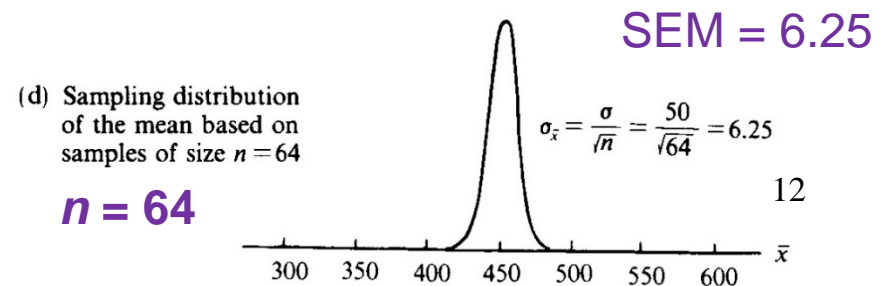
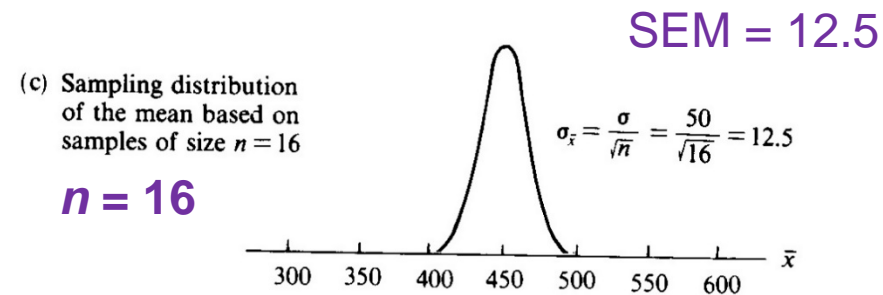
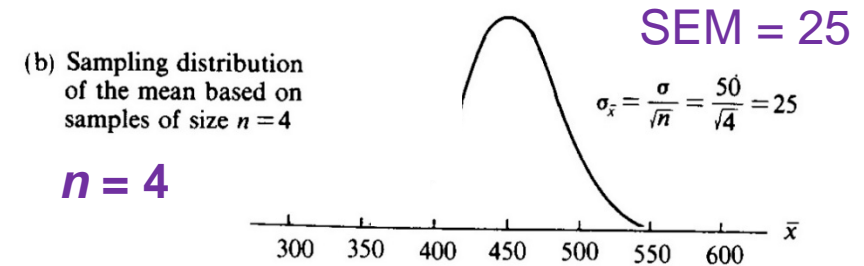
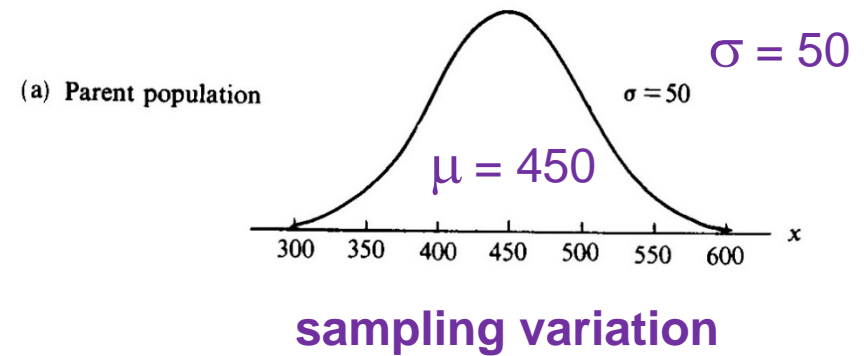
# Sample size and sampling error (the standard error, or SEM)

- affects precision of estimate of underlying ("true") mean

$$SEM = \frac{s}{\sqrt{n}}$$

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

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



# Sample size, $s$ , and estimates of $\mu$

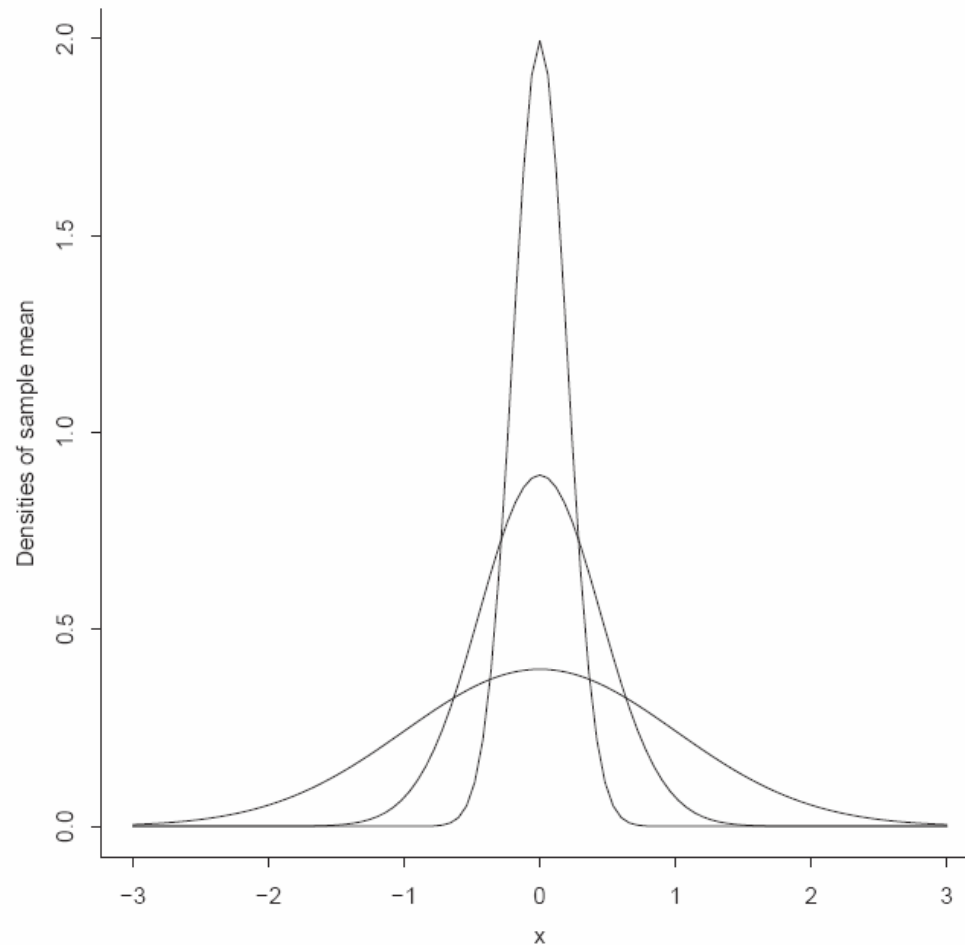
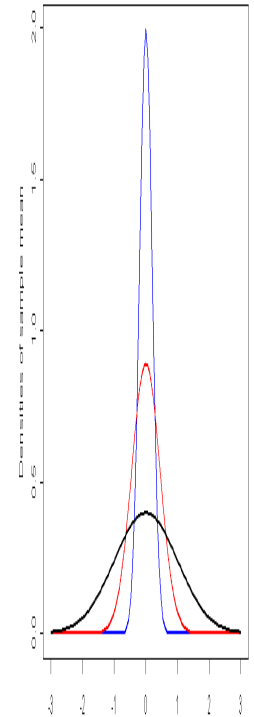
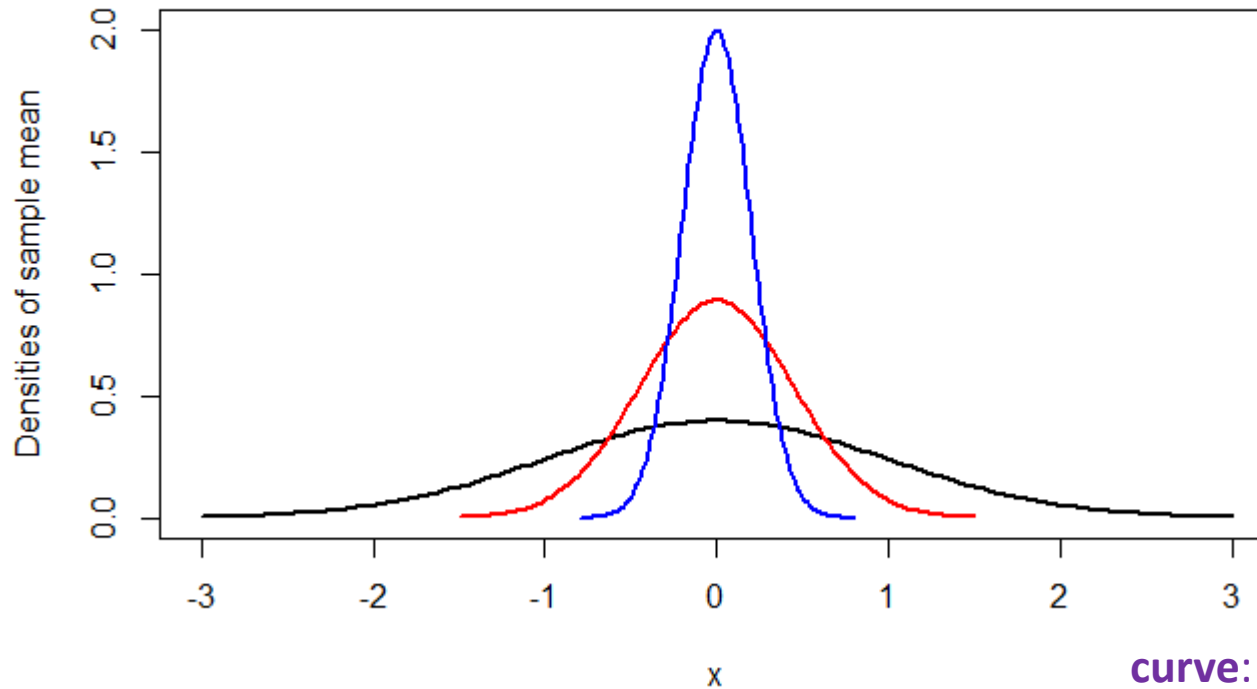


Figure 6.9: Density of  $\bar{X}$  for  $n = 5$  and  $n = 25$  along with parent population  $\text{Normal}(0,1)$ . As  $n$  increases, the density concentrates on  $\mu$ .

# Sample size, $s$ , and estimates of $\mu$

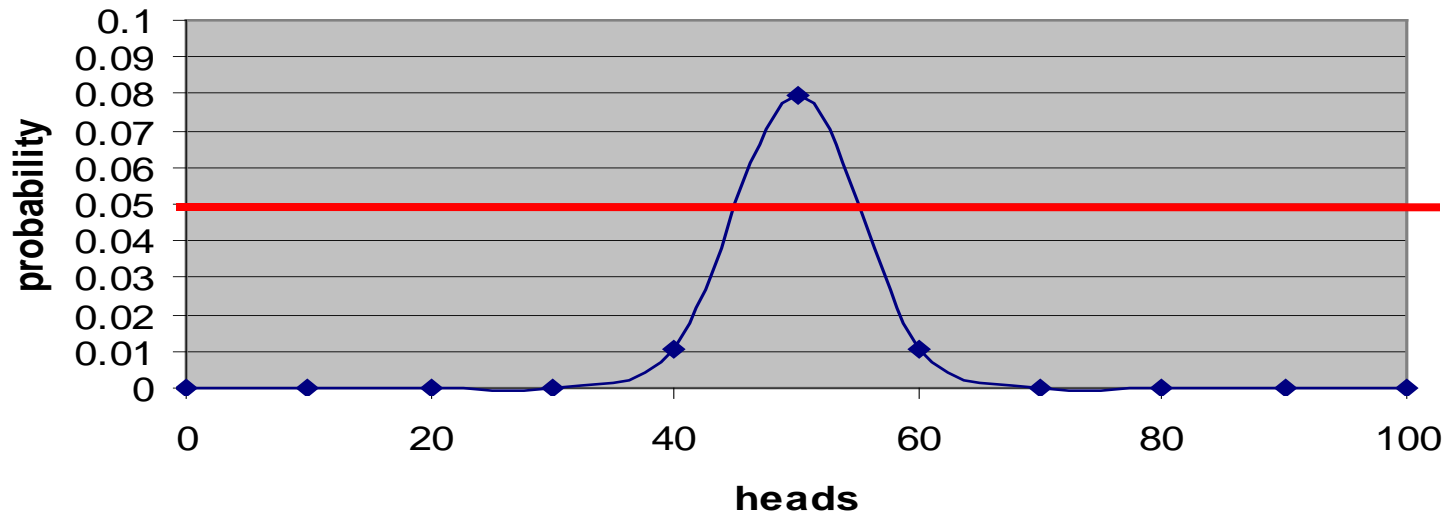
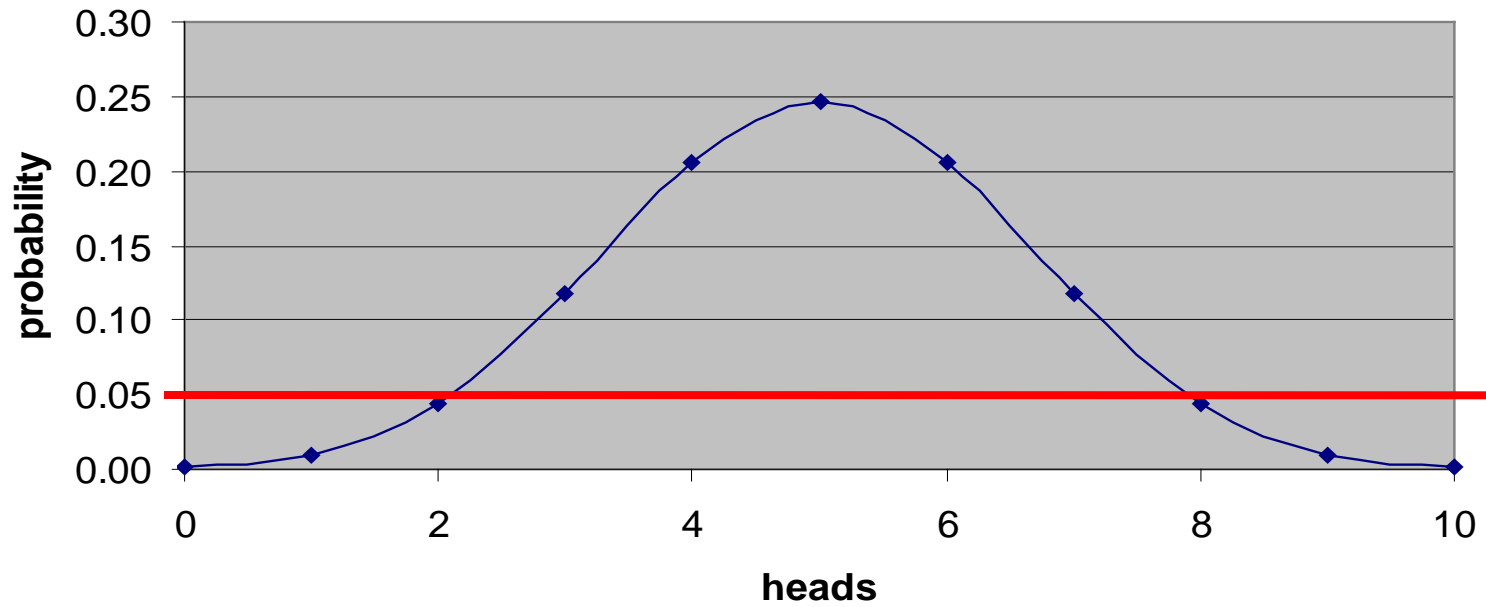
Sampling Distribution by sample size: black = 1 (SE); red = 5; blue = 25



**curve:** draw a continuous line  
**lwd:** line width  
**col:** color of line

```
# draw a continuous curve; # black is default color ;  
n <- 2;  
curve(dnorm(x, mean=0, sd=1), -3, 3, lwd=2, xlab="x", ylab="Densities of sample mean", ylim =  
c(0,2), main = 'Sampling Distribution by sample size: black = 1 (SE); red = 5; blue = 25; ');  
n <- 5;  
curve(dnorm(x, mean=0, sd=1/sqrt(n)), -1.5, 1.5, lw = 2, col="red", add=T);  
n <- 25;  
curve(dnorm(x, mean=0, sd=1/sqrt(n)), -0.8, 0.8, add=TRUE, #min, max;  
lwd = 2, col="blue" );
```

**A larger sample size will provide a better estimate of the population mean**



# Generating Distributions:

## `norm` (normal) `unif` (uniform)

The `r`-prefixed functions generate random samples from a distribution (simulated data)

`rnorm`(sample size,mu,sigma)

```
rnorm(1000, 0, 1) # 1000 values with the standard normal dist.
```

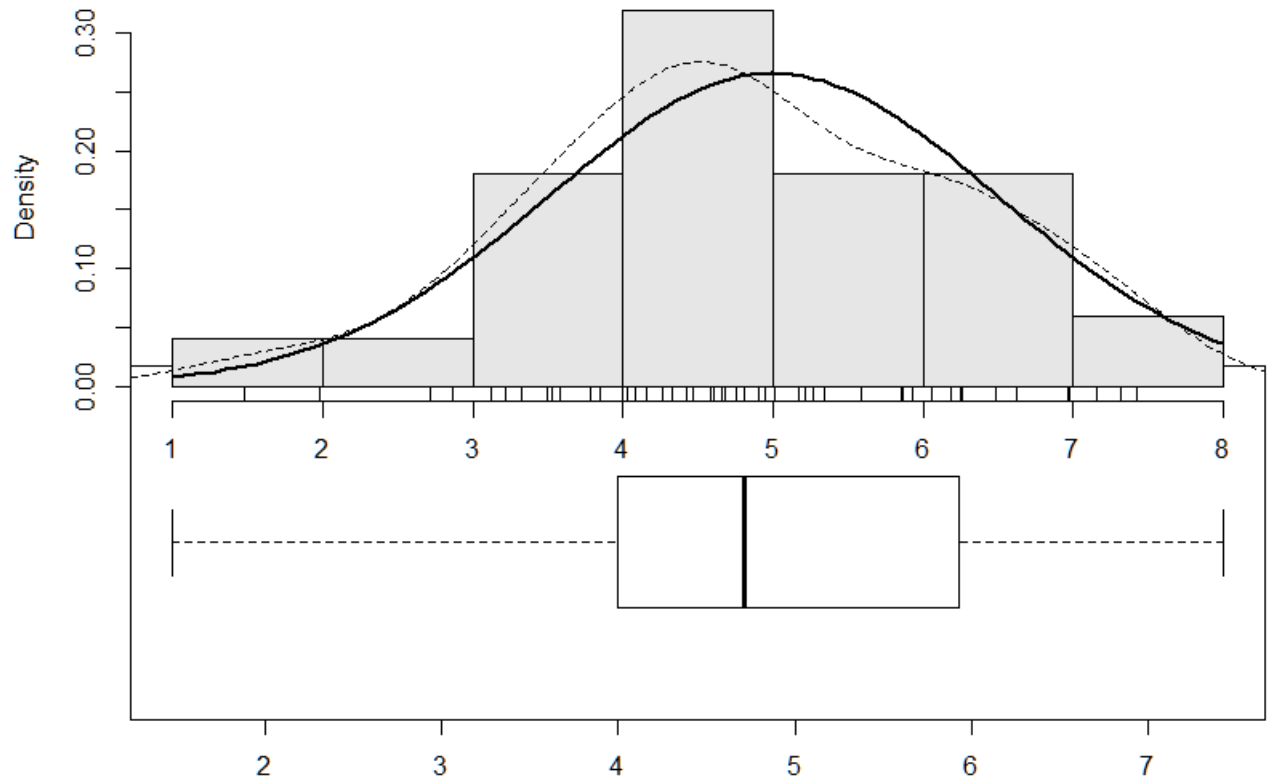
```
[1] -0.4977155242  0.3033878261 -0.9957592175 -0.7123912704 -1.0769687830
[6]  1.3419767315 -0.3820969494  1.5714187853 -1.6720294931  0.7684332130
[11] -0.0528299924 -0.6125091898 -0.4992395170 -0.2834088364 -0.3512385199
[16] -0.0752119177  1.8875435053  0.6819433616  0.3019784925  0.0754323090
[21] -0.5565305978  1.0123568885  0.8308080903  0.0409428125  0.4913253717
```

...

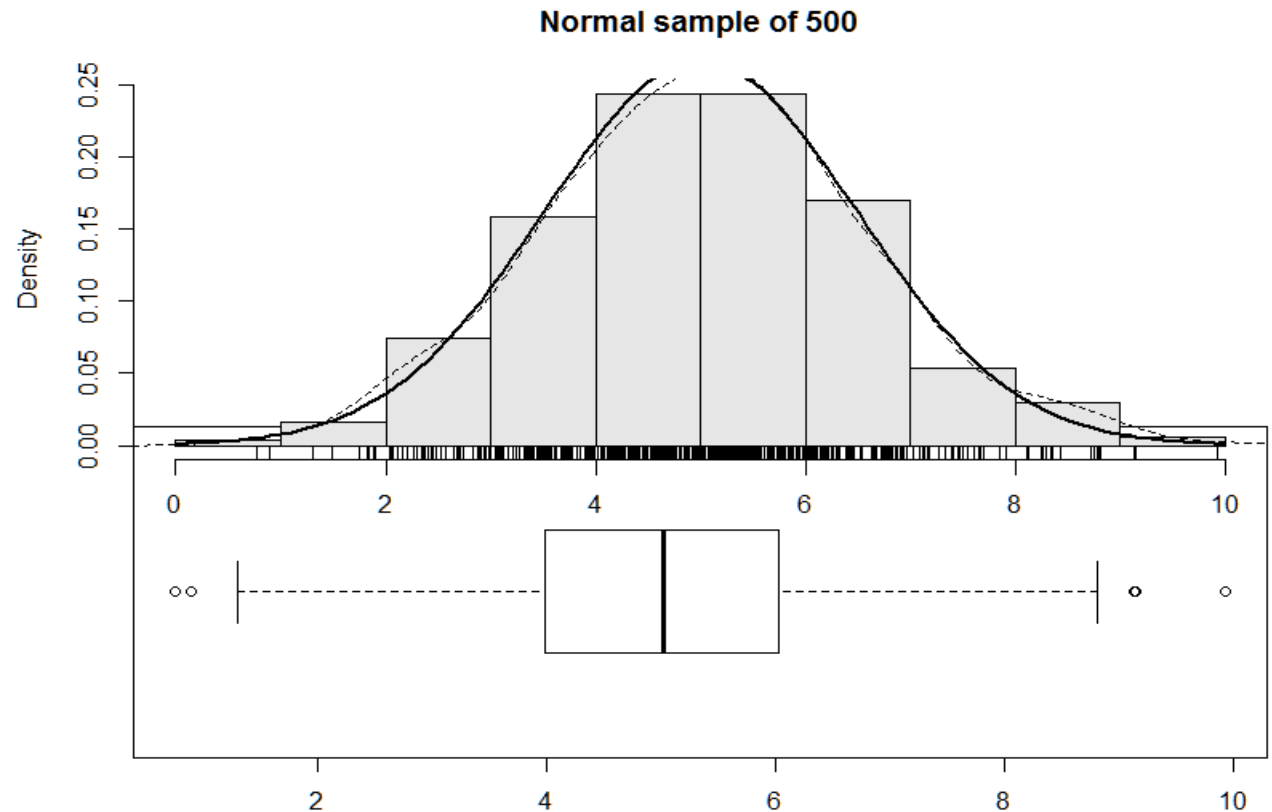
```
[966]  0.1198891010 -1.7024513393  2.1762204747  1.5456761101  0.7914776548
[971] -0.4682414263 -1.2786616170 -0.2673307183  1.6355726608 -0.3652457954
[976] -0.3348244291 -0.1180272508 -0.4967495956 -0.2886845277  1.4448755748
[981] -0.4941943580  0.2084222562 -0.2338059368 -0.0858165442 -0.9353715496
[986] -2.0107239248 -0.3214571421 -1.9189423283 -0.0949458847 -0.8162877404
[991]  1.1693028597 -0.5520440646 -1.6131976373  0.0392985886 -1.8615985082
[996] -0.3097877808  0.0202657697 -0.2114900414  0.0403419063  1.0530774921
```



Normal sample of 50



```
# normal distribution;
# small sample
res <- rnorm(50, 5, 1.5);
## fig= setting uses bottom 40% of diagram;
par(fig=c(0,1,0,0.6));
boxplot(res,horizontal=TRUE, bty="n", xlab='');
## fig= setting uses top 70% of figure;
par(fig=c(0,1,0.3,.95), new=TRUE);
hist(res, prob=TRUE, main="Normal sample of 50", col=gray(.9), xlab = '');
lines(density(res),lty=2);
curve(dnorm(x, 5, 1.5), lwd=2, add=TRUE);
rug(res);
par(fig=c(0,1,0,1)); # reset plot dimensions
```



```
# normal distribution;
# small sample
res = rnorm(500, 5, 1.5);
## fig= setting uses bottom 40% of diagram;
par(fig=c(0,1,0,0.6));
boxplot(res, horizontal=TRUE, bty="n", xlab='');
## fig= setting uses top 70% of figure;
par(fig=c(0,1,0.3,.95), new=TRUE);
hist(res, prob=TRUE, main="Normal sample of 500", col=gray(.9), xlab = '');
lines(density(res), lty=2);
curve(dnorm(x, 5, 1.5), lwd=2, add=TRUE);
rug(res);
par(fig=c(0,1,0,1)); # reset plot dimensions
```

# The CLT and dice

## One die: Uniform distribution

Probability =  $1/6$

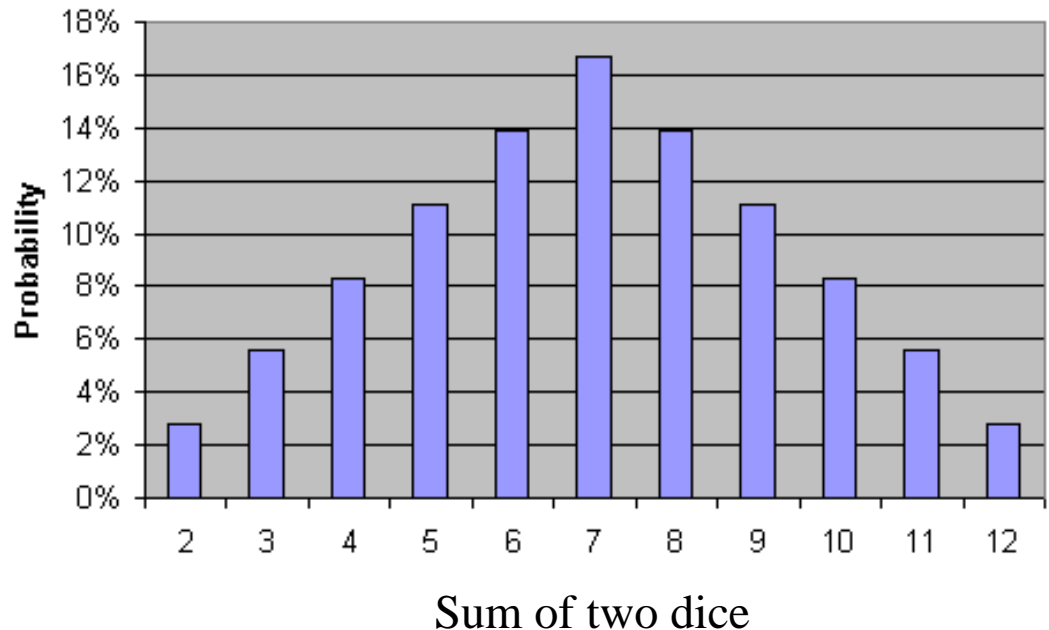
Mean = 3.5

## Sum of two dice:

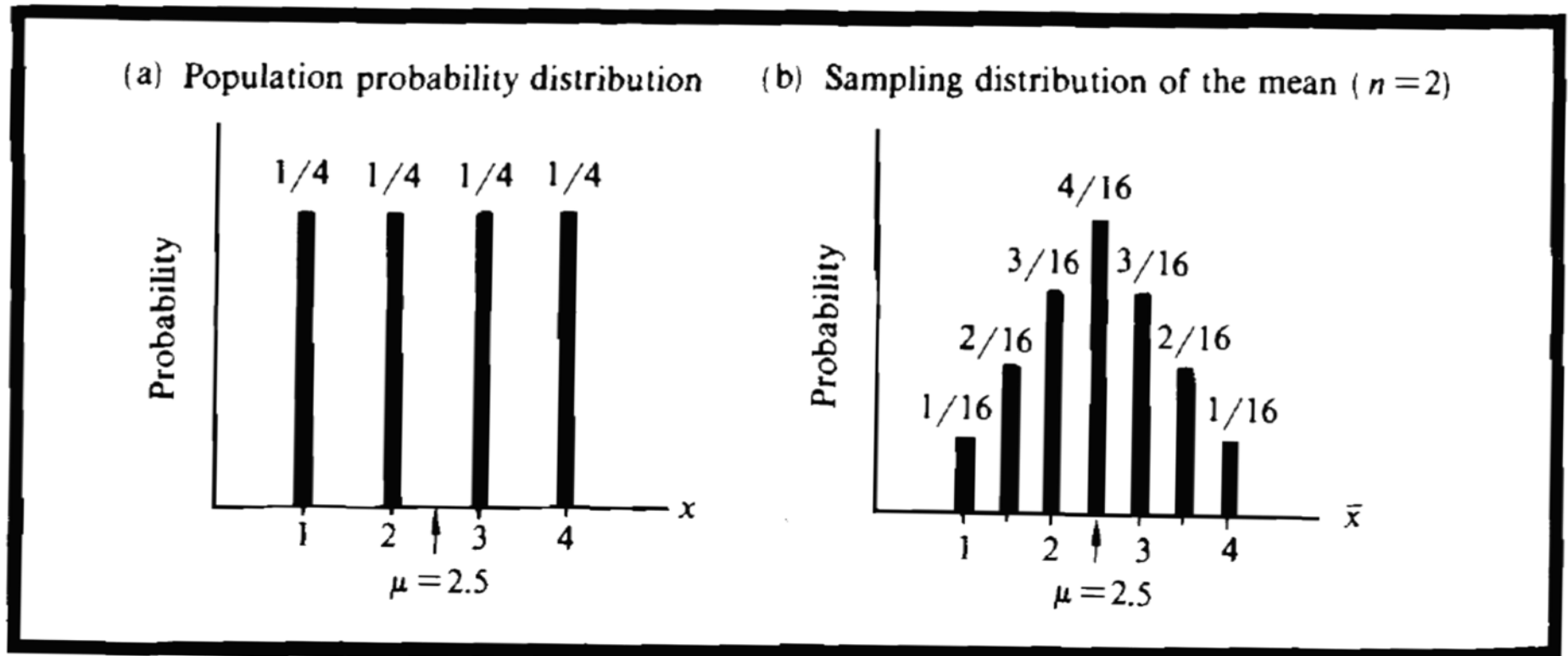
Looks close to normal

- same for summing coin flips

Mean = 7



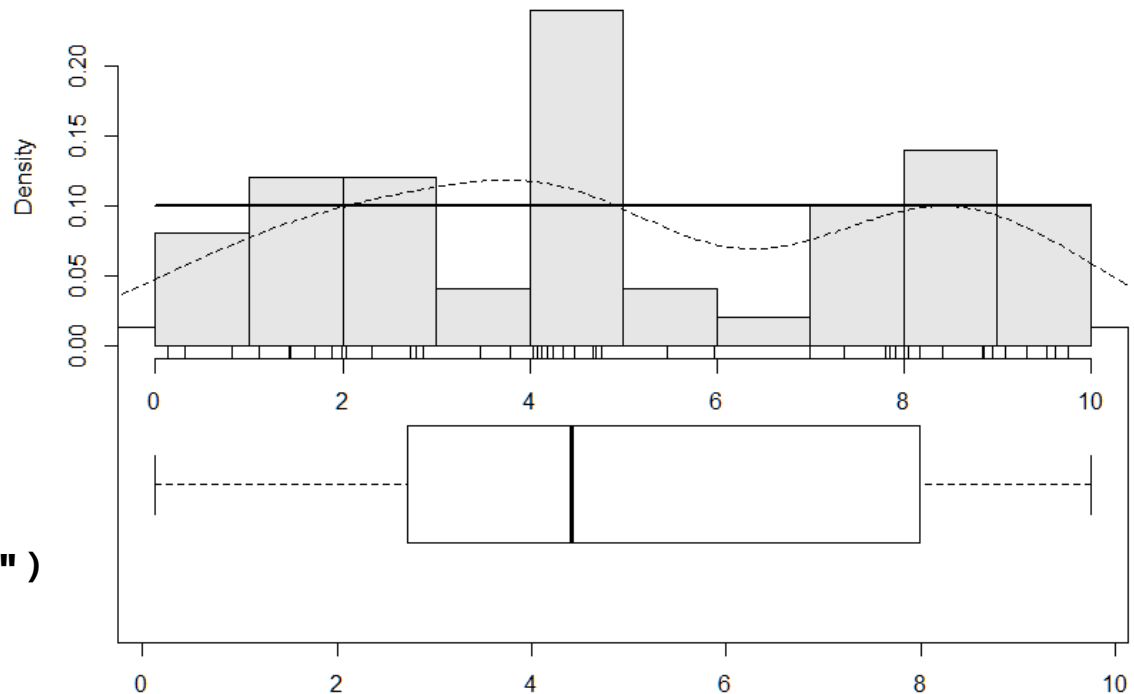
# The CLT and summing four possibilities



**Figure 14** The (a) parent population, and (b) sampling distribution of the mean based on a sample size of  $n=2$  (See Table 5).

# Plotting simulated distributions

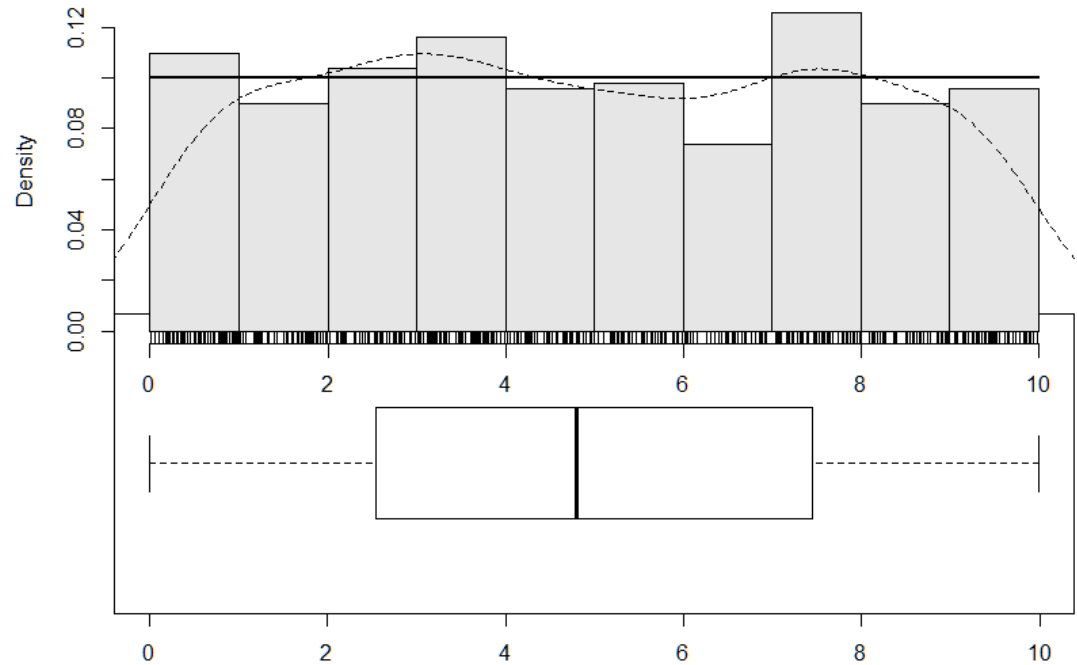
Uniform sample of 50



```
# install.packages("e1071")
library(e1071)
# uniform distribution;
res <- runif(50, min=0, max=10);
## fig= setting uses bottom 40% of diagram;
par(fig=c(0,1,0,0.6));
boxplot(res,horizontal=TRUE, bty="n", xlab='');
## fig= setting uses top 70% of figure;
par(fig=c(0,1,.30,0.95), new=TRUE);
hist(res,prob=TRUE,main="Uniform sample of 50",xlab = '',col=gray(.9));
lines(density(res),lty=2);
curve(dunif(x, min=0, max=10), lwd=2, add=TRUE);
rug(res); mean(res); sd(res); skewness(res); # mean should be near 5
```

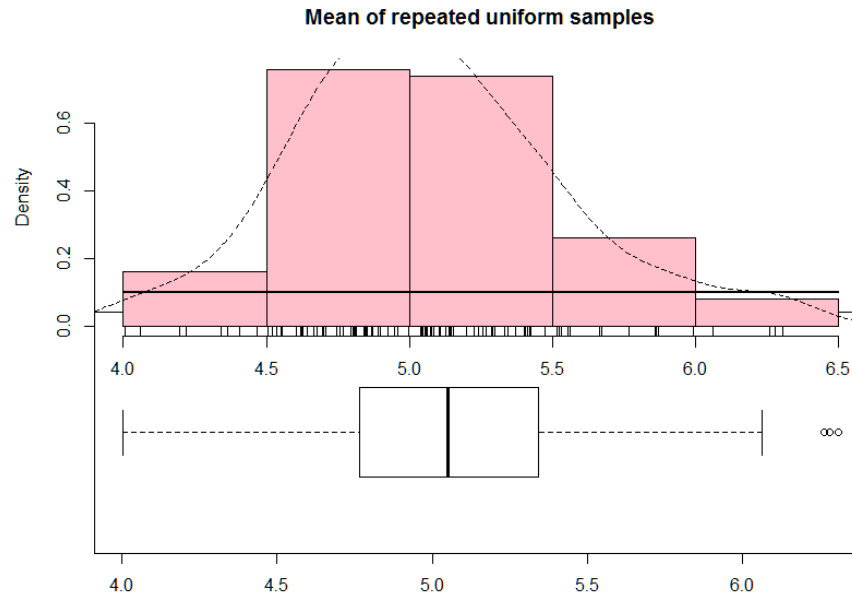
# Plotting simulated distributions

Uniform sample of 500



```
# uniform distribution;
res = runif(500, min=0, max=10);
## fig= setting uses bottom 40% of diagram;
par(fig=c(0,1,0,0.6));
boxplot(res, horizontal=TRUE, bty="n", xlab='');
## fig= setting uses top 70% of figure;
par(fig=c(0,1,.30,0.95), new=TRUE);
hist(res, prob=TRUE, main="Uniform sample of 500", xlab = '', col=gray(.9));
lines(density(res), lty=2);
curve(dunif(x, min=0, max=10), lwd=2, add=TRUE);
rug(res); mean(res); sd(res); skewness(res);
```

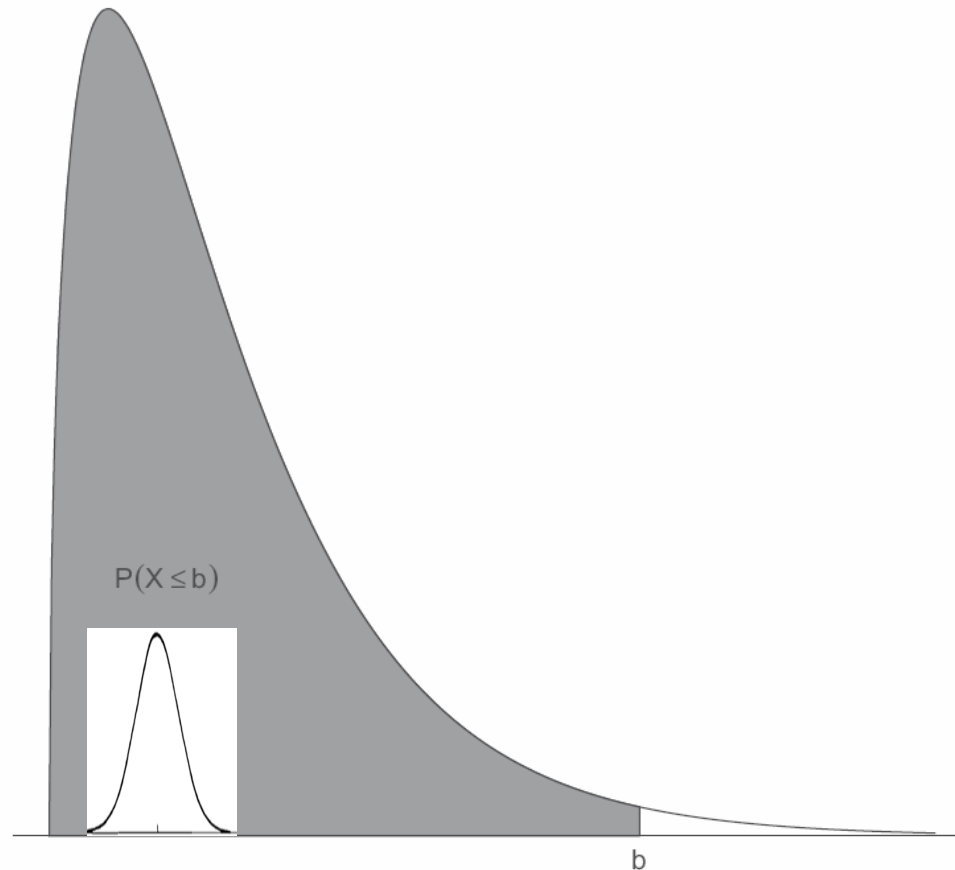
# The means of many simulated uniform distributions



```
res <- 0; #initialize res , sometimes needed
for ( i in c(1:100) ) { res[i] = mean(runif(50, 0, 10)) };
par(fig=c(0,1,0,0.6));
boxplot(res,horizontal=TRUE, bty="n", xlab='');
par(fig=c(0,1,.30,0.95), new=TRUE);
hist(res, prob=TRUE, main="Mean of repeated uniform samples", col='pink', xlab = '');
lines(density(res),lty=2);
curve(dunif(x, min=0, max=10), lwd=2, add=TRUE);
rug(res);
par(fig=c(0,1,0,1)); # reset plot dimensions
mean(res); sd(res); skewness(res); res <- 0; # reset;
```

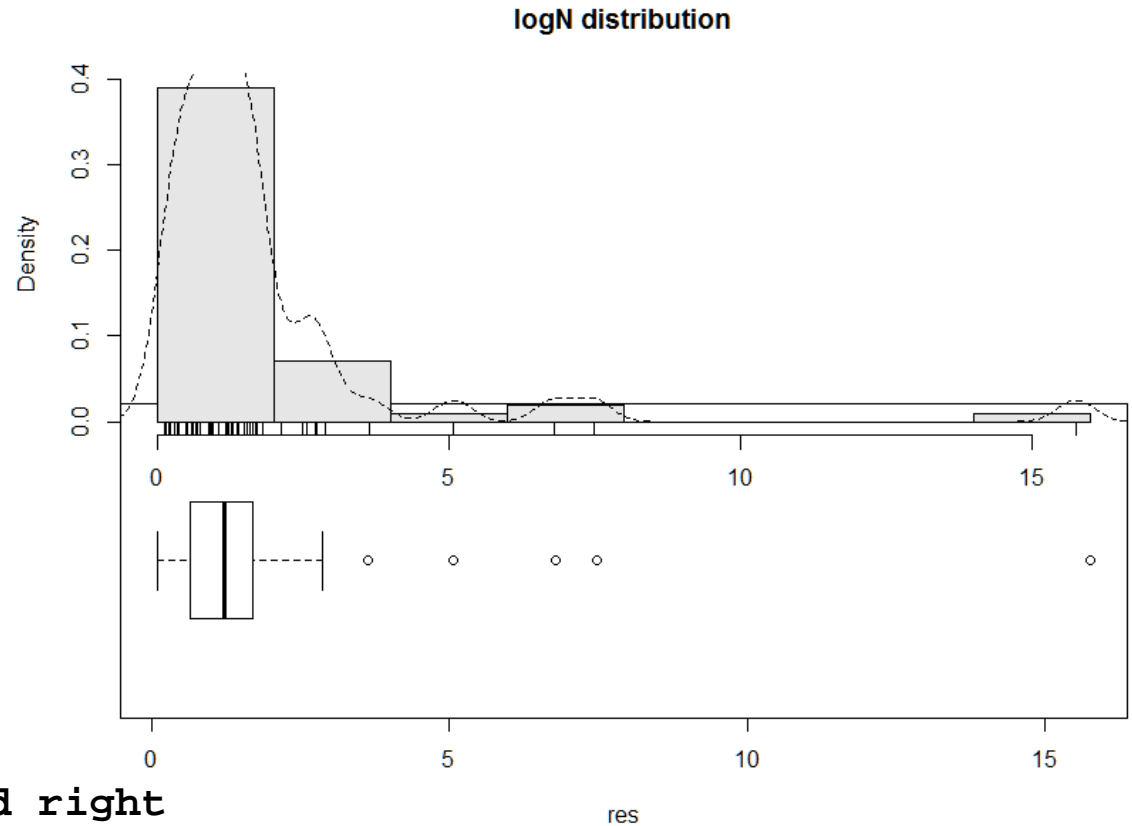
# Sampling from skewed distributions

According to the CLT,  
the sampling distribution should be approximately normal



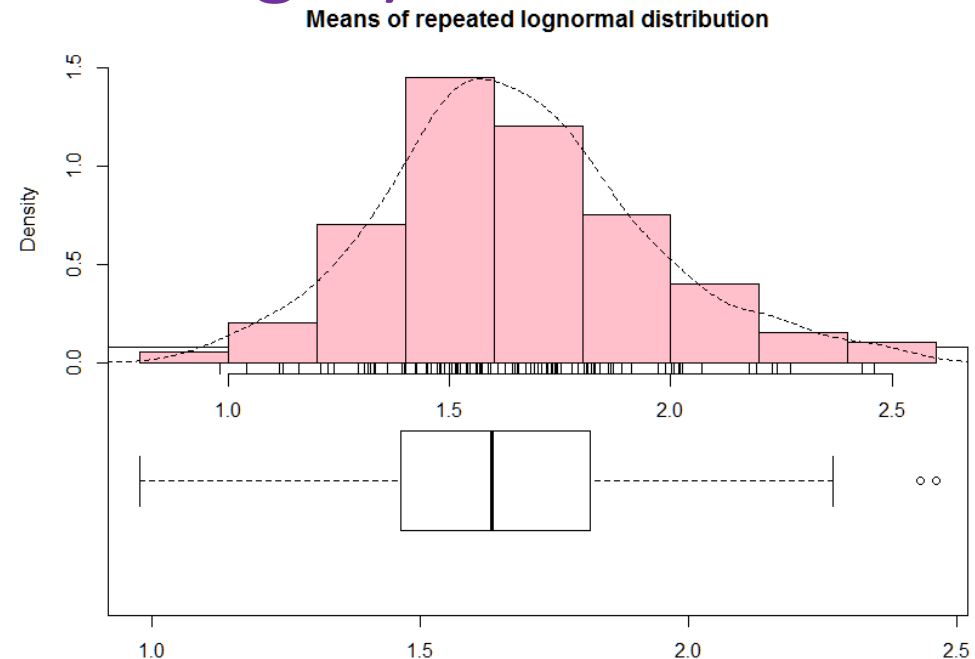


# Generating right-skewed distributions



```
# lognormal distribution;
# rlnorm() function, skewed right
res = rlnorm(50, 0, 1);
par(fig=c(0,1,0,0.6));
boxplot(res, horizontal=TRUE, bty="n", xlab = 'res');
par(fig=c(0,1,.30,0.95), new=TRUE);
hist(res, prob=TRUE, main="lognormal distribution (skewed right)", col=gray(.9),
xlab = '');
lines(density(res), lty=2);
rug(res);
par(fig=c(0,1,0,1)); # reset plot dimensions;
```

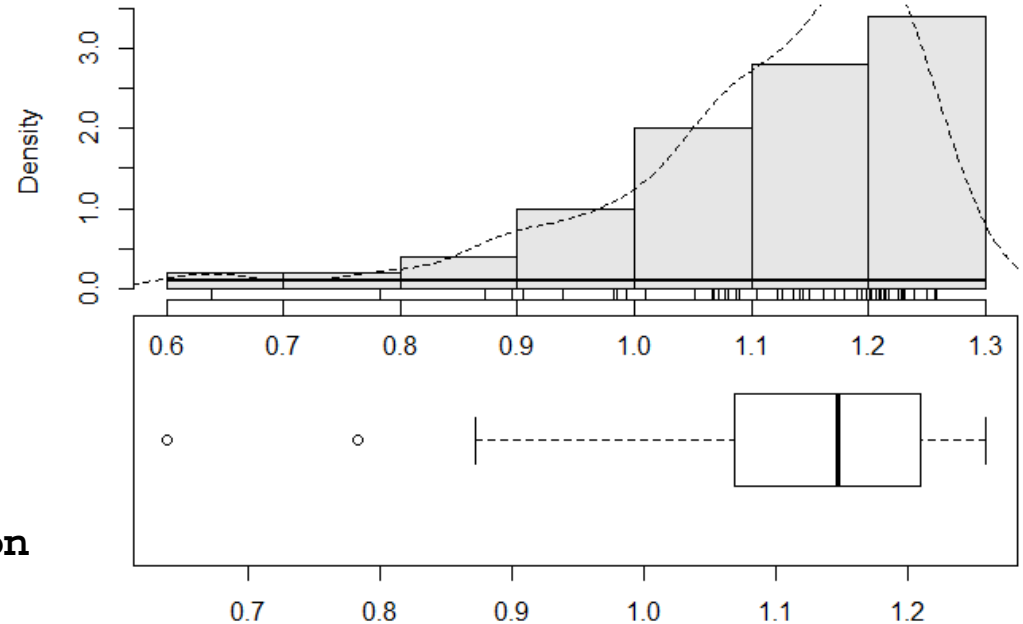
# The means of many lognormal (skewed right) distributions



```
# get the mean of 100 lognormal distributions of size 50;
for ( i in c(1:100)) { res[i] = mean(rlnorm(50, 0, 1)) };
par(fig=c(0,1,0,0.6));
boxplot(res, horizontal=TRUE, bty="n", xlab = '');
par(fig=c(0,1,.30,0.95), new=TRUE);
hist(res, prob=TRUE, main="Means of repeated lognormal distribution", col="pink",
xlab = '');
lines(density(res), lty=2);
rug(res);
par(fig=c(0,1,0,1)); # reset plot dimensions
res = 0 # clear all res values
```

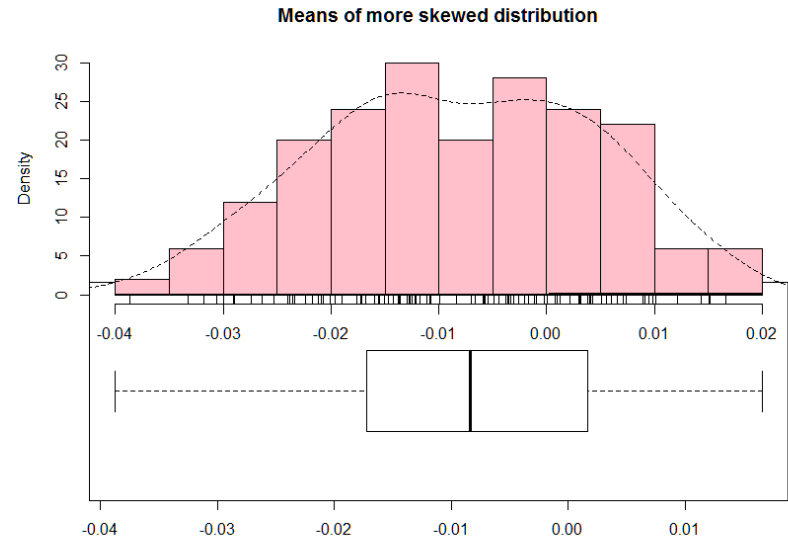
# Generating left-skewed distributions

Left-skewed distribution (10th rt)



```
# Start with uniform distribution
res = runif(50, min=0, max=10);
par(fig=c(0,1,0,0.6));
# Calculate the tenth root of res, will be left skewed;
res <- res^(0.10);
boxplot(res, horizontal=TRUE, bty="n", xlab='');
par(fig=c(0,1,.30,0.95), new=TRUE);
hist(res, prob=TRUE, xlab = '', main="Left-skewed distribution (10th rt)",
col=gray(.9));
lines(density(res), lty=2);
curve(dunif(x, min=0, max=10), lwd=2, add=TRUE);
rug(res); skewness(res);
par(fig=c(0,1,0,1)); # reset plot dimensions;
```

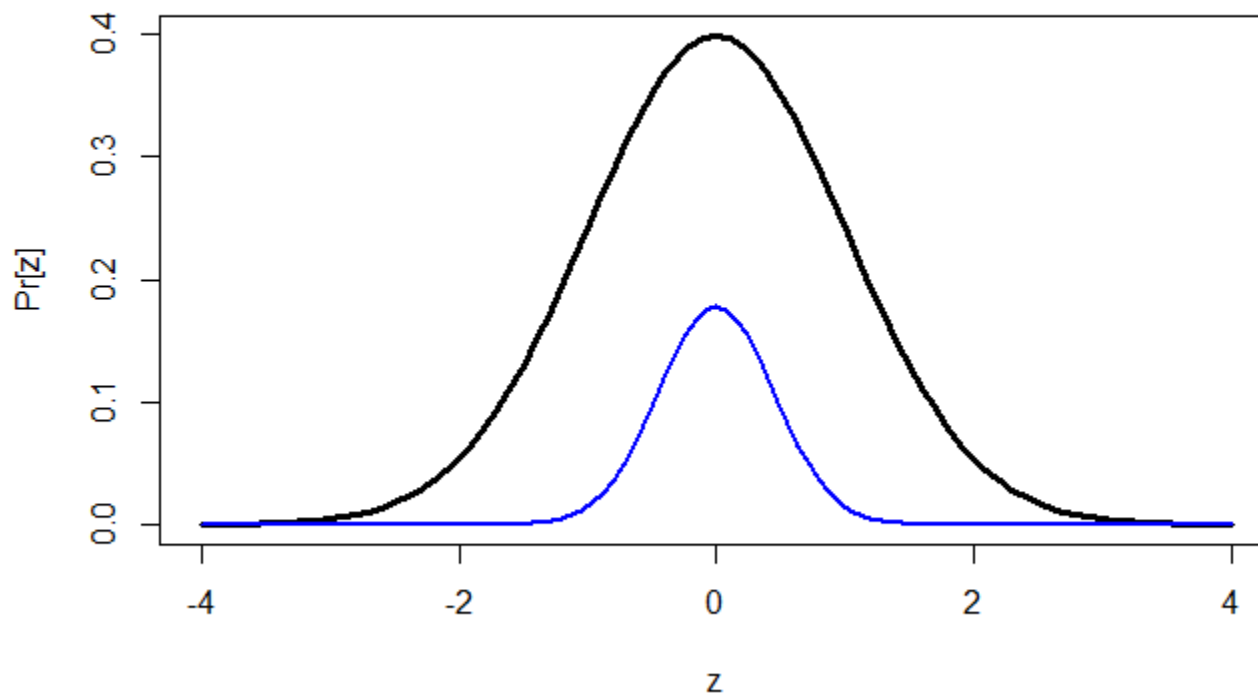
# The means of many left-skewed distributions



```
# get mean of tenth root of uniform values 100 times
for ( i in c(1:100)) { res[i] = mean(runif(50, 0, 1)^(0.10)) };
par(fig=c(0,1,0,0.6));
boxplot(res,horizontal=TRUE, bty="n", xlab='');
par(fig=c(0,1,.30,0.95), new=TRUE);
hist(res, prob=TRUE, xlab = '', main="Means of right-skewed distribution",
col='pink');
lines(density(res),lty=2);
curve(dunif(x, min=0, max=10), lwd=2, add=TRUE);
rug(res); skewness(res)
par(fig=c(0,1,0,1)); # reset plot dimensions;
res <- 0 ;
```

# The SEM is always $STDEV/\sqrt{n}$

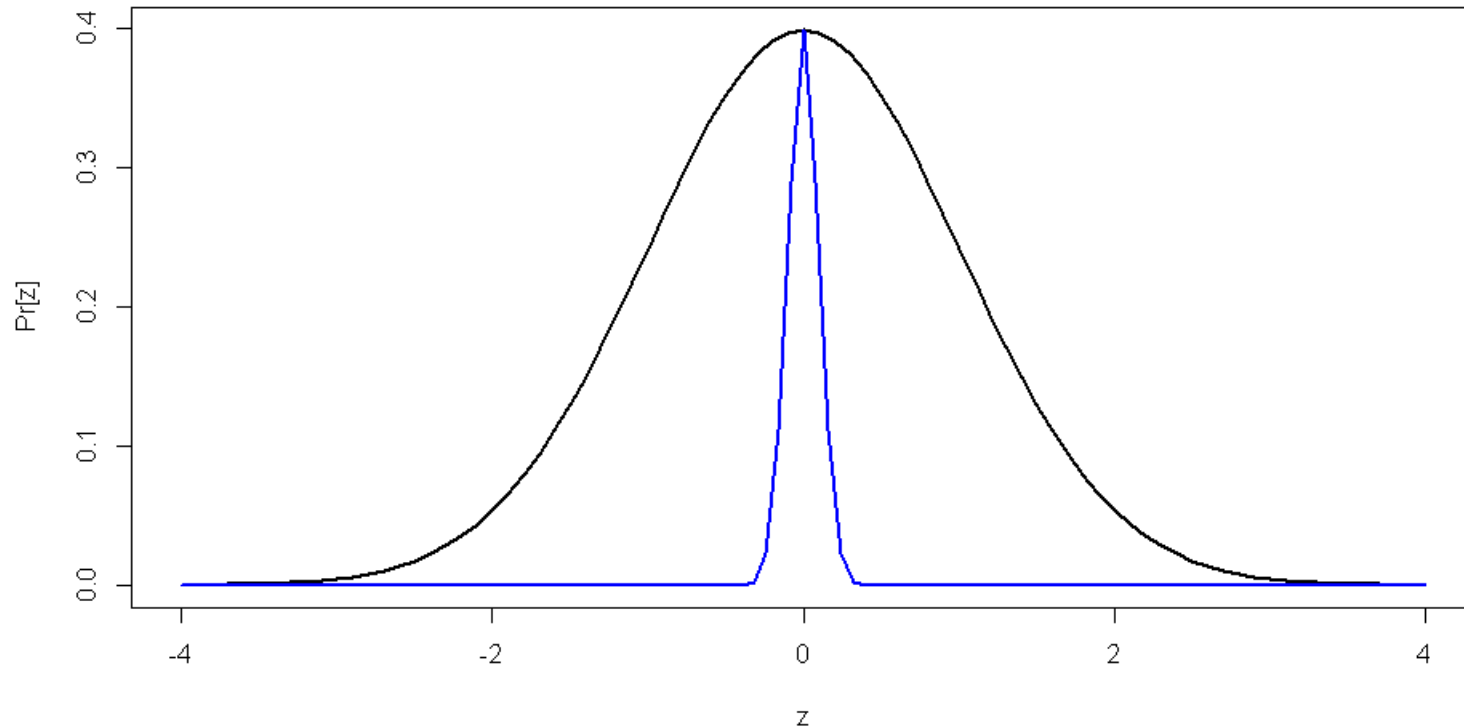
The STD and SEM (n=5) of a normal distribution



```
x <- seq(-4,4,0.1);  
# lwd = is line width (default = 1)  
plot(x,dnorm(x), type="l",xlab="z",ylab="Pr[z]",  
main="The STD and SEM (n=5) of a normal distribution", lwd = 3);  
curve(dnorm(x, mean=0, sd=1/sqrt(5) )/5, lwd = 2, col = "blue", add=T);
```

# The SEM is always $STDEV/\sqrt{n}$

The STD and SEM (n=100) of a normal distribution



```
x <- seq(-4,4,0.1);  
plot(x,dnorm(x), type="l",xlab="z",ylab="Pr[z]",  
main="The STD and SEM (n=100) of a normal distribution",lwd=2);  
curve(dnorm(x, mean=0, sd=1/sqrt(100) )/10, add=T, col="blue", lwd=2);
```

# Population mean and sample means, standard deviation and standard error

Sampling distribution of  $\bar{x}$  is centered on  $\mu$

- if data normally distributed,  $\bar{x}$  is normally distributed

(Estimated  $\sigma$  is  $s$ , but we "know"  $\sigma$  in these examples)

## Rules of thumb:

**+/- 2 standard deviations** from the mean will include 95% of the population's **individuals** (68-95-99)

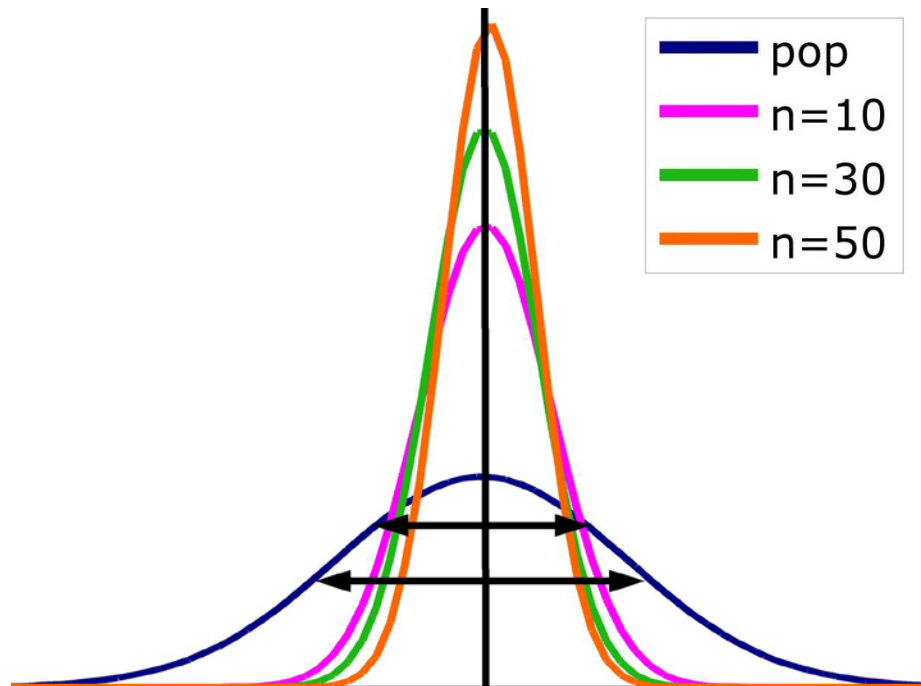
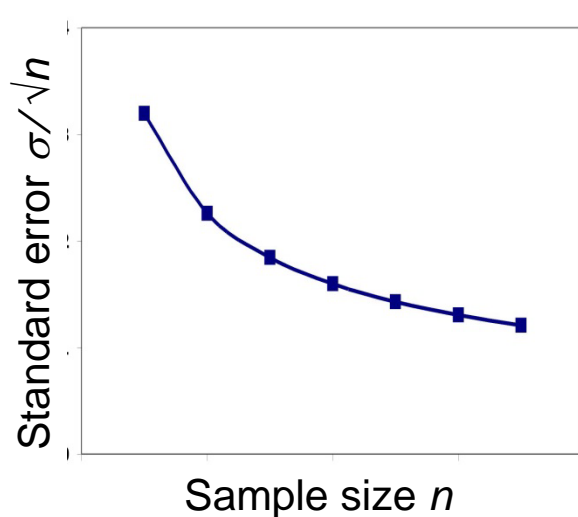
**+/- 2 standard errors from the sample mean** will include the actual **population mean** 95% of the time (68-95-99)

# Impact of Sample Size on the standard error (S.E.)



The size of the S.E. is a function of the sample size.

- The larger the sample size, the smaller the standard error.
- The spread decreases at a rate equal to  $\sqrt{n}$ .





# Confidence Interval for $\mu$



**The Big Idea:** The sampling distribution of  $\bar{x}$  tells us how close to  $\mu$  the sample mean  $\bar{x}$  is likely to be. All confidence intervals we construct will have a form similar to this:

$$\text{estimate} \pm \text{margin of error}$$

A **confidence interval (CI)** for a parameter has two parts:

- An **interval** calculated from the data, which has the form:

$$\text{estimate} \pm \text{margin of error}$$

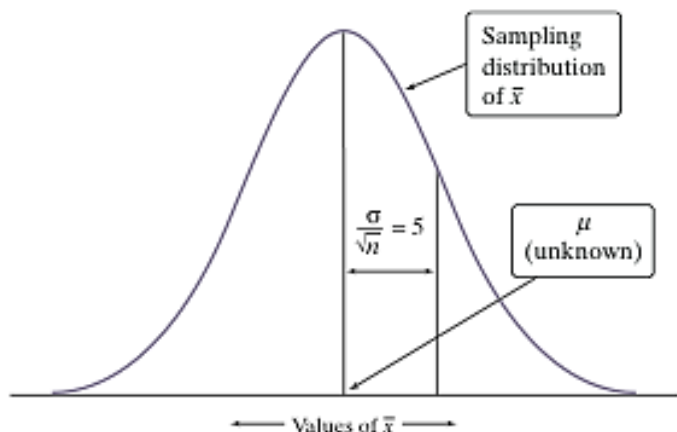
- A **confidence level**, the probability that the interval will capture the true parameter value in repeated samples.

We usually choose a confidence level of 90% or higher because we want to be quite sure of our conclusions. The most common confidence level is 95%, and the "margin of error" is usually a 95% CI.

# Confidence Interval for $\mu$

Previously, we estimated the population mean  $\mu$  by constructing a confidence interval using the sample mean  $(\bar{x}) = 240.79$ .

To calculate a 95% confidence interval for  $\mu$ , we use the formula:  
estimate  $\pm$  (z value)  $\cdot$  (standard error)



$$\begin{aligned}\bar{x} \pm z^* \cdot \frac{\sigma}{\sqrt{n}} &= 240.79 \pm 1.96 \cdot \frac{20}{\sqrt{16}} \\ &= 240.79 \pm 9.8 \\ &= (230.99, 250.59)\end{aligned}$$

## Confidence Interval for the Mean of a Normal Population

Choose an SRS of size  $n$  from a population having unknown mean  $\mu$  and known standard deviation  $\sigma$ . A level **C confidence interval for  $\mu$**  is:

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

The critical value  $z^*$  is found from the standard Normal distribution.

# The Impact of choosing a CI



The larger the CI, the more likely the population mean is actually contained.

There is no way around a small sample size. A 95% CI from a small sample will be larger than one from a larger sample from the same population.

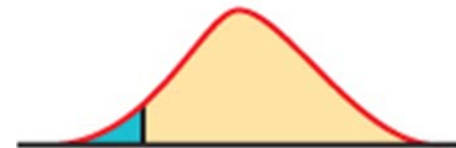
We want to be precise, but we should not deceive:

**A 50% CI is more precise but more likely incorrect than a 95% CI.**

Some small opinion polls may show a candidate with a 53% to 47% lead.

- but the **margin of error** could be  $\pm 5\%$
- so no clear leader

# Getting Probability Values: norm (normal distribution)



The “p” functions returns the **cumulative probability** (percentile/proportion) at given value (using CDF)

`pnorm(value, mu, sigma)`

`pnorm(0,0,1)` # returns cdf/perc when  $x \sim N(0,1)$  (Z scores)

```
[1] 0.5
```

`pnorm(98,120,20)` # returns cdf/perc of 98 when  $x \sim N(120,20)$

```
[1] 0.1356661
```

The “q” functions return the **value** for a cumulative probability (percentile/proportion) (0 to 1)

`qnorm(quant, mu, sigma)`

`qnorm(0.25,0,1)` # standard normal z, 25<sup>th</sup> percentile,  $x \sim N(0,1)$

```
[1] -0.6744898
```

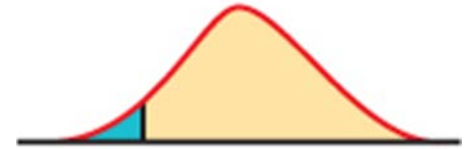
`qnorm(0.50,0,1)` # standard normal z, 50<sup>th</sup> percentile,  $x \sim N(0,1)$

```
[1] 0
```

`qnorm(0.10,160,19)` # 10<sup>th</sup> percentile value when  $x \sim N(160,19)$

```
[1] 135.6505
```

# Getting Probability Values: norm (normal distribution)



Let's experiment with different probabilities

# what is the Z score for prob = 0.25 for myoon ?

`qnorm(quant, mu, SEM)` # for CI for mean

```
qnorm(0.25,0,1)
```

```
[1] -0.6744898
```

# works for any mean and se (mean = 100, se = 10)

```
qnorm(0.25,100,10) # 25th percentile for myoo
```

```
[1] 93.2551
```

$$93.25 = 100 - (10 * |z|) \quad (z = -0.675)$$

# what is the Z score for prob = 0.1

```
qnorm(0.10,0,1)
```

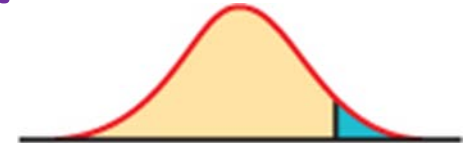
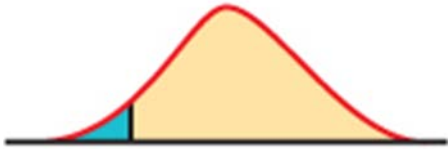
```
[1] -1.281552
```

```
qnorm(0.10,100,10) # 10th percentile for myoo
```

```
[1] 87.18448
```

$$87.2 = 100 - 10 * |z| \quad (z = -1.28)$$

# Getting Probability Values: norm (normal distribution)



Let's experiment with different probabilities

# Symmetry

```
qnorm(0.10,0,1) # 10th percentile z
```

```
[1] -1.281552
```

```
qnorm(0.10,100,10) # 10th percentile
```

```
[1] 87.18448
```

100 - 12.8

```
qnorm(0.90,0,1) # 90th percentile z
```

```
[1] 1.281552
```

```
qnorm(0.90,100,10) # 90th percentile
```

```
[1] 112.8155
```

100 + 12.8

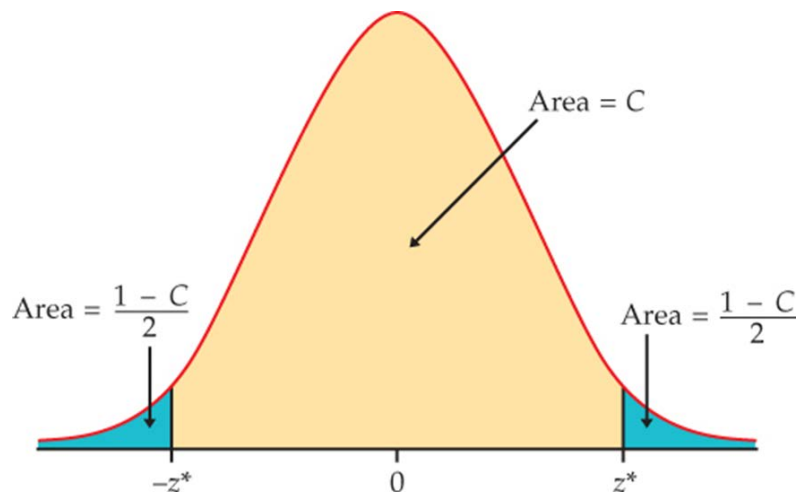
What proportion is between 0.10 and 0.90? (between 87.2 and 112.8)



# Getting intervals from mean and SE

$\bar{x} = 100, \text{SEM} = 10$

How do we get a 95 % CI  
on population mean?



```
# Symmetry: 2.5% in each tail (1 - 0.95)/2;
```

```
# qnorm(1-CI)/2,mu, SEM) # lower bound;
```

```
qnorm(0.025,100,10) # 2.5% ;
```

```
[1] 80.40036
```

```
qnorm(0.975,100,10) # 97.5 %;
```

```
[1] 119.5996
```

```
diff( c(qnorm(0.025,100,10), qnorm(0.975,100,10) ) )
```

```
[1] 39.19928
```

```
#get +/- ;
```

```
diff( c(qnorm(0.025,100,10), qnorm(0.975,100,10) ) ) / 2
```

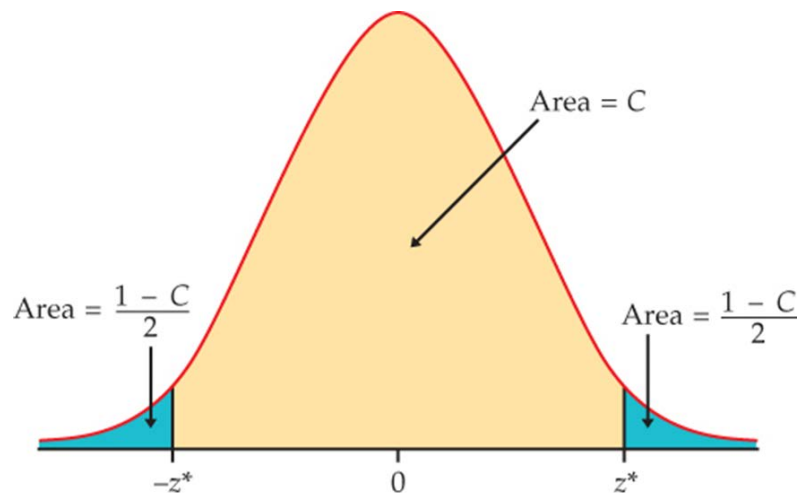
```
[1] 19.59964
```

- the estimated population mean is 100 with 95% CI +/- 19.6

# Getting intervals from mean and SE

**x = 100, SEM = 10**

**How do we get a 50 % CI  
on population mean?**



**Symmetry: 25% in each tail  $(1 - CI)/2$ ;**

**CI = 0.50 # or CI <- 0.50;**

**# qnorm(1-CI)/2,mu, SEM) # lower bound;**

**qnorm(0.25,100,10) # 25% ; code: qnorm(CI/2,100,10);**

**[1] 93.2551**

**# qnorm(CI + (1-CI),mu, se) # upper bound;**

**qnorm(0.75,100,10) # 75%;**

**[1] 106.7449**

**106.7 - 93.3;**

**[1] 13.4**

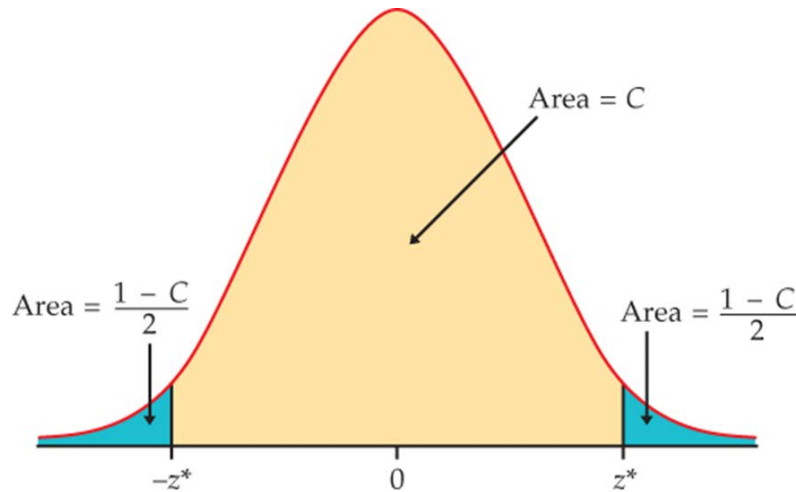
**the estimated population mean is 100 with 50% CI +/- 6.7**



# Getting intervals from mean and SE

$x = 100$ ,  $SEM = 10$

How do we get a 80 % CI  
on population mean?



Symmetry: 10% in each tail  $(1 - CI)/2$ ;

```
CI <- 0.80;
```

```
# qnorm(1-CI)/2,mu, SEM) # lower bound;
```

```
qnorm(0.10,100,10) # 10th percentile;
```

```
[1] 87.18448
```

```
# qnorm(CI + (1-CI)/2,mu, se) # upper bound;
```

```
qnorm(0.90,100,10) # 90th percentile;
```

```
[1] 112.8155
```

```
112.8 - 87.2
```

```
[1] 25.6
```

the estimated population mean is 100 with 80% CI  $\pm 12.8$

## Example 6.3 in Moore et al.

- We want to estimate the mean SAT score for CA seniors
- but 200,000 students take it out of 500,000
  - so we get some money to sample 500 RANDOM students

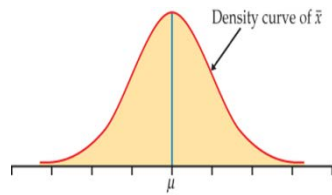
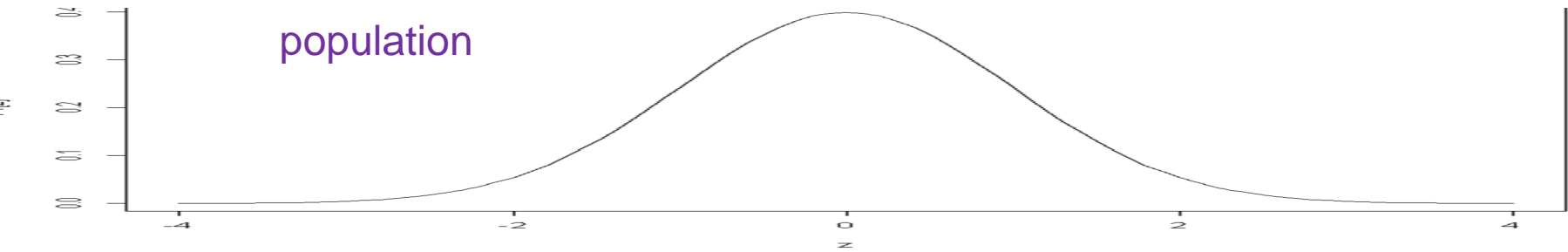
$$\bar{x} = 485$$

we "know"  $\sigma$  is 100

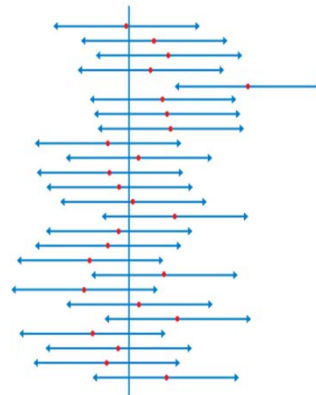
What is our estimate of  $\mu$  in the population?

The **point estimate** is  $\bar{x}$ , 485

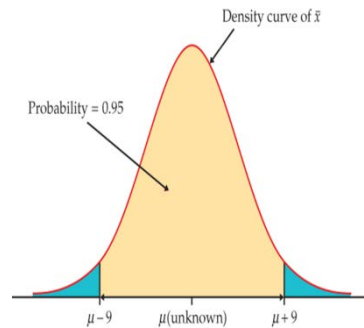
Is that enough? Why not?



repeated samples  
estimated means



estimated mean  
with CI



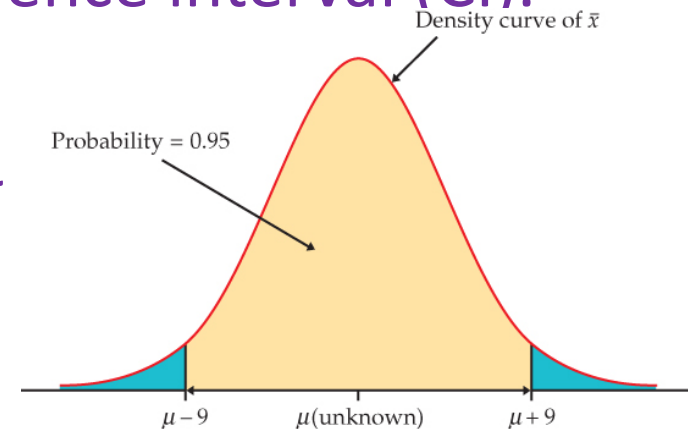
## Example 6.3 in Moore et al.

The mean is not enough because **every sample has some uncertainty**.

So we use the SEM to generate a confidence interval (CI).

Rule of thumb:  $\pm 2 \text{ SEM} = 95\% \text{ CI on } \mu$

$$\text{SEM} = \frac{100}{\sqrt{500}} = 4.5$$



So our estimate of  $\mu$  is:  $\bar{x} \pm (2 * 4.5) = 485 \pm 9$

We conclude  $\mu$  is between 476 and 494 with 95% confidence.