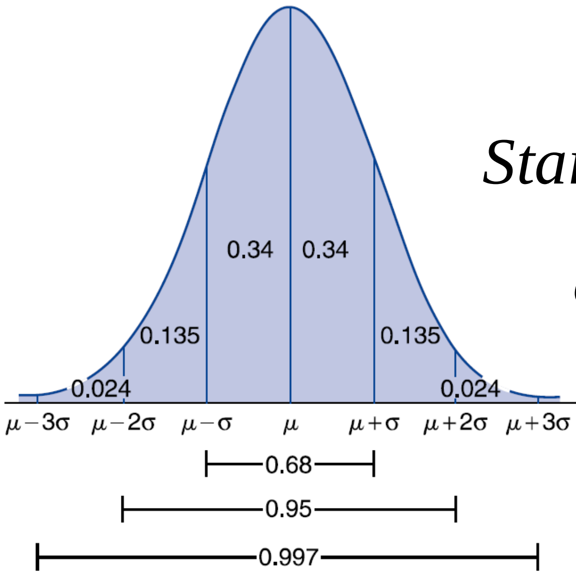
Review

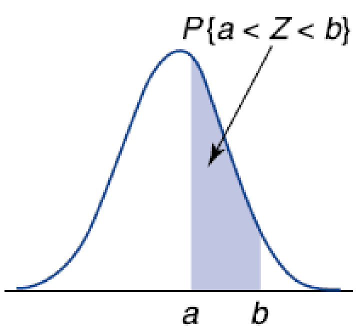


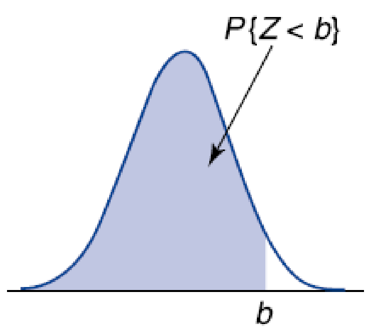
Standard Normal Distribution Z

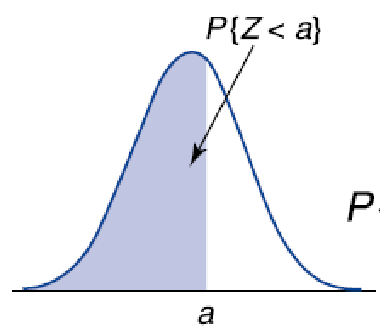
Conversion of Normal **Distribution to Standard** $Z = \dfrac{X-\mu}{\sigma}$

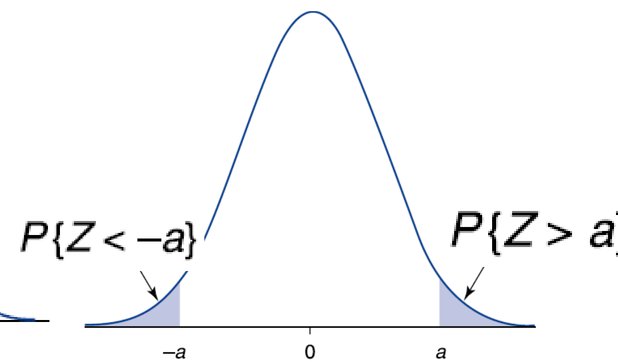
$P\{Z < -a\} = P\{Z > a\}$ *and* $P\{-a < Z < a\} = 2P\{Z < a\} - 1$

$$P\{|Z| > a\} = P\{Z > a\} + P\{Z < -a\} \;= 2P\{Z > a\}$$









Percentile of Normal Random Variables $P\{Z > z_\alpha\} = \alpha$

z_α is the intercept on X- Axis and* is negative when *α is more than 0.5

z_α is the 100(1 − α) percentile of the standard normal *distribution*

Review

Conversion of Normal Distribution to Standard

$$Z = \frac{X - \mu}{\sigma}$$

$$\mathbf{Z = \frac{[RV] - E[RV]}{SD[RV]}}$$

Continuity Correction

Then X is said to be a binomial random variable with parameters n and p.

$$P\{X = i\} = \frac{n!}{i!\,(n-i)!}p^i(1-p)^{n-i}$$

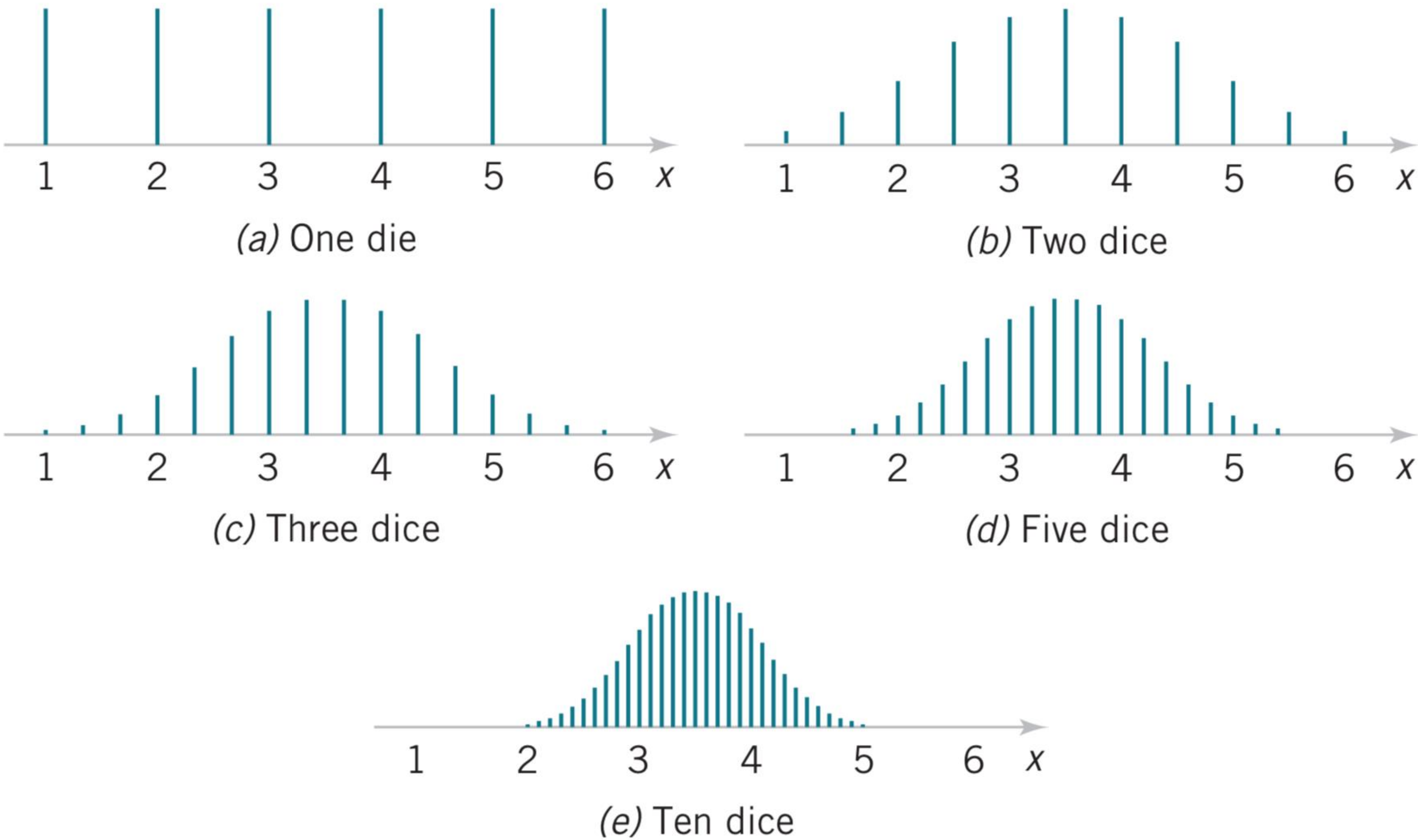
where P(x) is the probability of x successes out of n trials

Cumulative probabilities $$P\{X \le j\} = \sum_{i=0}^{j} P\{X = i\} = \sum_{i=0}^{j} \frac{n!}{i!\,(n-i)!}p^i(1-p)^{n-i}$$

Case: 60 or more heads in 100 flips

$$= P(60) + P(61) + \ldots\ldots + P(100)$$

Continuity Correction



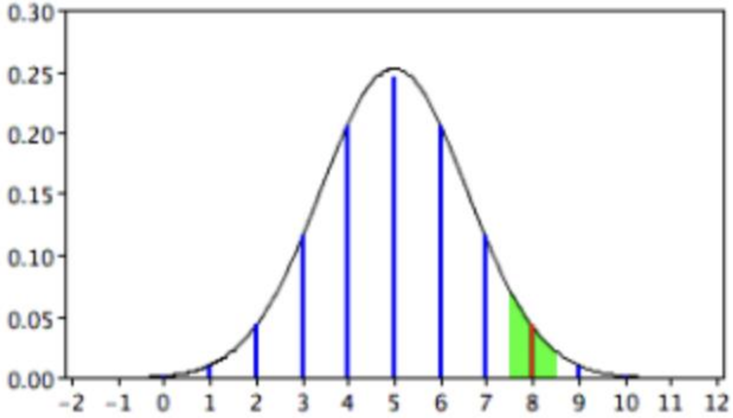
Continuity Correction

Probability of 8 heads out of 10 flips of a fair coin

$n = 10, \; p = 0.5$ and $X = 8$

$\mathcal{E}[X] = np = 10(0.5) = 5$

$SD = [np(1-p)]^{1/2} = 1.897$



What is the probability of getting a value exactly 1.897 standard deviations above the mean

$P[X = 8] = P[7.5 < X < 8.5]$: Continuity Correction

Now we have to find $P[7.5 < X < 8.5]$ with $\mathcal{E}[X] = 5$ and SD of 1.897

$P[7.5 < X < 8.5] = 4.4\ \%$

Probability of 8 heads out of 10 flips of a fair coin = 4.4 %

Continuity Correction

Normal Approximation: Both np and np(1-p) should be equal or higher than 5

Probability	Normal Approximation
$P[X = n]$	$P[n - 0.5 < X < n + 0.5]$
$P[X > n]$	$P[X > n + 0.5]$
$P[X < n]$	$P[X < n - 0.5]$
$P[X \leqslant n]$	$P[X < n + 0.5]$
$P[X \geqslant n]$	$P[X > n - 0.5]$

Continuity Correction

Now for a binomial distribution

$$E[X] = np \quad \text{and} \quad \mathrm{SD}(X) = \sqrt{np(1-p)}$$

Since $\overline{X}$, the proportion of the sample that has the characteristic, is equal to X/n, we see that

$$E[\overline{X}] = \frac{E[X]}{n} = p$$

and

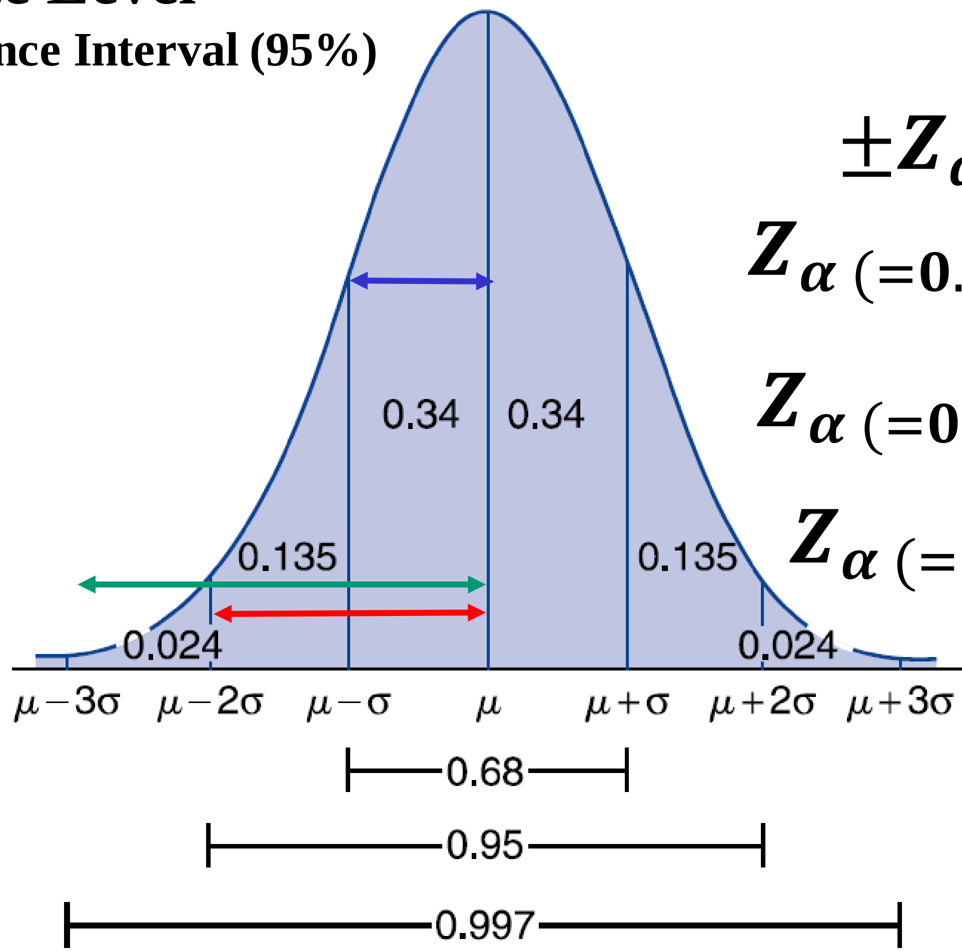
$$\mathrm{SD}(\overline{X}) = \frac{\mathrm{SD}(X)}{n} = \sqrt{\frac{p\,(1-p)}{n}}$$

$$P\{X \le j\} = \sum_{i=0}^{j} P\{X = i\} = \sum_{i=0}^{j} \frac{n!}{i!\,(n-i)!} p^i (1-p)^{n-i}$$

Sampling Statistics

α = Significance Level

100(1- α) = Confidence Interval (95%)



$$\pm Z_{\alpha/2}\sigma$$

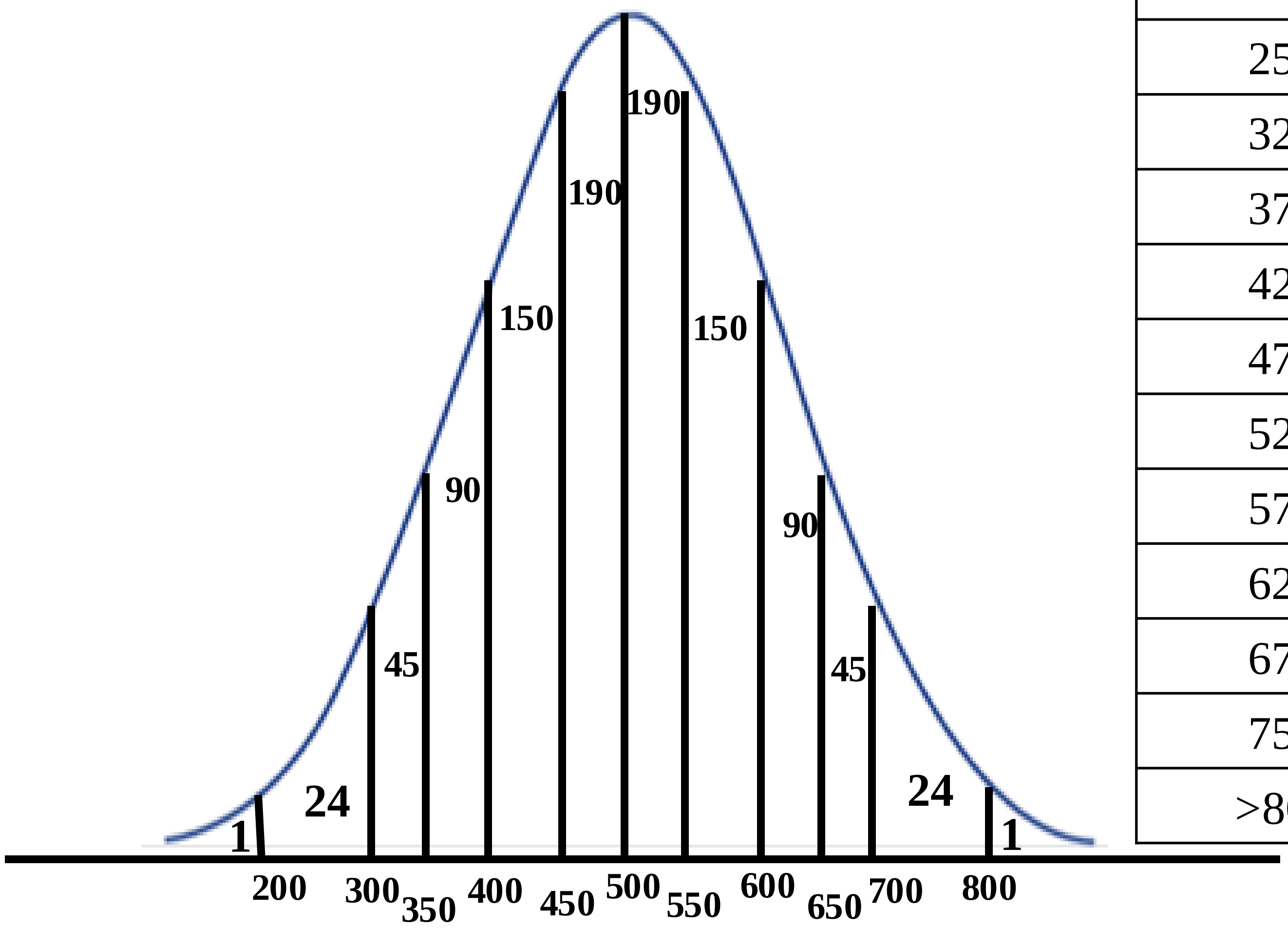
$$Z_{\alpha\,(=0.05)/2}$$

$$Z_{\alpha\,(=0.003)/2}$$

$$Z_{\alpha\,(=0.16)/2}$$

Sampling Statistics

Population Size =1000, $\mu = 500$ $\sigma = 100$

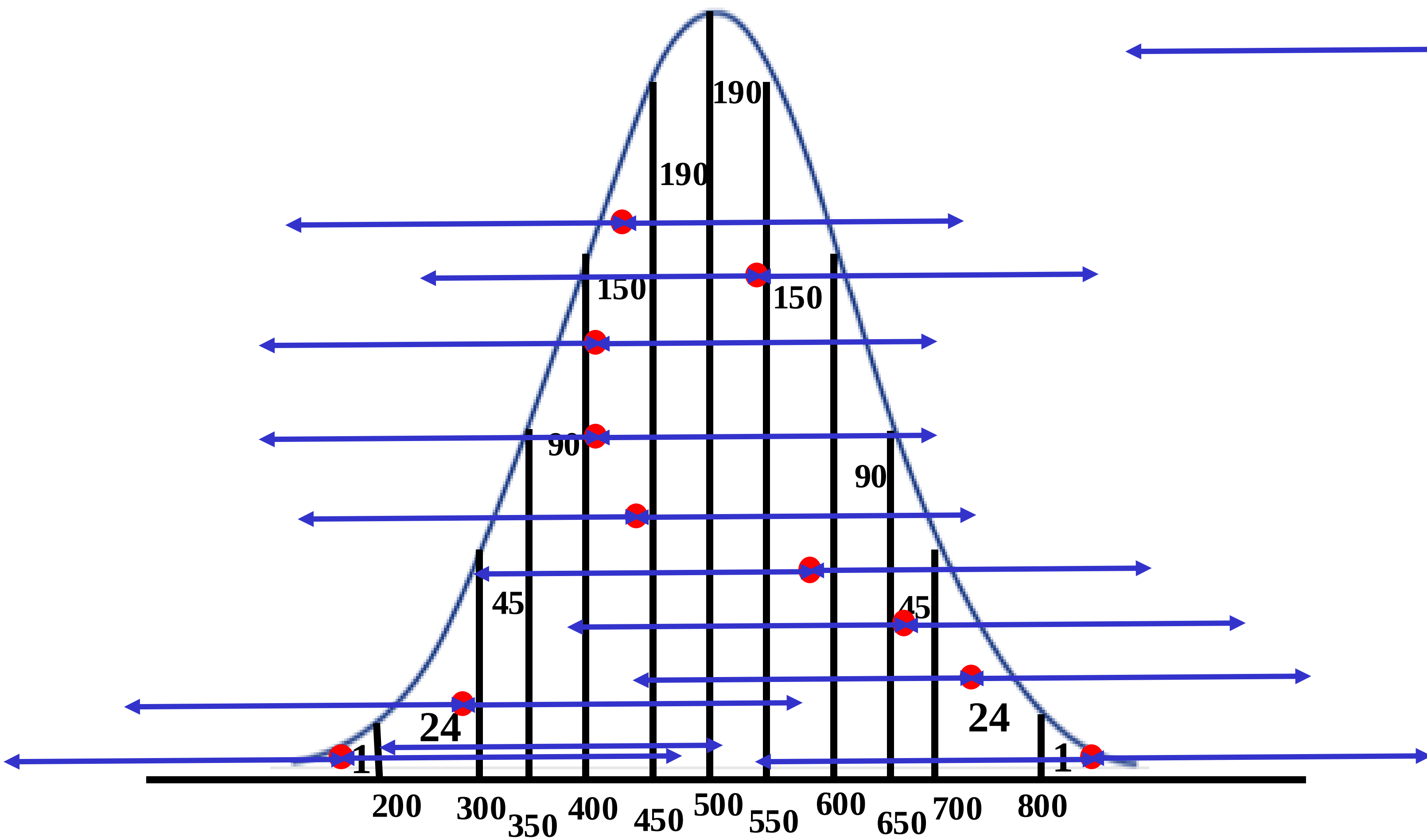


Income	Number
< 200	1
250	24
325	45
375	90
425	150
475	190
525	190
575	150
625	90
675	45
750	24
>800	1

Distribution of approximate number of members in each range

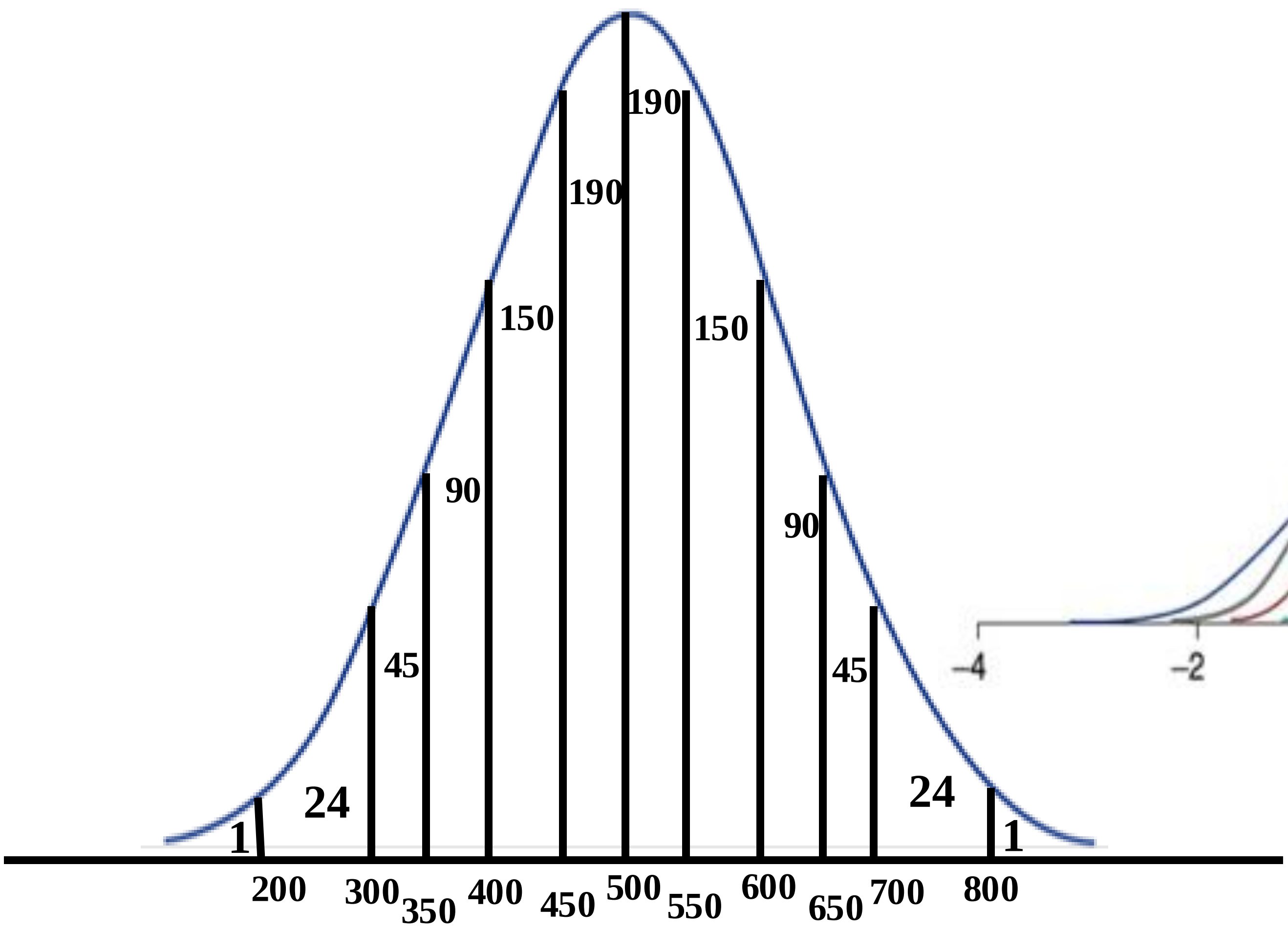
Population Size =1000, $\mu = 500$ $\sigma = 100$

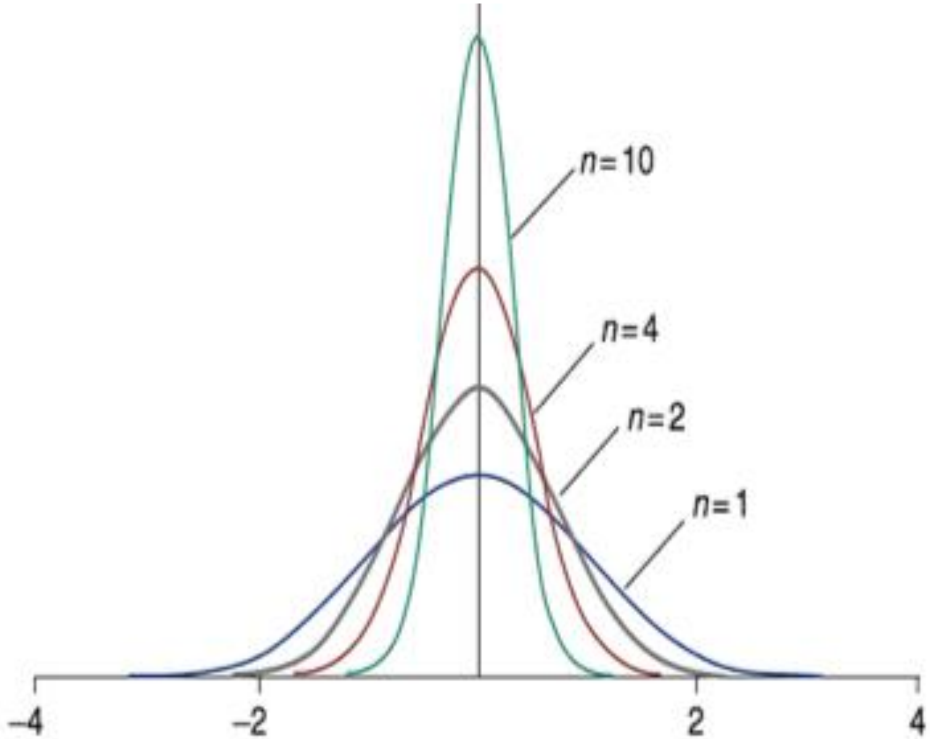
$$Sample\ Mean \pm Z_{\alpha/2}\sigma$$



Distribution of approximate number of members in each range

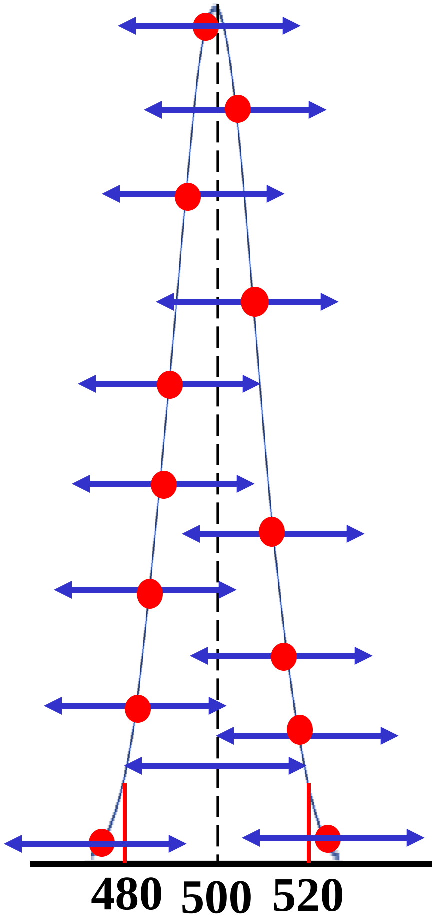
Population Size =1000, μ = 500 σ = 100

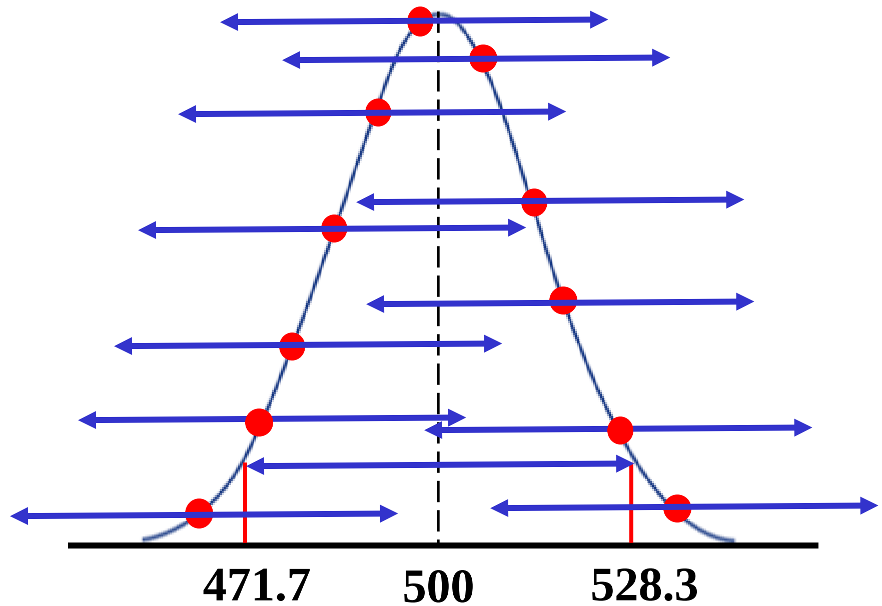




Distribution of approximate number of members in each range

Sampling Statistics





Random Variable

Sample Size vs Sample Numbers

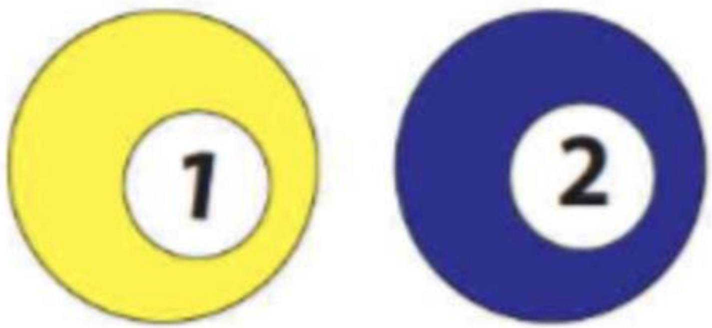
Sampling Distribution

Expected Value of Sampling Distribution

Standard deviation of Sampling Distribution

Sampling Statistics

Considering a population whose values are equally likely to be either 1 or 2. That is, if *X is the value of a member of* the population, then



$$P\{X = 1\} = \frac{1}{2} \qquad P\{X = 2\} = \frac{1}{2}$$

Sampling Statistics

Calculate the expected value and variance of this population

$$\mu = E[X] = 1\left(\frac{1}{2}\right) + 2\left(\frac{1}{2}\right) = 1.5$$

$$\sigma^2 = \mathrm{Var}(X) = E[(X-\mu)^2] \quad = (1-1.5)^2\left(\frac{1}{2}\right) + (2-1.5)^2\left(\frac{1}{2}\right) \ = \frac{1}{4}$$

Sampling Statistics

Take a sample size of 2 from this population and obtain the probability distribution of the sample mean $(X_1 + X_2)/2$, *and calculate its expected value and variance*

Since the pair of values X_1, X_2 *can assume any of four possible pairs of values, the sample space will be*

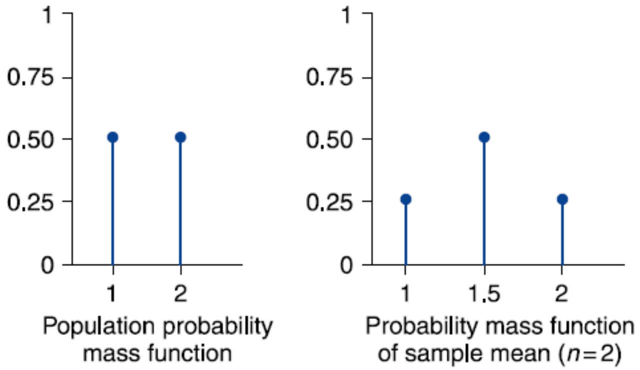
$$(1, 1), (1, 2), (2, 1), (2, 2)$$

$$P\{\overline{X} = 1\} = P\{(1, 1)\} = \frac{1}{4}$$

$$P\{\overline{X} = 1.5\} = P\{(1, 2) \text{ or } (2, 1)\} = \frac{2}{4} = \frac{1}{2}$$

$$P\{\overline{X} = 2\} = P\{(2, 2)\} = \frac{1}{4}$$

Sampling Statistics



$$E[\overline{X}] = 1\left(\frac{1}{4}\right) + 1.5\left(\frac{1}{2}\right) + 2\left(\frac{1}{4}\right) = \frac{6}{4} = 1.5$$

Also

$$\begin{aligned}\mathrm{Var}(\overline{X}) &= E\left[(\overline{X} - 1.5)^2\right] \\ &= (1 - 1.5)^2\left(\frac{1}{4}\right) + (1.5 - 1.5)^2\left(\frac{1}{2}\right) + (2 - 1.5)^2\left(\frac{1}{4}\right) \\ &= \frac{1}{16} + 0 + \frac{1}{16} = \frac{1}{8}\end{aligned}$$

Sampling Statistics

Obtain the probability distribution of the sum

SUM: Probability = 2:1/4   3:1/2   4:1/4

E[SUM] = 3,

Variance [SUM] = 1/2

Sampling Statistics

Obtain the probability distribution of the sample mean if sample size is 3 *and calculate its expected value and variance*

For Sample Size 3

$$E[\bar{X}] = 1.5$$

$$Variance\,[\bar{X}] = \frac{1}{12}$$

Obtain the probability distribution of the sum if sample size is 3

SUM: Probability = 3:1/8   4:3/8   5:3/8   6:1/8

E[SUM] = 4.5, Variance [SUM] = 3/4

Sampling Statistics

Obtain the probability distribution of the sample mean if sample size is 4 *and calculate its expected value and variance*

For Sample Size 4

$$E[\bar{X}] = 1.5$$

$$Variance\ [\bar{X}] = \frac{1}{16}$$

Obtain the probability distribution of the sum and calculate its expected value and variance if sample size is 4

E[SUM] = 6.0, Variance [SUM] = 1

Sampling Statistics

Population	Mean = ?	Variance = ?

Sample Size	E[Mean]	Variance
2	?	?
3	?	?
4	?	?

Sample Size	E[Sum]	Variance
2	?	?
3	?	?
4	?	?

Sampling Statistics

Population	Mean = 1.5	Variance = 1/4

Sample Size	E[Mean]	Variance
2	1.5	1/8
3	1.5	1/12
4	1.5	1/16

Sample Size	E[Sum]	Variance
2	3.0	1/2
3	4.5	3/4
4	6.0	1

Sampling Statistics

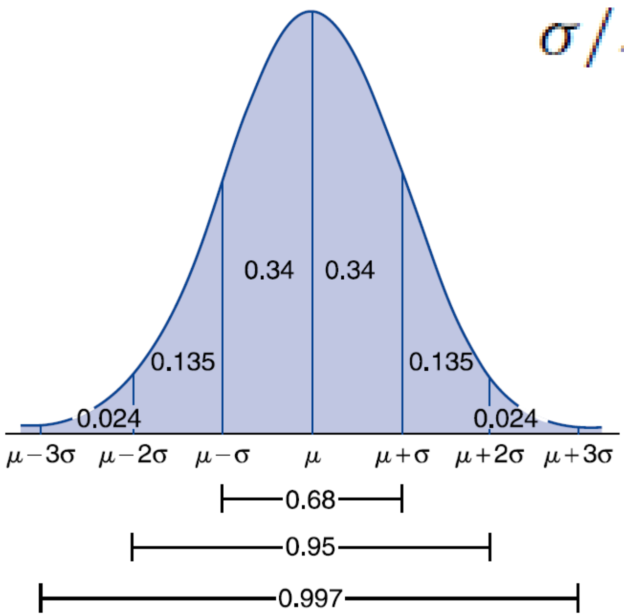
Size	X	Sample Mean	Probability	E(Mean)	Variance
2	2	1	1/4	1.5	1/8
	3	1.5	1/2		
	4	2	1/4		
3	3	1	1/8	1.5	1/12
	4	1.33	3/8		
	5	1.7	3/8		
	6	2	1/8		
4	4	1	1/16	1.5	1/16
	5	1.25	4/16		
	6	1.5	6/16		
	7	1.75	4/16		
	8	2	1/16		

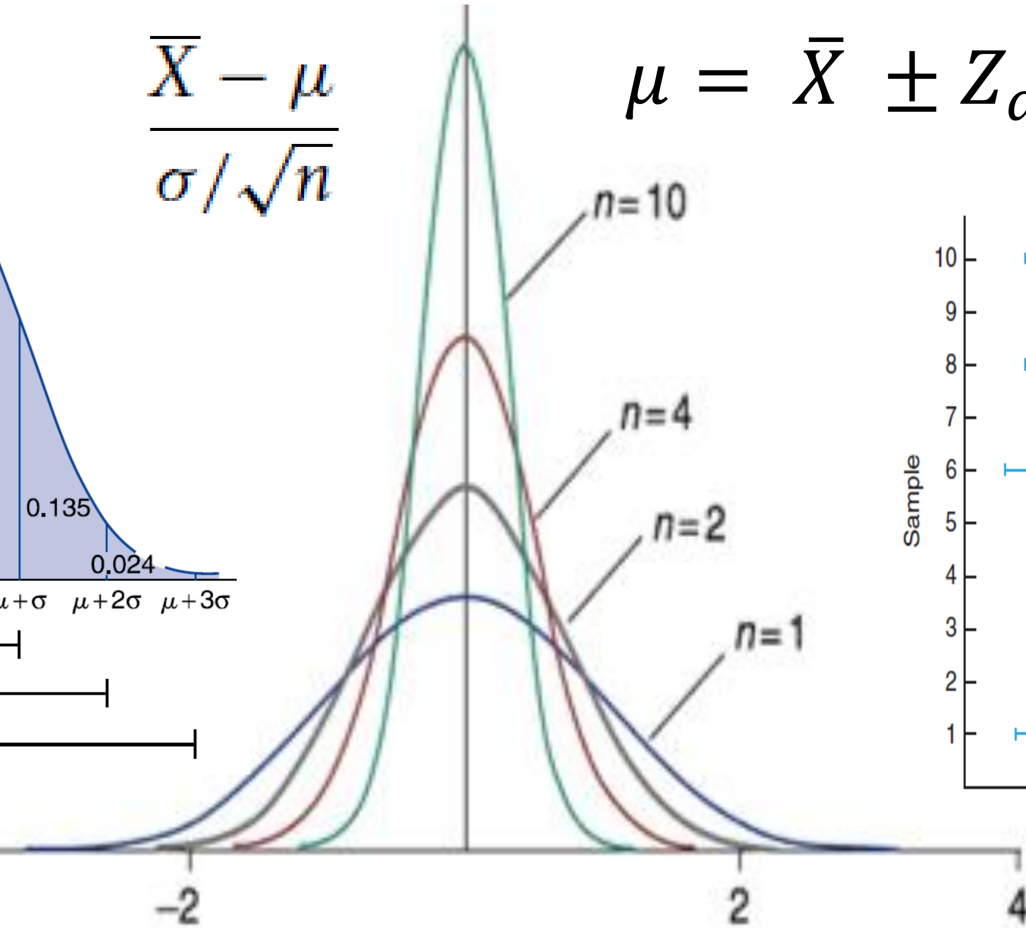
Sampling Statistics

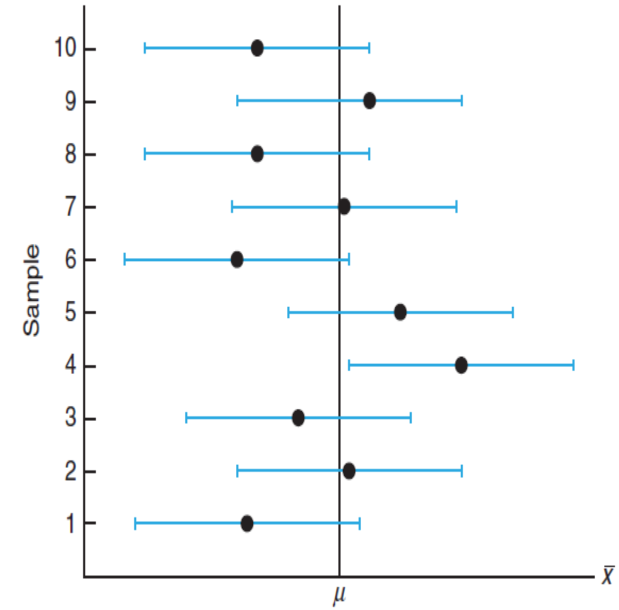
$$Z = \frac{[RV] - E[RV]}{SD[RV]}$$

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

$$\mu = \bar{X} \pm Z_{\alpha/2}\sigma/\sqrt{n}$$







PROBABILITY OF PICKING UP EXTREME SAMPLE DECREASES AS SAMPLE SIZE INCREASES

As Sample Size Increases, Variance Decreases (sample spread decreases) and hence picking up Samples closer to Mean increases

Sampling Statistics

Ability to learn about the underlying population distribution by observing the sample data

If $X_1, \ldots, X_n$ are independent random variables having a common probability distribution, we say they constitute a sample from that distribution.

Let $X_1, \ldots, X_n$ be a sample of values from the population with mean μ and variance σ^2

Then *the sample mean is* defined by

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n}$$

Since the value of the sample mean $\overline{X}$ *is determined by the values of the random* variables in the sample, it follows that $\overline{X}$ *is also a random variable. Its expectation* can be shown to be

$$E[\overline{X}] = \mu$$

That is, the expected value of the sample mean $\overline{X}$ *is equal to the population* mean μ.

In addition, it can be shown that the variance of the sample mean is

$$\text{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

Sampling Statistics

We can conclude that sample mean *is also centred on the population mean μ, but its spread becomes more and* more reduced as the sample size increases.

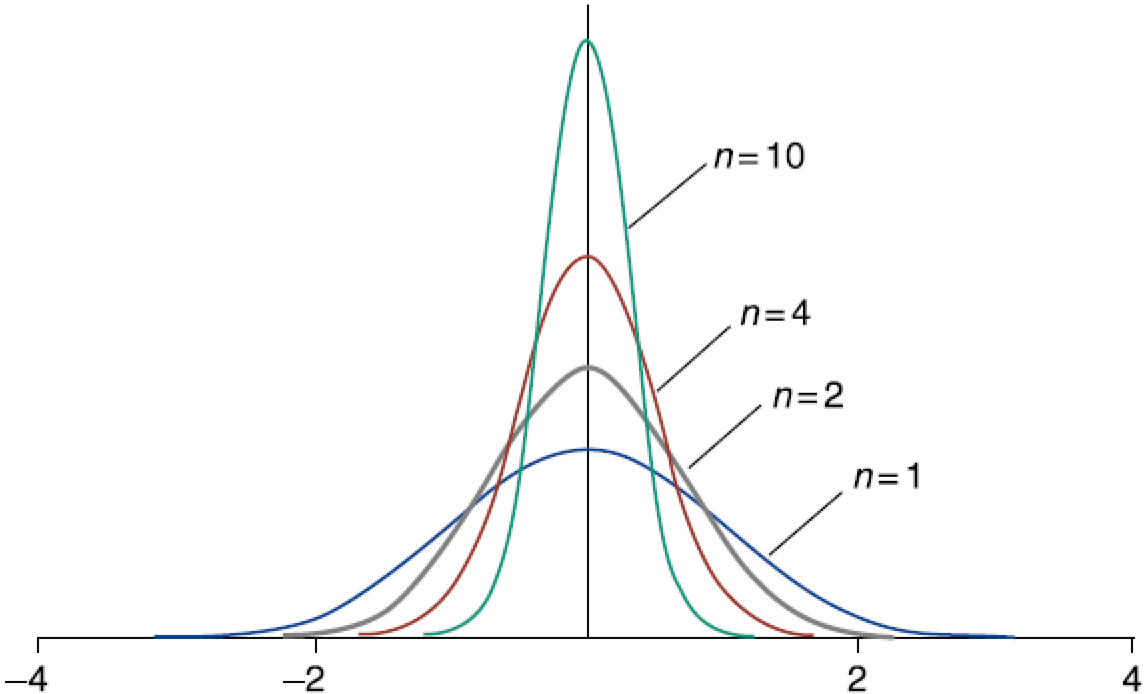


FIGURE 7.1

Densities of sample means from a standard normal population.

Sampling Distribution

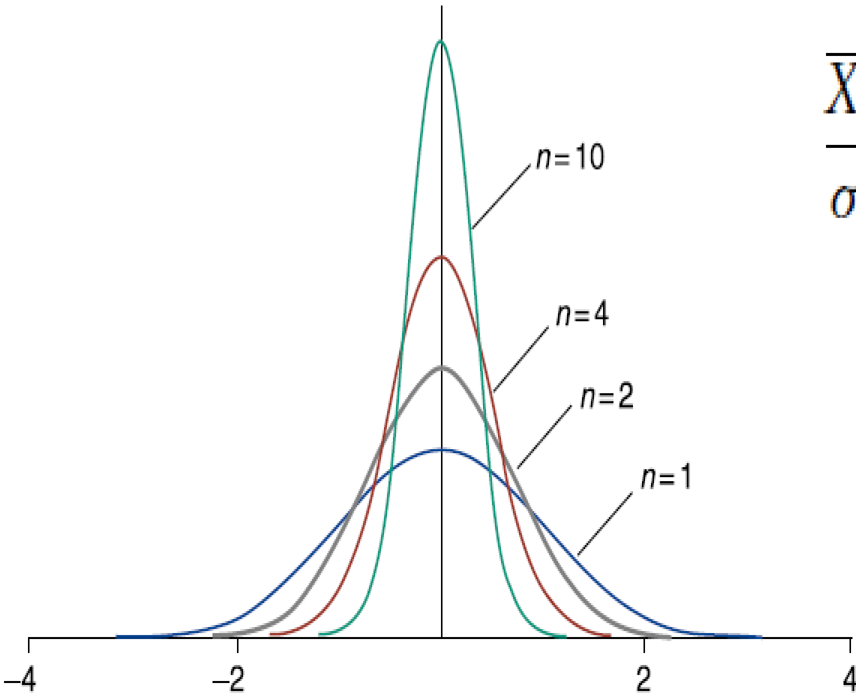


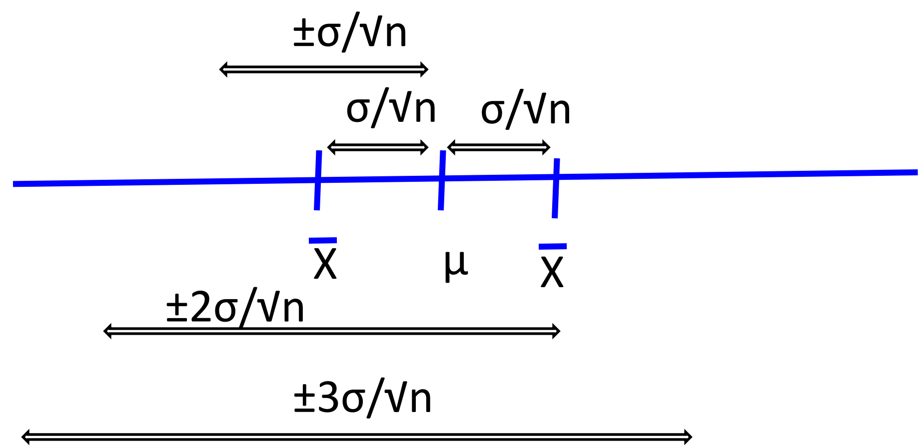
FIGURE 7.1

Densities of sample means from a standard normal population.

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

$$\mathrm{SD}(\overline{X}) = \frac{\sigma}{\sqrt{n}}$$

standard error of $\overline{X}$ as an estimator of the mean.



Central Limit Theorem

One of the most important results in probability theory, known as the *central limit theorem, which states that the sum (and thus also the average)* of a large number of independent random variables is approximately normally distributed.

Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be a sample from a population having mean μ and standard deviation σ. For n large, the sum

$$X_1 + X_2 + \cdots + X_n$$

will approximately have a normal distribution with mean $n\mu$ and standard deviation $\sigma\sqrt{n}$.

Provided that all the random variables tend to be of roughly the same magnitude so that none of them tends to dominate the value of the sum, it can be shown that the sum of a large number of independent random variables will have an approximately normal distribution

Distribution of the Sample Mean

Let $X_1, \ldots, X_n$ be a sample of values from the population with mean μ and variance σ^2

Then *the sample mean is* defined by

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n}$$

it follows from the central limit theorem that $\overline{X}$ *also will be approximately normal when the* sample size *n is large. Since* $\overline{X}$ *has expectation μ and standard deviation* $\sigma/\sqrt{n}$, the standardized variable

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

has an approximately standard normal distribution

Distribution of the Sample Mean

Let $\overline{X}$ be the sample mean of a sample of size n from a population having mean μ and variance σ^2. By the central limit theorem,

$$P\{\overline{X} \le a\} = P\left\{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le \frac{a-\mu}{\sigma/\sqrt{n}}\right\}$$

$$\approx P\left\{Z \le \frac{a-\mu}{\sigma/\sqrt{n}}\right\}$$

where Z is a standard normal.

Suppose that exactly 46 percent of the population favours a particular candidate. If a random sample of size 200 is chosen, what is the probability that at least 100 favour this candidate?

If *X is the number who favour the candidate, then X is a binomial random* variable with parameters *n = 200 and p = 0.46. The desired probability is P{X ≥ 100}.*

Therefore, to compute *P{X ≥ 100}, we should use the normal approximation on the equivalent* probability *P{X ≥ 99.5}. Considering the standardized variable*

$$\frac{X - 200(0.46)}{\sqrt{200(0.46)(0.54)}} = \frac{X - 92}{7.0484}$$

$$P\{X \geq 100\} = P\{X \geq 99.5\} \quad = P\left\{\frac{X - 92}{7.0484} \geq \frac{99.5 - 92}{7.0484}\right\} \approx P\{Z > 1.0641\}$$

$$= 0.144$$

Distribution of the Sample Mean

An insurance company has 10,000 ($=10^4$) automobile policyholders. If the expected yearly claim per policyholder is \$260 with a standard deviation of \$800, approximate the probability that the total yearly claim exceeds \$2.8 million ($=\$2.8 \times 10^6$).

Number the policyholders, and let X_i denote the yearly claim of policyholder i, $i = 1, \ldots, 10^4$. By the central limit theorem, $X = \sum_{i=1}^{10^4} X_i$ will have an approximately normal distribution with mean $10^4 \times 260 = 2.6 \times 10^6$ and standard deviation $800\sqrt{10^4} = 800 \times 10^2 = 8 \times 10^4$. Hence,

$$\begin{aligned}
P\{X > 2.8 \times 10^6\} &= P\left\{\frac{X - 2.6 \times 10^6}{8 \times 10^4} > \frac{2.8 \times 10^6 - 2.6 \times 10^6}{8 \times 10^4}\right\} \\
&\approx P\left\{Z > \frac{0.2 \times 10^6}{8 \times 10^4}\right\} \\
&= P\left\{Z > \frac{20}{8}\right\} \\
&= P\{Z > 2.5\} = 0.0062
\end{aligned}$$

where $\approx$ means "is approximately equal to." That is, there are only 6 chances out of 1000 that the total yearly claim will exceed \$2.8 million. ■

Distribution of the Sample Mean

The blood cholesterol levels of a population of workers have mean 202 and standard deviation 14.

(a) If a sample of 36 workers is selected, approximate the probability that the sample mean of their blood cholesterol levels will lie between 198 and 206.

(b) Repeat (a) for a sample size of 64.

(a) It follows from the central limit theorem that $\overline{X}$ is approximately normal with mean $\mu = 202$ and standard deviation $\sigma/\sqrt{n} = 14/\sqrt{36} = 7/3$. Thus the standardized variable

$$W = \frac{\overline{X} - 202}{7/3}$$

has an approximately standard normal distribution. To compute $P\{198 \leq \overline{X} \leq 206\}$, first we must write the inequality in terms of the standardized variable W. This results in the equality

$$P\{198 \leq \overline{X} \leq 206\} = P\left\{\frac{198 - 202}{7/3} \leq \frac{\overline{X} - 202}{7/3} \leq \frac{206 - 202}{7/3}\right\}$$

Distribution of the Sample Mean

$$= P\{-1.714 \leq W \leq 1.714\}$$

$$\approx P\{-1.714 \leq Z \leq 1.714\}$$

$$= 2P\{Z \leq 1.714\} - 1$$

$$= 0.913$$

For a sample size of 64, the sample mean $\overline{X}$ will have mean 202 and standard deviation $14/\sqrt{64} = 7/4$. Hence, writing the desired probability in terms of the standardized variable

$$\frac{\overline{X} - 202}{7/4}$$

yields

$$P\{198 \leq \overline{X} \leq 206\} = P\left\{\frac{198 - 202}{7/4} \leq \frac{\overline{X} - 202}{7/4} \leq \frac{206 - 202}{7/4}\right\}$$

$$\approx P\{-2.286 \leq Z \leq 2.286\} \quad = 0.978$$

Thus, we see that increasing the sample size from 36 to 64 increases the probability that the sample mean will be within 4 of the population mean from 0.913 to 0.978

Interval Estimator

When we estimate a parameter by a point estimator, we do not expect the resulting estimator to exactly equal the parameter, but we expect that it will be "close" to it. To be more specific, we sometimes try to find an interval about the point estimator in which we can be highly confident that the parameter lies. Such an interval is called an *interval estimator.*

Definition *An interval estimator of a population parameter is an interval that is predicted to contain the parameter. The confidence we ascribe to the interval is the probability that it will contain the parameter.*

$$\mu = \bar{X} \pm Z_{\alpha/2}\sigma/\sqrt{n}$$

Interval Estimator

Interval Estimator of the Mean of a normal population with known population variance

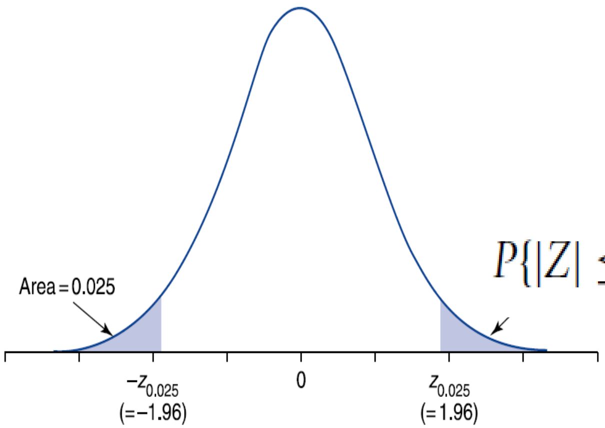
Let $X_1, \ldots, X_n$ be a sample of size n from a normal population having known standard deviation σ, and suppose we want to utilize this sample to obtain a 95 percent confidence interval estimator for the population mean μ. To obtain such an interval, we start with the sample mean $\overline{X}$, which is the point estimator

of μ. We now make use of the fact that $\overline{X}$ is normal with mean μ and standard deviation $\sigma/\sqrt{n}$, which implies that the standardized variable

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \sqrt{n}\frac{\overline{X} - \mu}{\sigma}$$

has a standard normal distribution.

Interval Estimator



$$P\{|Z| \le 1.96\} = P\{-1.96 \le Z \le 1.96\} = 0.95.$$

Now, since $z_{0.025}$ *= 1.96, it follows that* 95 percent of the time the absolute value of *Z is less than or equal to 1.96*

Thus, we can write

$$P\left\{\frac{\sqrt{n}}{\sigma}\left|\overline{X}-\mu\right| \le 1.96\right\} = 0.95$$

Upon multiplying both sides of the inequality by $\sigma/\sqrt{n}$, we see that the preceding equation is equivalent to

$$P\left\{\left|\overline{X}-\mu\right| \le 1.96\frac{\sigma}{\sqrt{n}}\right\} = 0.95$$

Interval Estimator

From the preceding statement we see that, with 95 percent probability, μ and $\overline{X}$ will be within $1.96\sigma/\sqrt{n}$ of each other. But this is equivalent to stating that

$$P\left\{\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right\} = 0.95$$

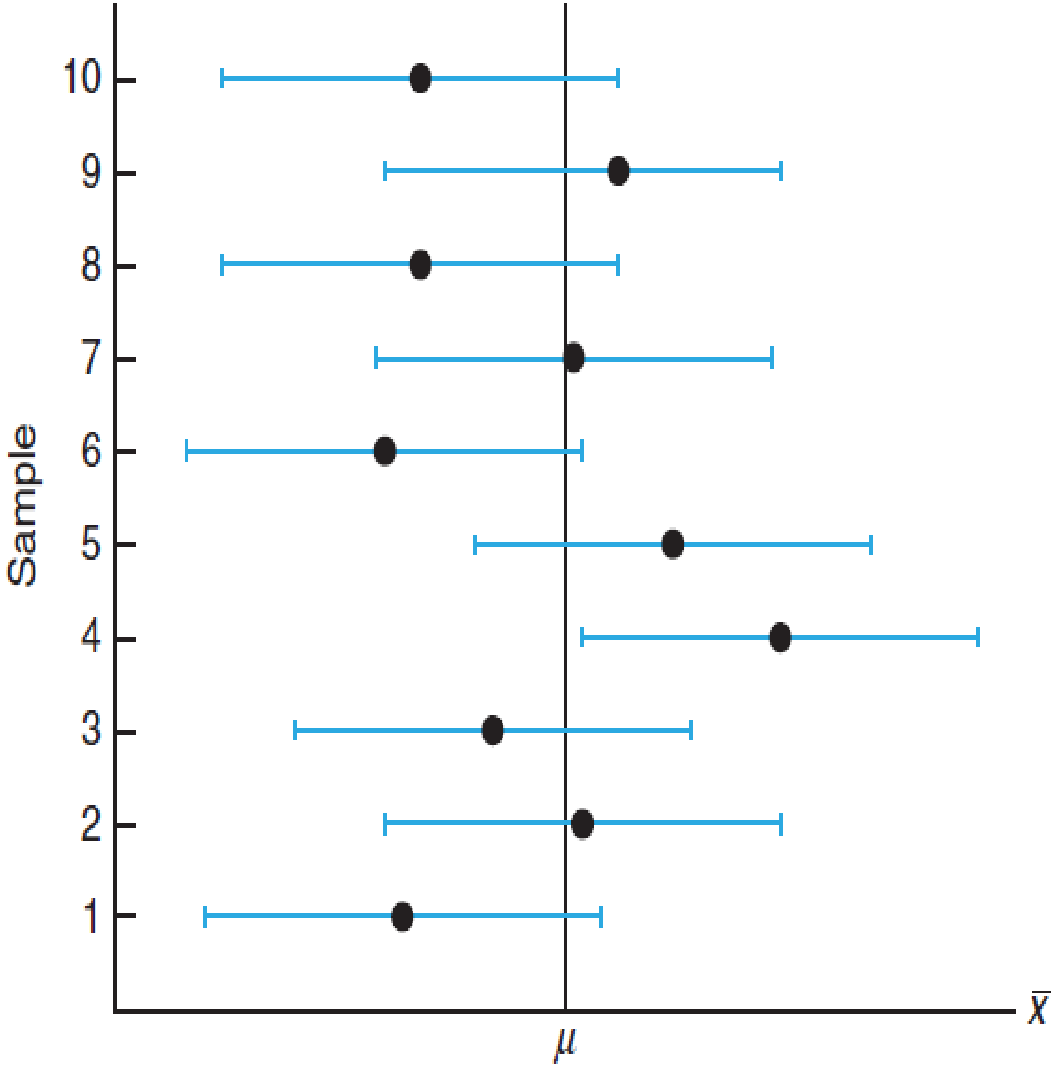
That is, with 95 percent probability, the interval $\overline{X} \pm 1.96\sigma/\sqrt{n}$ will contain the population mean.

The interval from $\overline{X} - 1.96\sigma/\sqrt{n}$ to $\overline{X} + 1.96\sigma/\sqrt{n}$ is said to be a 95 *percent confidence interval estimator* of the population mean μ. If the observed value of $\overline{X}$ is $\overline{x}$, then we call the interval $\overline{x} \pm 1.96\sigma/\sqrt{n}$ a 95 *percent confidence interval estimate* of μ.

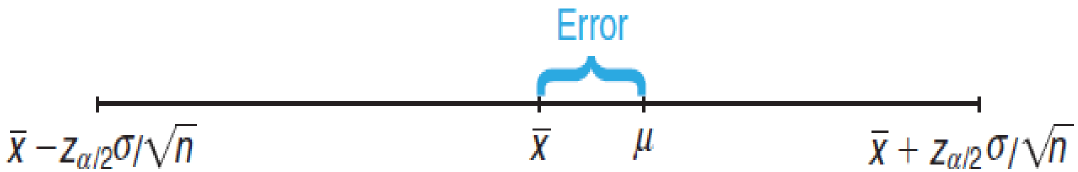
In the long run, 95 percent of the interval estimates so constructed will contain the mean of the population from which the sample is drawn.

$$\mu = \bar{X} \pm Z_{\alpha/2}\sigma/\sqrt{n}$$

Interval Estimator



Interval Estimator



If $\bar{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error will not exceed $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

Suppose that if a signal having intensity μ originates at location A, then the intensity recorded at location B is normally distributed with mean μ and standard deviation 3. That is, due to "noise," the intensity recorded differs from the actual intensity of the signal by an amount that is normal with mean 0 and standard deviation 3. To reduce the error, the same signal is independently recorded 10 times. If the successive recorded values are

$$17, 21, 20, 18, 19, 22, 20, 21, 16, 19$$

construct a 95 percent confidence interval for μ, the actual intensity.

The value of the sample mean is

$$\frac{17 + 21 + 20 + 18 + 19 + 22 + 20 + 21 + 16 + 19}{10} = 19.3$$

Since σ = 3, it follows that a 95 percent confidence interval estimate of μ is given by

$$\mathbf{19.3 \pm 1.96\,(3/\sqrt{10}) = 19.3 \pm 1.86}$$

That is, we can assert with 95 percent confidence that the actual intensity of the signal lies between 17.44 and 21.16

For any value of α between 0 and 1, the probability that a standard normal lies in the interval between $-z_{\alpha/2}$ *and* $z_{\alpha/2}$ *is equal to 1 – α*

$$P\left\{\frac{\sqrt{n}}{\sigma}\left|\overline{X}-\mu\right| \leq z_{\alpha/2}\right\} = 1-\alpha$$

By the same logic used previously when $\alpha = 0.05(z_{0.025} = 1.96)$, we can show that, with probability $1-\alpha$, μ will lie in the interval $\overline{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$.

The interval $\overline{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$ is called a $100(1-\alpha)$ *percent confidence interval estimator* of the population mean.

Table 8.1 Confidence Level Percentiles

Confidence level $100(1-\alpha)$	Corresponding value of α	Value of $z_{\alpha/2}$
90	0.10	$z_{0.05} = 1.645$
95	0.05	$z_{0.025} = 1.960$
99	0.01	$z_{0.005} = 2.576$

For $n = 10$, $\bar{X} = 19.3$, and $\sigma = 3$, calculate 90 and 99 percent confidence interval estimate of μ

We are being asked to construct a $100(1 - \alpha)$ confidence interval estimate, with $\alpha = 0.10$ in part (a) and $\alpha = 0.01$ in part (b). Now

$$z_{0.05} = 1.645 \quad \text{and} \quad z_{0.005} = 2.576$$

and so the 90 percent confidence interval estimator is

$$\overline{X} \pm 1.645 \frac{\sigma}{\sqrt{n}} \quad = 19.3 \pm 1.56$$

and the 99 percent confidence interval estimator is

$$\overline{X} \pm 2.576 \frac{\sigma}{\sqrt{n}} \quad = 19.3 \pm 2.44$$

Note that the larger the confidence coefficient $100(1 - \alpha)$, the larger the length of this interval.

For instance, suppose we want to determine an interval of length at most *b* *that, with 95 percent certainty, contains the* population mean. How large a sample is needed?

To answer this, note that since $z_{0.025}$ *= 1.96, a 95 percent confidence interval for μ based on a sample of size n* is

$$\overline{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

Since the length of this interval is

$$\text{Length of interval} = 2(1.96)\frac{\sigma}{\sqrt{n}} = 3.92\frac{\sigma}{\sqrt{n}}$$

we must choose n so that

$$\frac{3.92\sigma}{\sqrt{n}} \leq b$$

or, equivalently,

$$\sqrt{n} \geq \frac{3.92\sigma}{b}$$

Upon squaring both sides we see that the sample size n must be chosen so that

$$n \geq \left(\frac{3.92\sigma}{b}\right)^2$$

If the population standard deviation is σ = 2 and we want a 95 percent confidence interval estimate of the mean μ that is of size less than or equal to *b = 0.1, how large a sample is needed?*

We have to select a sample of size n, where

$$n \geq \left(\frac{3.92 \times 2}{0.1}\right)^2 = (78.4)^2 = 6146.6$$

That is, a sample of size 6147 or larger is needed.

This indicates that for a population with standard deviation is σ = 2, we need a sample size of 6147 to be 95 percent confident that the mean lies within a value of actual mean ± 0.05.

Determining the Necessary Sample Size

The length of the $100(1-\alpha)$ percent confidence interval estimator of the population mean will be less than or equal to b when the sample size n satisfies

$$n \geq \left(\frac{2z_{\alpha/2}\sigma}{b}\right)^2$$

The confidence interval estimator is

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Thus, the standard deviation of $\bar{X}$, or *standard error* of $\bar{X}$, is $\sigma/\sqrt{n}$. Simply put, the standard error of an estimator is its standard deviation. For $\bar{X}$, the computed confidence limit

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \text{ is written as } \bar{x} \pm z_{\alpha/2}\ \text{s.e.}(\bar{x}),$$

In the case where σ is known and sampling is from a normal distribution, the confidence limits on μ are

$$\bar{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \ is\ written\ as\ \bar{x} \pm Z_{\alpha/2}\, s.e\,(\bar{x})$$

Sampling Statistics

The standard deviation of a random variable, which is equal to the square root of its variance, is a direct indicator of the spread in the distribution. It follows from the identity

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

that $\mathrm{SD}(\overline{X})$, the standard deviation of the sample mean $\overline{X}$, is given by

$$\mathrm{SD}(\overline{X}) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

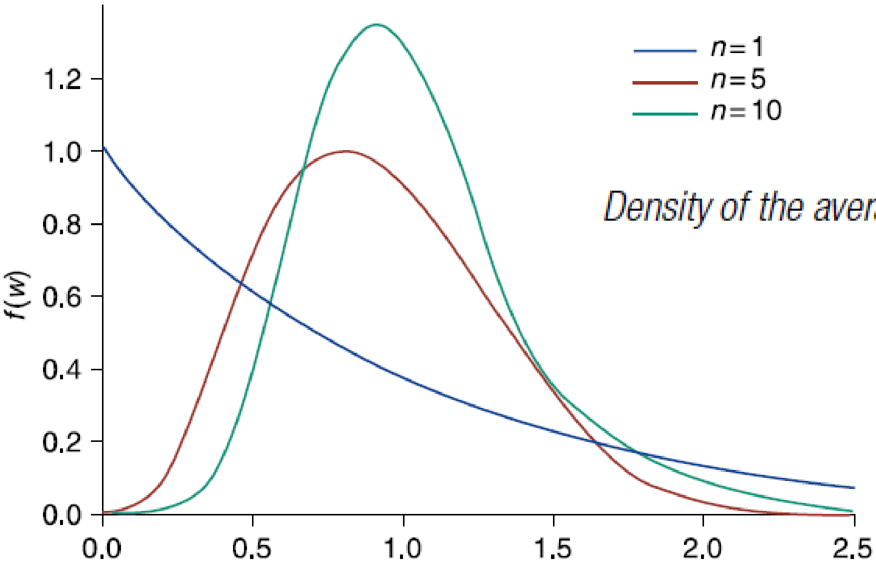
In the preceding formula, σ is the population standard deviation, and n is the sample size.

The *standard deviation* of the sample mean is equal to the population standard deviation divided by the square root of the sample size.

Sampling Proportions

How Large a Sample Is Needed?

The central limit theorem leaves open the question of how large the sample size *n needs to be for the normal approximation to be valid, and indeed the answer* depends on the population distribution of the sample data. For instance, if the underlying population distribution is normal, then the sample mean *will also* be normal, no matter what the sample size is. **A general rule of thumb is that you can be confident of the normal approximation whenever the sample size *n is at* least 30**. That is, practically speaking, no matter how non normal the underlying population distribution is, the sample mean of a sample size of at least 30 will be approximately normal. In most cases the normal approximation is valid for much smaller sample sizes. Indeed, usually a sample size of 5 will suffice for the approximation to be valid.



Density of the average of n exponential random variables.

Point Estimator of a Population Proportion

Suppose that we are trying to estimate the proportion of a large population that is in favour of a given proposition.

Let *p denote the unknown proportion.*

To estimate p, a random sample should be chosen, and then p should be estimated by the proportion of the sample that is in favour. Calling the estimator $\hat{p}$,

we can express it by
$$\hat{p} = \frac{X}{n}$$

where** X is the number of members of the sample who are in favour of the* **proposition and** *n is the size of the sample.

$$E[\hat{p}] = p$$

$$\mathrm{SD}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$$

The standard deviation of $\hat{p}$ is also called the *standard error* of $\hat{p}$

Point Estimator of a Population Proportion

Since $$p(1-p) \leq \frac{1}{4}$$

it follows that

$$\mathrm{SD}(\hat{p}) \leq \sqrt{\frac{1}{4n}} = \frac{1}{2\sqrt{n}}$$

For instance, suppose a random sample of size 900 is chosen.

Then no matter what proportion of the population is actually in favour of the proposition,

it follows that the standard error of the estimator of this proportion is less than or equal to

$$1/(2\sqrt{900}) = 1/60.$$