21DS636-SM

Probability Theory

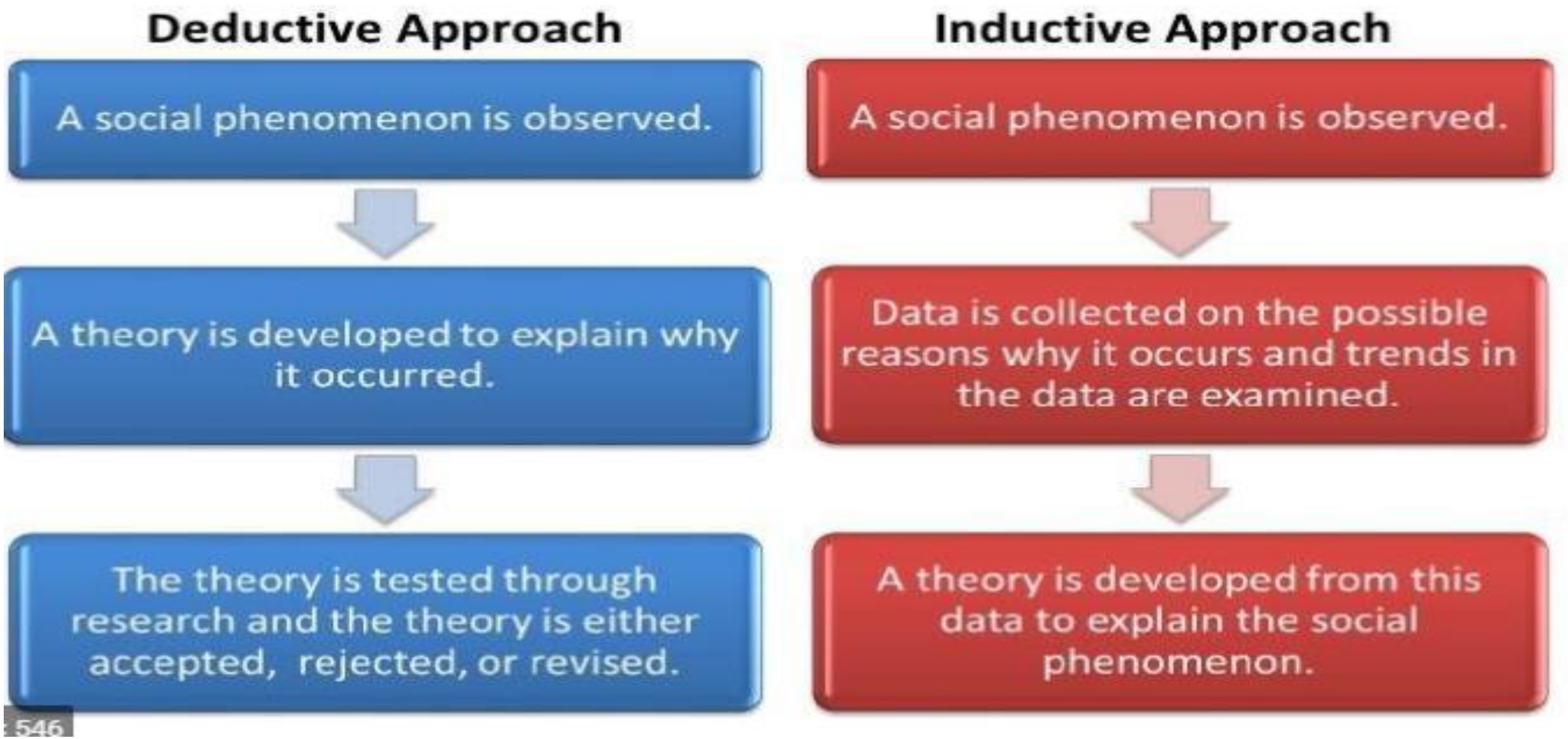
Sampling Distribution Models and the Central Limit Theorem

Dr. Deepa Gupta, professor
Human Language Technology Lab
Department of Computer Science & Engineering
ASE-Bangalore

Sampling Distribution Models and the Central Limit Theorem

Transition from Data Analysis and Probability to Statistics

Deductive and Inductive Approaches

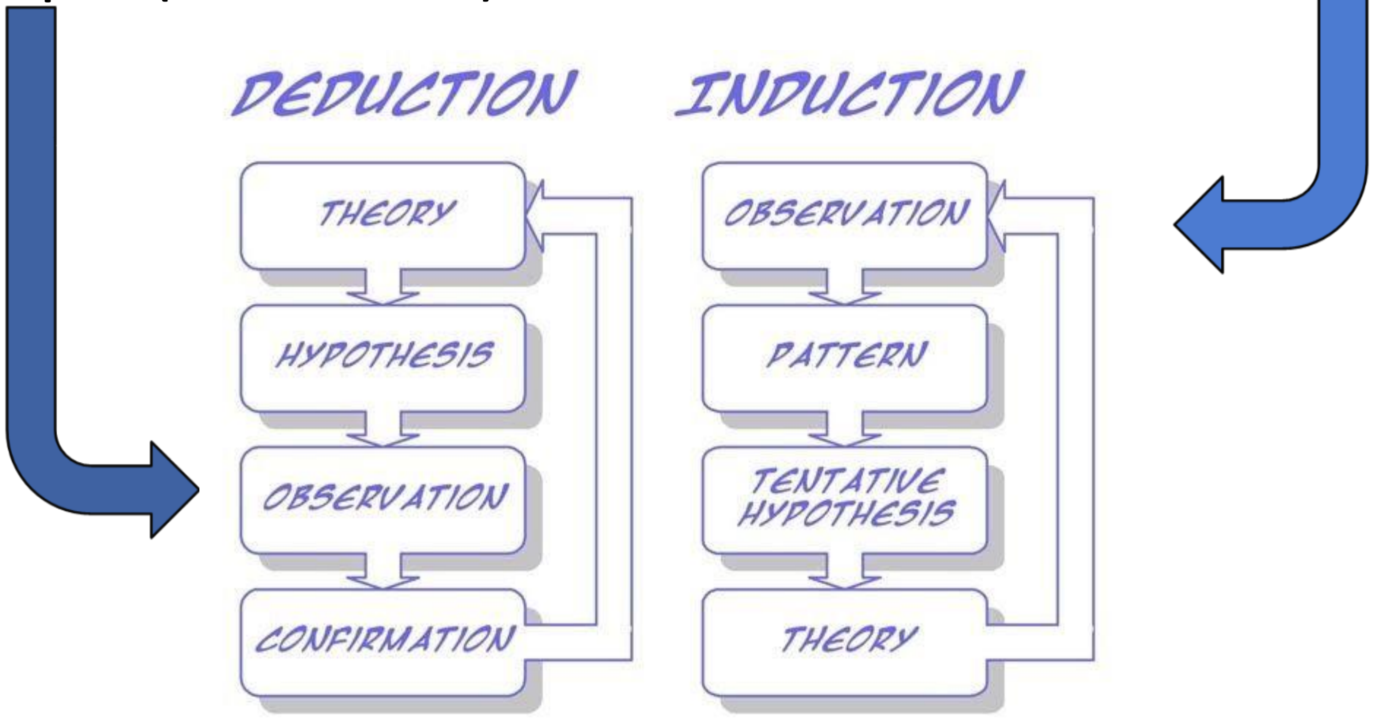


Probability:

- From population to sample (deduction)

Statistics:

- From sample to the population (induction)

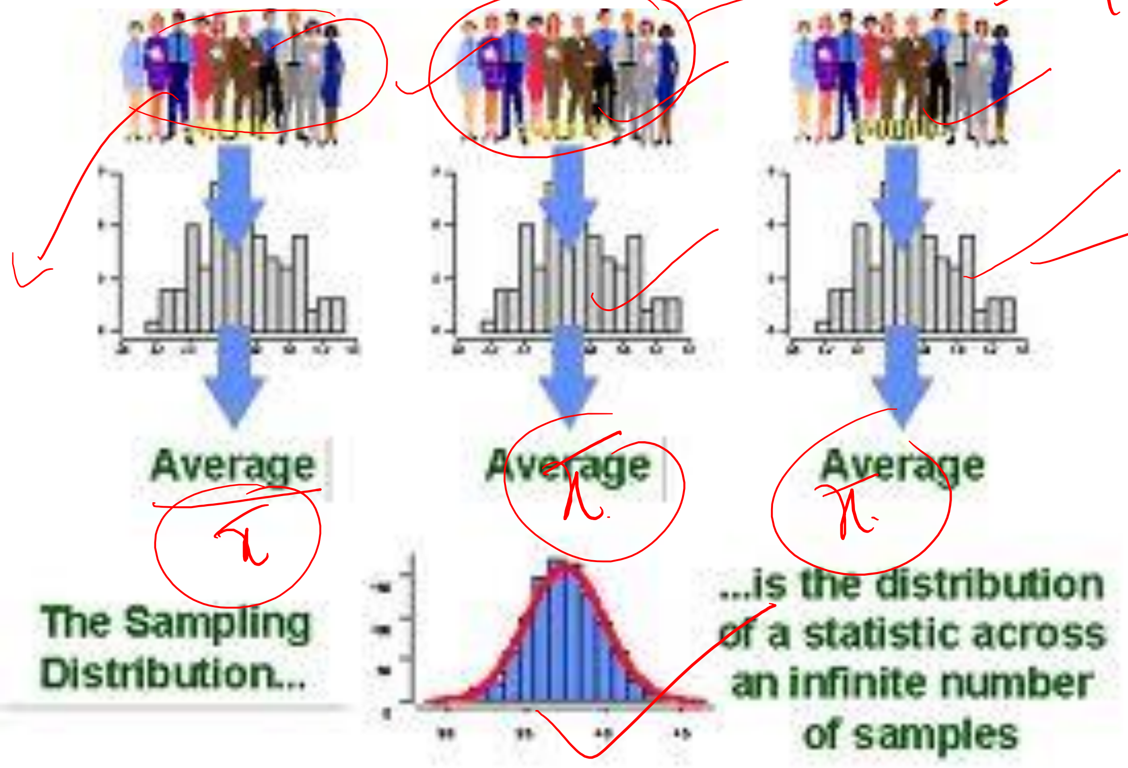


Basic Terminology

- **Inferential statistics**: the part of statistics that allows researchers to generalize their findings beyond data collected.
- **Statistical inference:** a procedure for making inferences or generalizations about a larger population from a sample of that population
  - **Population:** any collection of entities that have at least one characteristic in common
  - **Parameter:** the numbers that describe characteristics of scores in the population (mean, variance, s.d., etc.)
  - **Sample:** a part of the population
  - **Statistic:** the numbers that describe characteristics of scores in the sample (mean, variance, s.d., correlation coefficient, reliability coefficient, etc.)

Sampling Distribution

- The sampling distribution of a sample statistic calculated from a sample of n measurements is the **probability distribution** of values taken by the statistic in all possible samples of size n taken from the same population



Based on all possible samples of size n.

100

n = 10

N = 100

n = 10

5.49

5.11

S.3

5.71

5.49

Sampling Distributions

- Suppose that we draw all possible samples of size n from a given population.

μ p σ^2 σ Population Parameter

- We compute a statistic (e.g., a mean, proportion, standard deviation) for each sample.

$\bar{x}$ $\hat{p}$ s s^2

- The probability distribution of this statistic is called a sampling distribution.

- The standard deviation of this statistic is called the standard error.

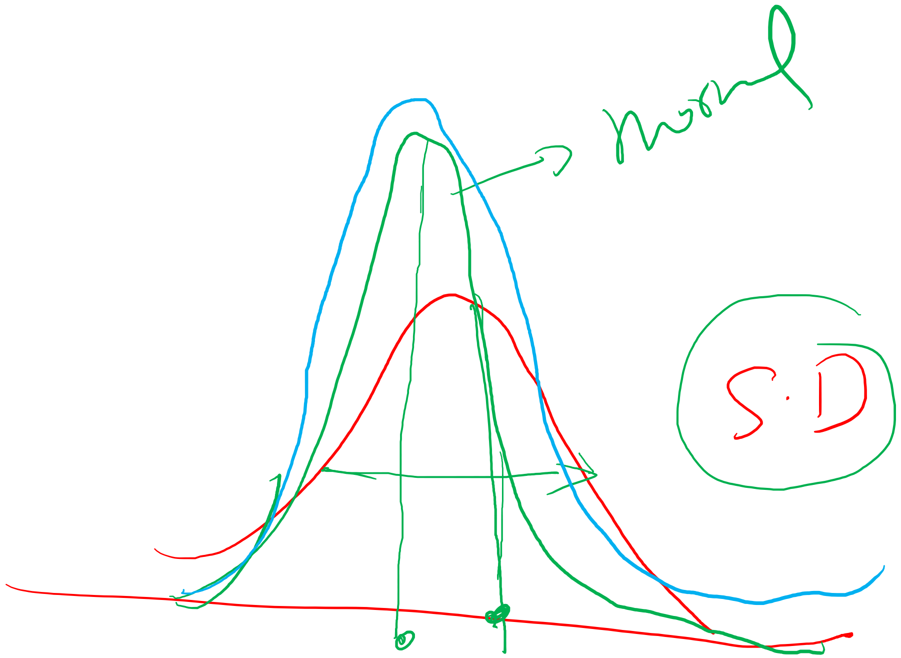
Parameters & Statistics

- In real life, parameters of populations are unknown and unknowable.
  - For example: the mean height of US adult (18+) men is unknown and unknowable
- Rather than investigating the whole population, we take a sample, calculate a statistic related to the parameter of interest, and make an inference.

$\bar{x}$ Sample data μ

Variability of a Sampling Distribution

- The variability of a sampling distribution is measured by its <u>variance</u> or its <u>standard deviation</u>. The variability of a sampling distribution depends on three factors:
  - N: The number of observations in the population.
  - n: The number of observations in the sample.
  - The way that the random sample is chosen.



Sampling Distributions

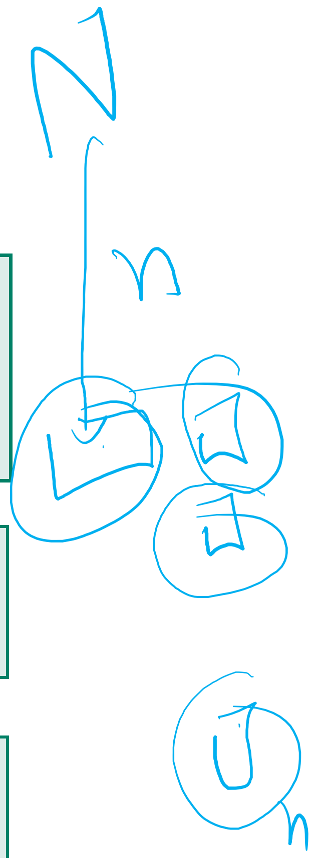
Statistical inference is concerned with making decisions about a population based on the information contained in a random sample from that population.

Definitions:

The random variables $X_1, X_2, \ldots, X_n$ are a **random sample** of size n if (a) the X_i's are independent random variables, and (b) every X_i has the same probability distribution.

A **statistic** is any function of the observations in a random sample.

The probability distribution of a statistic is called a **sampling distribution.**



Sampling Distribution Models of Sample Means

AMRITA VISHWA VIDYAPEETHAM

Population Parameter of Frequent Interest:

<u>The Population Mean μ</u>

P sample $\bar{x}$

- To estimate the unknown value of μ, the sample mean $\bar{x}$ is often used.

$\bar{x}$ $\bar{x}$ - - $\bar{x}$

- We need to examine the Sampling Distribution of the Sample Mean $\bar{x}$

$\bar{X}$

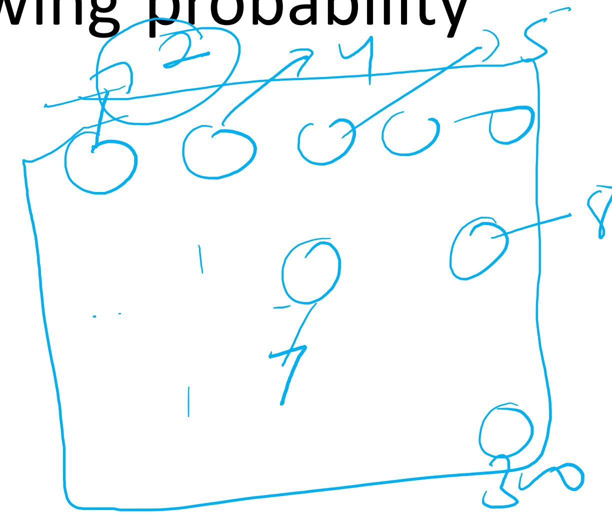
(the probability distribution of all possible values of $\bar{x}$ based on a sample of size n).

Example

- Professor Stickler has a large statistics class of over 300 students. He asked them the ages of their cars and obtained the following probability distribution:

P.D

x	2	3	4	5	6	7	8
p(x)	1/14	1/14	2/14	2/14	2/14	3/14	3/14



- Random Sample n=2 is to be drawn from pop.

- Find the sampling distribution of the sample mean $\overline{x}$ for samples of size n = 2.

Solution

2 — 8

- 7 possible ages (ages 2 through 8)

7 7

7^2

- Total of 7^2=49 possible samples of size 2
- All 49 possible samples with the corresponding sample means and probabilities are on the next slide

All 49 possible samples of size n = 2

x	2	3	4	5	6	7	8
p(x)	1/14	1/14	2/14	2/14	2/14	3/14	3/14

Population: ages of cars and their distribution

$\frac{1}{14} \times \frac{2}{14} = \frac{2}{196}$

$\frac{2+2}{2} = 2$ $\quad \frac{2+4}{2} = \frac{6}{2} = 3$

Sample	2,2	2,4	2,6	2,8	2,5	2,3	2,7	4,2	4,4	4,6	4,8	4,5	4,3	4,7	6,2	6,4	6,6
xbar	2	3	4	5	3.5	2.5	4.5	3	4	5	6	4.5	3.5	5.5	4	5	6
Prob	$\frac{1}{196}$	$\frac{2}{196}$	$\frac{2}{196}$	$\frac{3}{196}$	$\frac{2}{196}$	$\frac{1}{196}$	$\frac{3}{196}$	$\frac{2}{196}$	$\frac{4}{196}$	$\frac{4}{196}$	$\frac{6}{196}$	$\frac{4}{196}$	$\frac{2}{196}$	$\frac{6}{196}$	$\frac{2}{196}$	$\frac{4}{196}$	$\frac{4}{196}$

Sample	6,8	6,5	6,3	6,7	8,2	8,4	8,6	8,8	8,5	8,3	8,7	5,2	5,4	5,6	5,8	5,5
xbar	7	5.5	4.5	6.5	5	6	7	8	6.5	5.5	7.5	3.5	4.5	5.5	6.5	5
Prob	$\frac{6}{196}$	$\frac{4}{196}$	$\frac{2}{196}$	$\frac{6}{196}$	$\frac{3}{196}$	$\frac{6}{196}$	$\frac{6}{196}$	$\frac{9}{196}$	$\frac{6}{196}$	$\frac{3}{196}$	$\frac{9}{196}$	$\frac{2}{196}$	$\frac{4}{196}$	$\frac{4}{196}$	$\frac{6}{196}$	$\frac{4}{196}$

Sample	5,3	5,7	3,2	3,4	3,6	3,8	3,5	3,3	3,7	7,2	7,4	7,6	7,8	7,5	7,3	7,7
xbar	4	6	2.5	3.5	4.5	5.5	4	3	5	4.5	5.5	6.5	7.5	6	5	7
Prob	$\frac{2}{196}$	$\frac{6}{196}$	$\frac{1}{196}$	$\frac{2}{196}$	$\frac{2}{196}$	$\frac{3}{196}$	$\frac{2}{196}$	$\frac{1}{196}$	$\frac{3}{196}$	$\frac{3}{196}$	$\frac{6}{196}$	$\frac{6}{196}$	$\frac{9}{196}$	$\frac{6}{196}$	$\frac{3}{196}$	$\frac{9}{196}$

$P(X = 2, 4) = P(X=2) \cdot P(X=4) = \frac{1}{14} \times \frac{2}{14} = \frac{2}{196}$

P.D

24

49

42
24

7×7

49

34

43

Probability Distribution of the Sample Mean Age of 2 Cars

Sample	2,2	2,4	2,6	2,8	2,5	2,3	2,7	4,2	4,4	4,6	4,8	4,5	4,3	4,7	6,2	6,4	6,6
xbar	2	3	4	5	3.5	2.5	4.5	3	4	5	6	4.5	3.5	5.5	4	5	6
Prob	$\frac{1}{196}$	$\frac{2}{196}$	$\frac{2}{196}$	$\frac{3}{196}$	$\frac{2}{196}$	$\frac{1}{196}$	$\frac{3}{196}$	$\frac{2}{196}$	$\frac{4}{196}$	$\frac{4}{196}$	$\frac{6}{196}$	$\frac{4}{196}$	$\frac{2}{196}$	$\frac{6}{196}$	$\frac{2}{196}$	$\frac{4}{196}$	$\frac{4}{196}$
Sample	6,8	6,5	6,3	6,7	8,2	8,4	8,6	8,8	8,5	8,3	8,7	5,2	5,4	5,6	5,8	5,5	
xbar	7	5.5	4.5	6.5	5	6	7	8	6.5	5.5	7.5	3.5	4.5	5.5	6.5	5	
Prob	$\frac{6}{196}$	$\frac{4}{196}$	$\frac{2}{196}$	$\frac{6}{196}$	$\frac{3}{196}$	$\frac{6}{196}$	$\frac{6}{196}$	$\frac{9}{196}$	$\frac{6}{196}$	$\frac{3}{196}$	$\frac{9}{196}$	$\frac{2}{196}$	$\frac{4}{196}$	$\frac{4}{196}$	$\frac{6}{196}$	$\frac{4}{196}$	
Sample	5,3	5,7	3,2	3,4	3,6	3,8	3,5	3,3	3,7	7,2	7,4	7,6	7,8	7,5	7,3	7,7	
xbar	4	6	2.5	3.5	4.5	5.5	4	3	5	4.5	5.5	6.5	7.5	6	5	7	
Prob	$\frac{2}{196}$	$\frac{6}{196}$	$\frac{1}{196}$	$\frac{2}{196}$	$\frac{2}{196}$	$\frac{3}{196}$	$\frac{2}{196}$	$\frac{1}{196}$	$\frac{3}{196}$	$\frac{3}{196}$	$\frac{6}{196}$	$\frac{6}{196}$	$\frac{9}{196}$	$\frac{6}{196}$	$\frac{3}{196}$	$\frac{9}{196}$	

Sampling dist.

$\bar{x}$	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8
$p(\bar{x})$	1/196	2/196	5/196	8/196	12/196	18/196	24/196	26/196	28/196	24/196	21/196	18/196	9/196

Example (cont.)

- Population probability dist.

x	2	3	4	5	6	7	8
$p(x)$	1/14	1/14	2/14	2/14	2/14	3/14	3/14

- Sampling dist. of $\overline{x}$

$\overline{x}$	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8
$p(\overline{x})$	1/196	2/196	5/196	8/196	12/196	18/196	24/196	26/196	28/196	24/196	21/196	18/196	1/196

Population probability dist.

$2 \times \frac{1}{14}$

x	2	3	4	5	6	7	8
p(x)	1/14	1/14	2/14	2/14	2/14	3/14	3/14

$\sum x_i \, p(x_i)$

E(X)=2(1/14)+3(1/14)+4(2/14)+ … +8(3/14)=5.714

Population mean E(X)=μ = 5.714

5.714

Sampling dist. of $\bar{x}$

x	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8
p(x)	1/196	2/196	5/196	8/196	12/196	18/196	24/196	26/196	28/196	24/196	21/196	18/196	1/196

$P(\bar{X}=x)$

E($\bar{X}$)=2(1/196)+2.5(2/196)+3(5/196)+3.5(8/196)+4(12/196)+4.5(18/196)+5(24/196) +5.5(26/196)+6(28/196)+6.5(24/196)+7(21/196)+7.5(18/196)+8(1/196) = 5.714

$E[\bar{X}] = \sum \bar{x}_i \, P(\bar{X} = \bar{x}_i)$

Mean of sampling distribution of $\bar{x}$: E($\bar{X}$) = 5.714

5.714

Example (cont.)

Population from which sample is selected:

$$\mu = E(X) = 2(\tfrac{1}{14}) + 3(\tfrac{1}{14}) + 4(\tfrac{2}{14}) + \cdots + 8(\tfrac{3}{14}) = \mathbf{5.714}$$

$$\sigma^2 = Var(X) = \mathbf{3.4898}$$

$$\sigma = SD(X) = \sqrt{Var(X)} = \sqrt{3.4898} = \mathbf{1.8681}$$

●●●●●●●●●●●●●●●●●●●●

Sampling dist. of $\bar{X}$:

$$E(\bar{X}) = 2(\tfrac{1}{196}) + 2.5(\tfrac{2}{196}) + \ldots + 8(\tfrac{9}{196}) = \mathbf{5.714}$$

$$Var(\bar{X}) = 1.7449 = \frac{\mathbf{3.4898}}{2} = \frac{Var(X)}{2}$$

$$SD(\bar{X}) = \sqrt{Var(\bar{X})} = \sqrt{\frac{Var(X)}{2}} = \frac{SD(X)}{\sqrt{2}} = \frac{\mathbf{1.8681}}{\sqrt{2}} = 1.3209$$

$$V(X) = E[X^2] - (E[X])^2$$

$$V(\bar{X}) = V(Y)$$

Sampling Distribution of the Sample Mean $\overline{X}$: Example

- A fair 6-sided die is thrown; let X represent the number of dots showing on the upper face.
- The probability distribution of X is

x	1	2	3	4	5	6
p(x)	1/6	1/6	1/6	1/6	1/6	1/6

Population mean μ:

$\mu = E(X) = 1(1/6) + 2(1/6) + 3(1/6) + \ldots\ldots\ldots$

$= 3.5.$

Population variance σ^2

$\sigma^2 = V(X) = (1-3.5)^2(1/6) + (2-3.5)^2(1/6) +$

$\ldots\ldots\ldots$

$\ldots\ldots\ldots\ldots = 2.92$

- Suppose we want to estimate μ from the mean $\bar{x}$ of a sample of size n = 2.
- What is the sampling distribution of $\bar{x}$ in this situation?

1,1

Sample		Mean	Sample		Mean	Sample		Mean
1	1,1	1	13	3,1	2	25	5,1	3
2	1,2	1.5	14	3,2	2.5	26	5,2	3.5
3	1,3	~~2~~	15	3,3	3	27	5,3	4
4	1,4	~~2.5~~	16	3,4	3.5	28	5,4	4.5
5	1,5	~~3~~	17	3,5	4	29	5,5	5
6	1,6	~~3.5~~	18	3,6	4.5	30	5,6	5.5
7	2,1	~~1.5~~	19	4,1	2.5	31	6,1	3.5
8	2,2	2	20	4,2	3	32	6,2	4
9	2,3	2.5	21	4,3	3.5	33	6,3	4.5
10	2,4	3	22	4,4	4	34	6,4	5
11	2,5	3.5	23	4,5	4.5	35	6,5	5.5
12	2,6	4	24	4,6	5	36	6,6	6

2×3

36

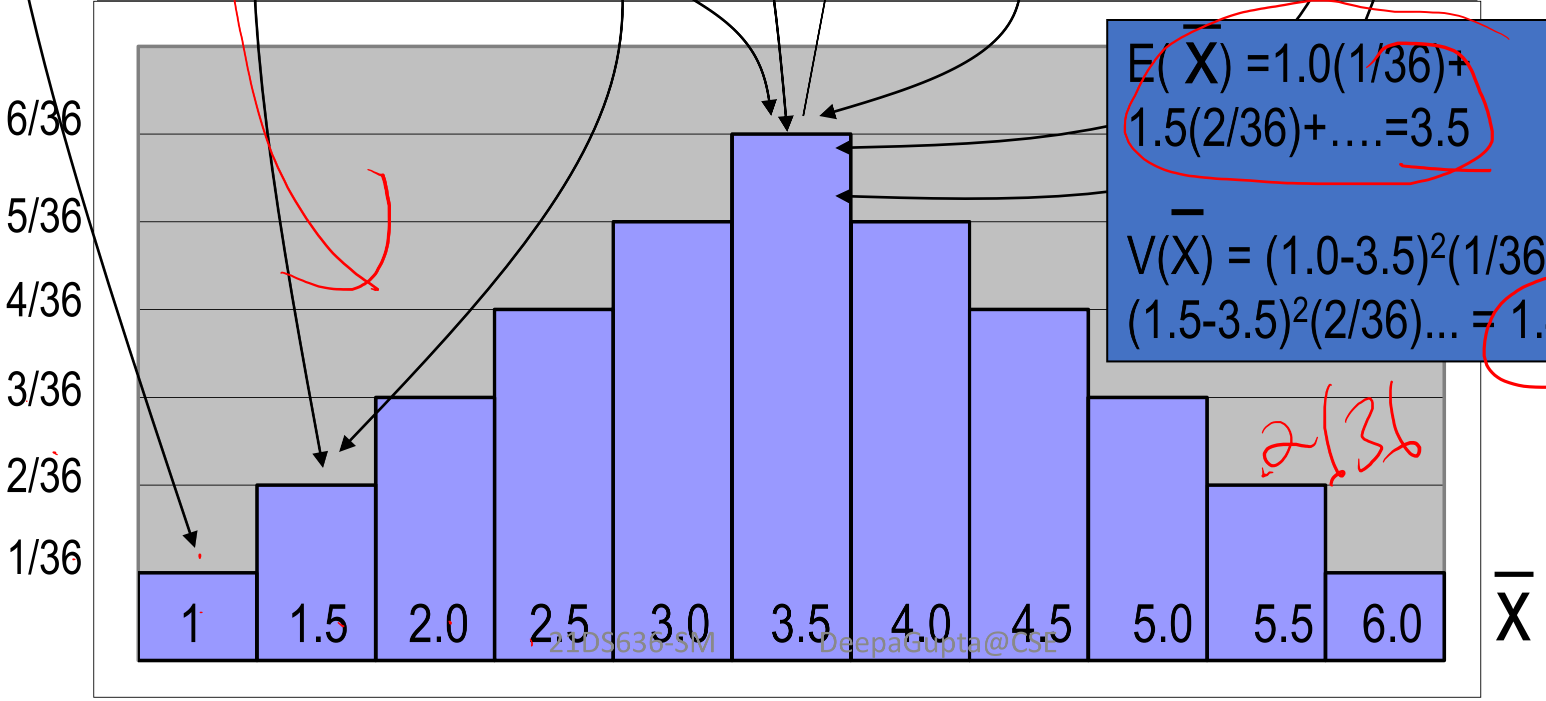
6×6

Sample		Mean	Sample		Mean	Sample		Mean
1	1,1	1	13	3,1	2	25	5,1	3
2	1,2	1.5	14	3,2	2.5	26	5,2	3.5
3	1,3	2	15	3,3	3	27	5,3	4
9	2,3	2.5	21	4,3	3.5	33	6,3	4.5
10	2,4	3	22	4,4	4	34	6,4	5
11	2,5	3.5	23	4,5	4.5	35	6,5	5.5
12	2,6	4	24	4,6	5	36	6,6	6

$\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

$$Note:\ E(\bar{X}) = E(X)\ and\ \ Var(\bar{X}) = \frac{Var(X)}{2}$$

5.5



$E(\bar{X}) = 1.0(1/36) + 1.5(2/36) + \ldots. = 3.5$

$V(\bar{X}) = (1.0-3.5)^2(1/36) + (1.5-3.5)^2(2/36)\ldots = 1.46$

$E(\bar{X})$

$= E[X - (E(X))]^2$

S.D

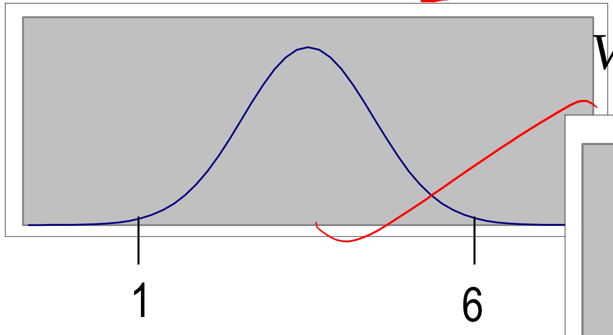
1.41 $\quad \frac{2.92}{2}$

1.46

$n = 5$

$E(\bar{X}) = 3.5$

$Var(\bar{X}) = .5833 \; (= \dfrac{Var(X)}{5})$

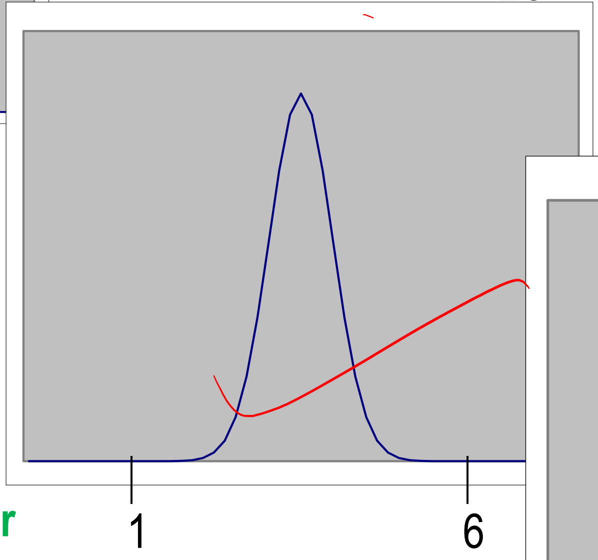


$3.5 - \sqrt{0.5831} \leq \mu < 3.5 + \sqrt{0.583}$

$n = 10$

$E(\bar{X}) = 3.5$

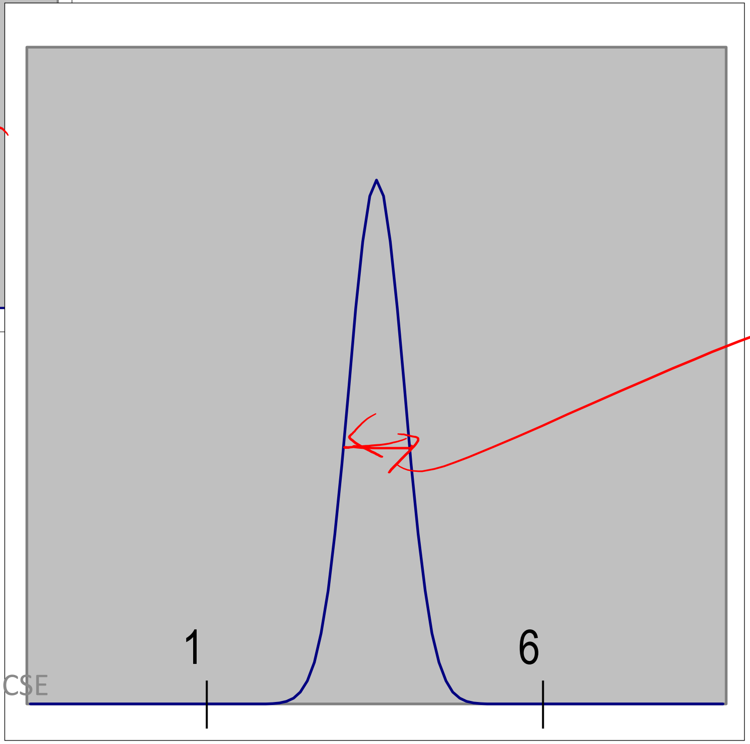
$Var(\bar{X}) = .2917 \; (= \dfrac{Var(X)}{10})$



$n = 25$

$E(\bar{X}) = 3.5$

$Var(\bar{X}) = .1167 \; (= \dfrac{Var(X)}{25})$



Notice that $Var(\bar{X})$ is smaller than *Var(X)*. The larger the sample size the smaller is $Var(\bar{X})$. Therefore, $\bar{X}$ tends to fall closer to μ, as the sample size increases.

Properties of the Sampling Distribution of $\bar{x}$

- $E(\bar{X}) = \mu$
  (the expected value of the sampling distribution of $\bar{X}$= the expected value μ of the sampled population
- $SD(\bar{X}) = \frac{SD(X)}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}}$
  where σ is the standard deviation of the pulation from which the sample is taken and n is the sample size.

$Var(\bar{x}) = \frac{Var(x)}{n}$

Sampling Distribution Models of Sample Proportions

Sampling distribution of p: the sample Example proportion

- If a coin is fair the probability of a head on any toss of the coin is **p = 0.5** (**p** is the population parameter)
- Imagine tossing this fair coin 4 times and calculating the proportion p of the 4 tosses that result in heads (note that $\hat{p}$ = x/4, where x is the number of heads in 4 tosses).
- **Objective**: determine the sampling distribution of $\hat{p}$, the proportion of heads in 4 tosses of a fair coin.

Example: Sampling distribution of $\hat{p}$

Possible values of $\hat{p}: \frac{0}{4}=0, \frac{1}{4}=.25, \frac{2}{4}=.50, \frac{3}{4}=.75, \frac{4}{4}=1$

There are 2^4 = 16 equally likely possible outcomes

(1 =head, 0 =tail)

(1,1,1,1) (1,1,1,0) (1,1,0,1) (1,0,1,1)

(0,1,1,1) (1,1,0,0) (1,0,1,0) (1,0,0,1)

(0,1,1,0) (0,1,0,1) (0,0,1,1) (1,0,0,0)

(0,1,0,0) (0,0,1,0) (0,0,0,1) (0,0,0,0)

$\frac{3/4}{16}$ $\frac{1}{16}$

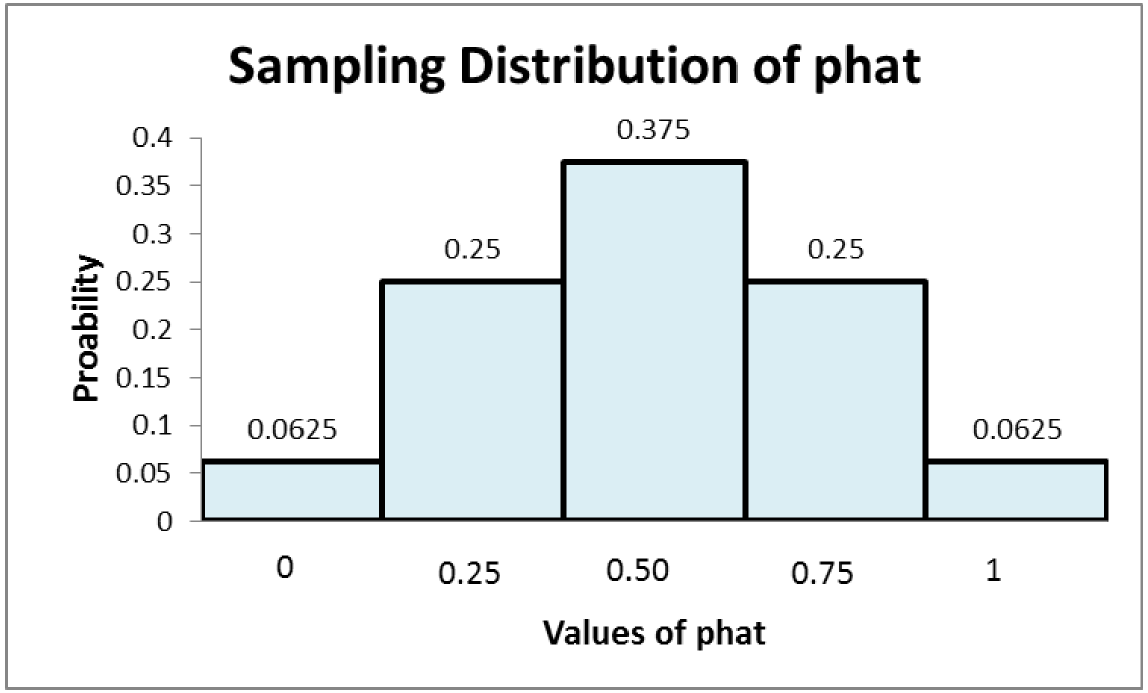
$\hat{p}$	0.0 (0 heads)	0.25 (1 head)	0.50 (2 heads)	0.75 (3 heads)	1.0 (4 heads)
$P(\hat{p})$	1/16= 0.0625	*4/16= 0.25*	*6/16= 0.375*	*4/16= 0.25*	*1/16= 0.0625*

X

P(X)

Sampling distribution of $\hat{p}$

$\hat{p}$	0.0 (0 heads)	0.25 (1 head)	0.50 (2 heads)	0.75 (3 heads)	1.0 (4 heads)
P($\hat{p}$)	1/16= 0.0625	*4/16= 0.25*	*6/16= 0.375*	*4/16= 0.25*	*1/16= 0.0625*



Sampling distribution of $\hat{p}$ (cont.)

$\hat{p}$	0.0 (0 heads)	0.25 (1 head)	0.50 (2 heads)	0.75 (3 heads)	1.0 (4 heads)
P($\hat{p}$)	1/16= 0.0625	*4/16= 0.25*	*6/16= 0.375*	*4/16= 0.25*	*1/16= 0.0625*

$E[X] = \sum x P(x)$

- E($\hat{p}$) =0*.0625+ 0.25*0.25+ 0.50*0.375 +0.75*0.25+ 1.0*0.0625 = 0.5 = p (the prob of heads)

$E(\hat{p})$

- Var($\hat{p}$) = $(0-0.5)^2 * 0.0625 + (.25-0.5)^2 * 0.25 + (0.5-0.5)^2 * 0.375$

$$+ (0.75-0.5)^2 * 0.25 + (1-0.5)^2 * 0.0625$$

$$= 0.0625$$

$V(x) = \sum (x - E(x))^2 \times P(x_i)$

$= E[x^2] - (E[x])^2$

- $$SD(\hat{p}) = \sqrt{Var(\hat{p})} = \sqrt{0.0625} = 0.25$$

- Note that $SD(\hat{p}) = \sqrt{\dfrac{p(1-p)}{n}} = \sqrt{\dfrac{.5*.5}{4}} = \dfrac{.5}{\sqrt{4}} = 0.25$

$\sqrt{\dfrac{p(1-p)}{n}}$

Expected Value and Standard Deviation of the Sampling Distribution of $\hat{p}$

- $E(\hat{p}) = p$
- $SD(\hat{p}) = \sqrt{\dfrac{p(1-p)}{n}}$

where p is the “success” probability in the sampled population and n is the sample size

P.P

$\mu \longrightarrow \bar{x}$

$\bar{X} \sim D$

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

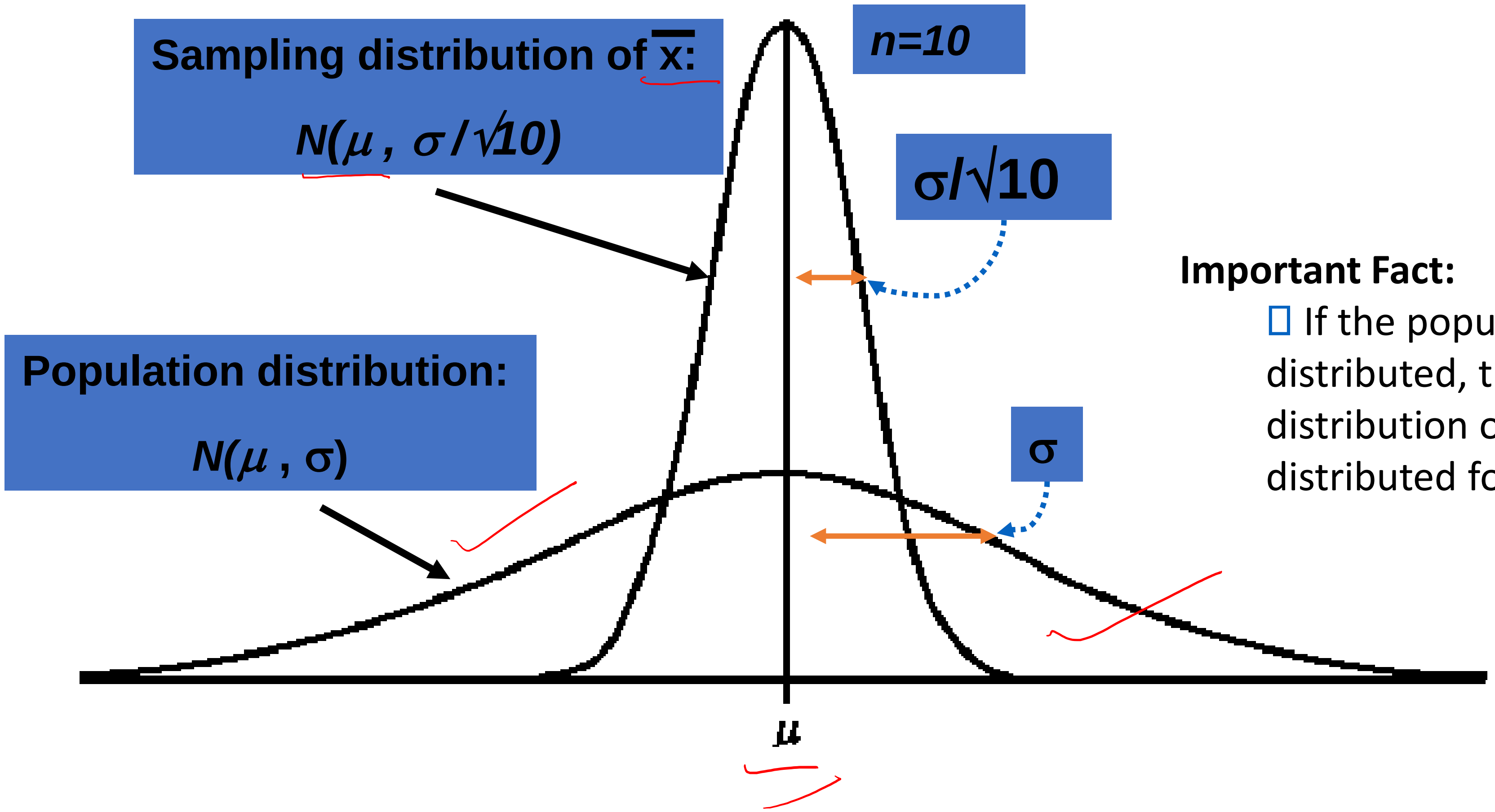
$V(\bar{X}) = \frac{\sigma}{\sqrt{n}}$

THE CENTRAL LIMIT THEOREM

The “World is Normal” Theorem

AMRITA VISHWA VIDYAPEETHAM

But first,…Sampling Distribution of $\bar{x}$- Normally Distributed Population

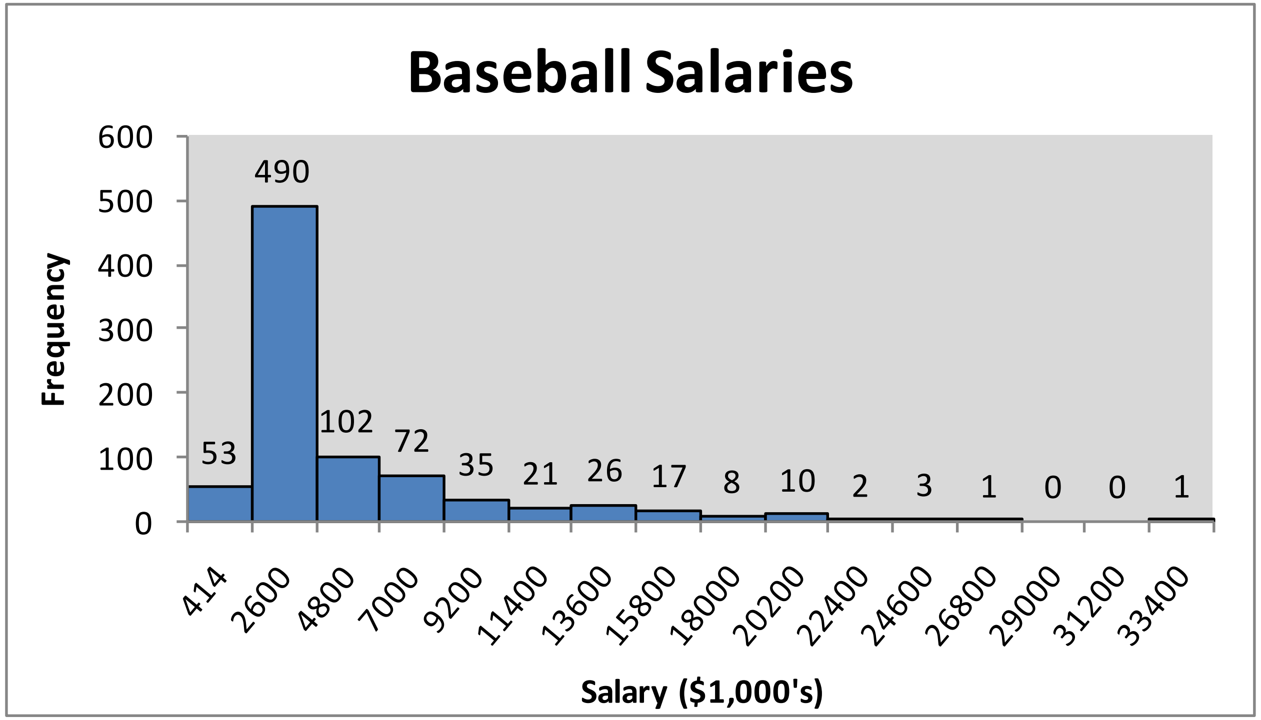


Important Fact:

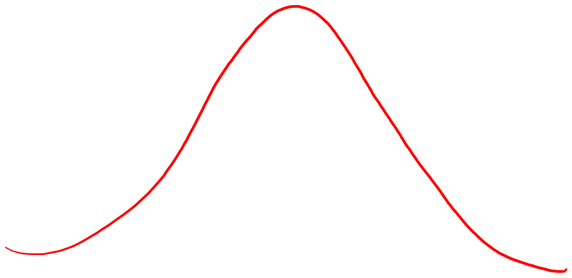
- If the population is normally distributed, then the sampling distribution of $\bar{x}$ is normally distributed for any sample size n.

Non-normal Populations

- What can we say about the shape of the sampling distribution of $\bar{x}$ when the population from which the sample is selected is not normal?



Population



The Central Limit Theorem (for the sample mean $\bar{x}$)

- If a random sample of n observations is selected from a population (**any** population), then when n is sufficiently large, the sampling distribution of $\bar{x}$ will be approximately normal.

$30 \geq 0$

$n = 30$

- (The larger the sample size, the better will be the normal approximation to the sampling distribution of $\bar{x}$.)

$n \geq 30$

$E(\bar{X}) = \mu$

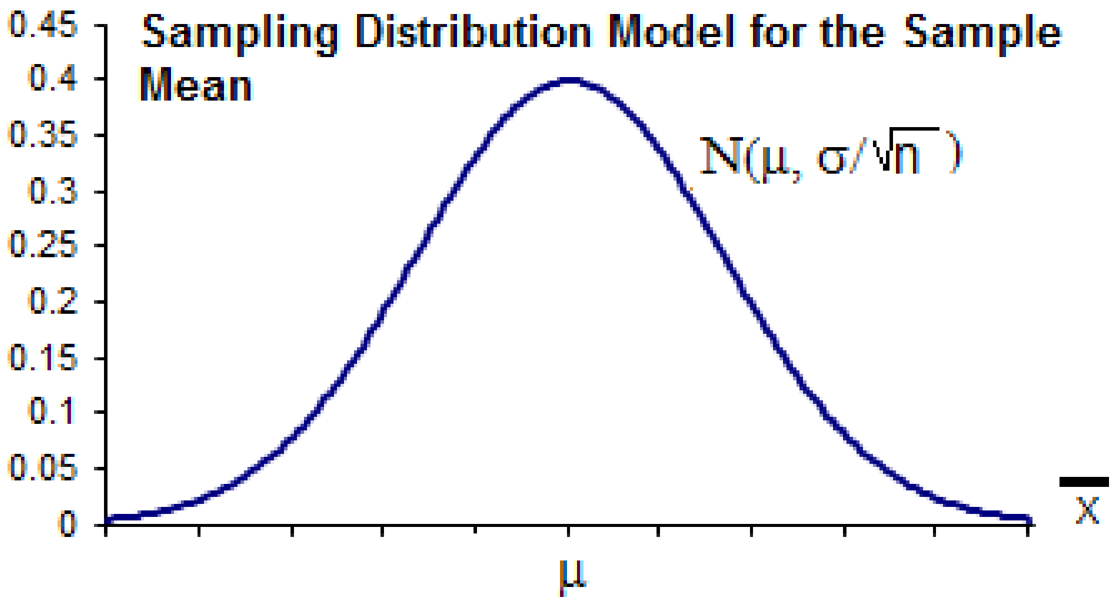
$\bar{X}$

$SD(\bar{X}) = \sqrt{V(\bar{X})} = \frac{\sigma}{\sqrt{n}}$

$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

The Importance of the Central Limit Theorem

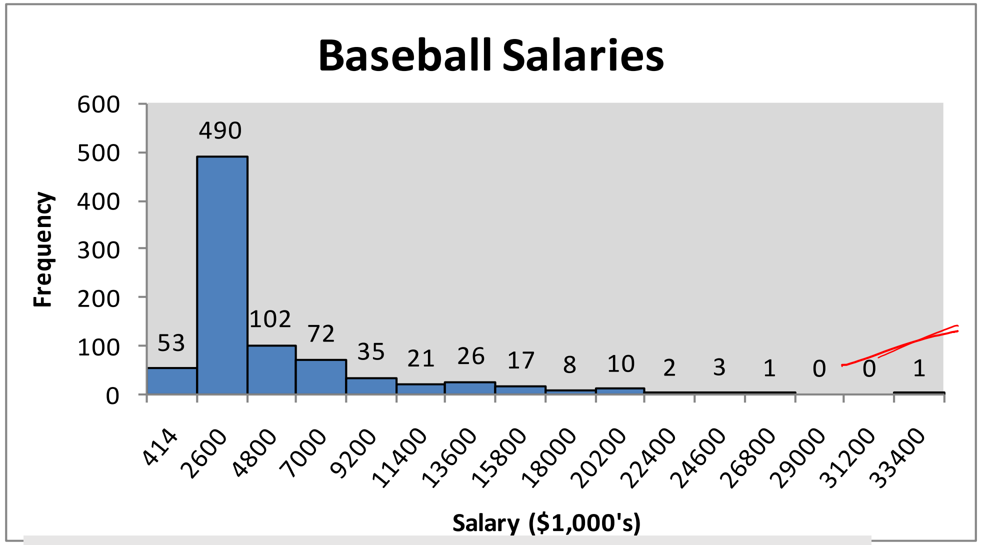
- When we select simple random samples of size n, the sample means $\bar{x}$ will vary from sample to sample.
- We can model the distribution of these sample means with a probability model that is …



$$\boxed{N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

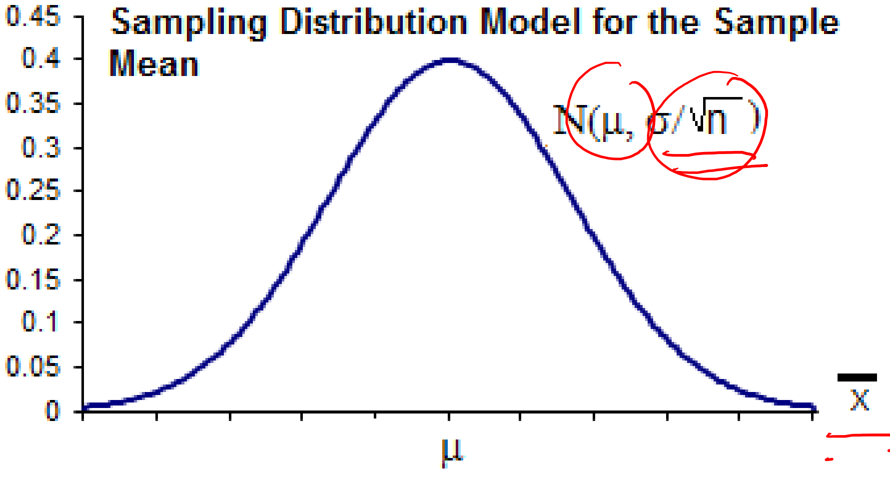
How Large Should n Be?

- For the purpose of applying the Central Limit Theorem, we will consider a sample size to be large when n > 30.



Even if the population from which the sample is selected looks like this …

… the Central Limit Theorem tells us that a good model for the sampling distribution of the sample mean $\overline{x}$ is …



The Central Limit Theorem (for the sample proportion $\hat{p}$)

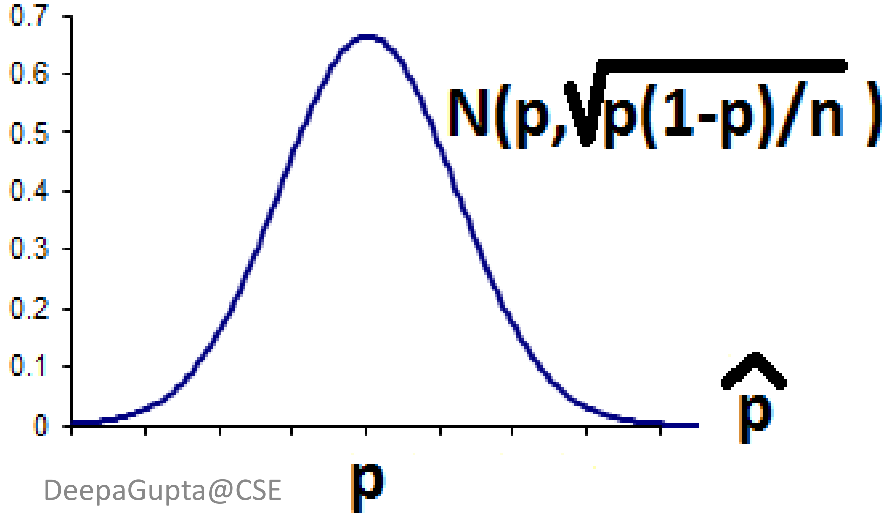
- If x "successes" occur in a random sample of n observations selected from a population (**any** population), then when n is sufficiently large, <u>the sampling distribution of $\hat{p}$ =x/n will be approximately normal.</u>

(The larger the sample size, the better will be the normal approximation to the sampling distribution of $\hat{p}$.)

The Importance of the Central Limit Theorem

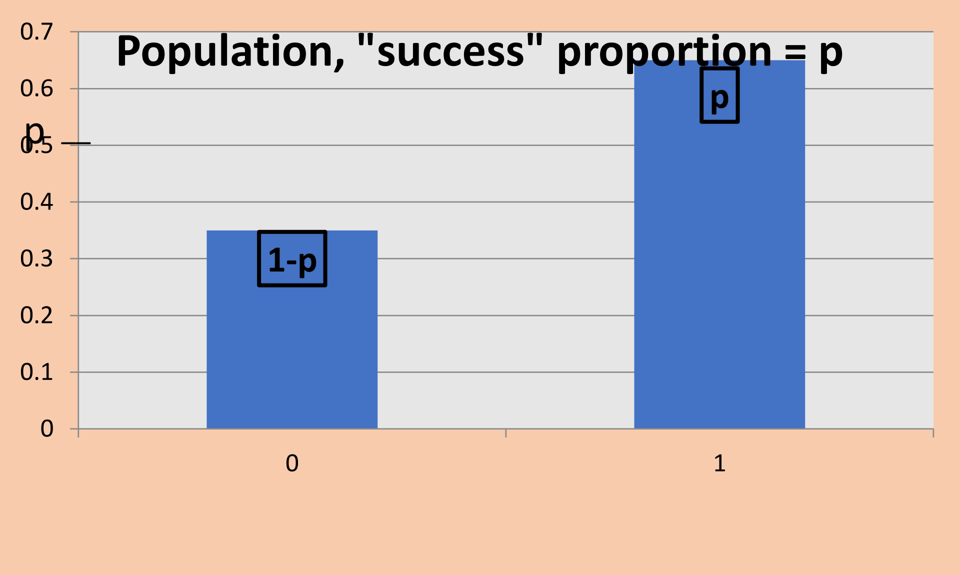
- When we select simple random samples of size n from a population with "success" probability p and observe x "successes", the sample proportions $\hat{p}$ =x/n will vary from sample to sample.
- We can model the distribution of these sample proportions with a probability model that is…

$$\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$



How Large Should n Be?

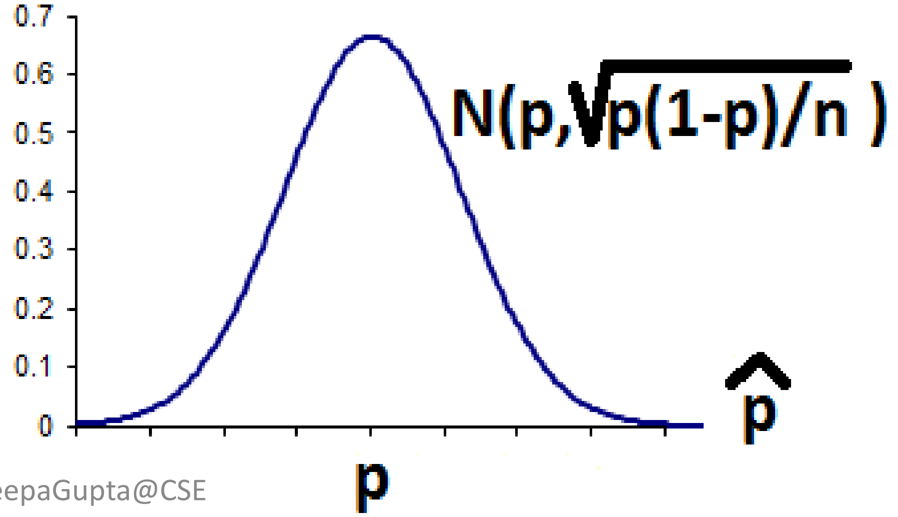
- For the purpose of applying the central limit theorem, we will consider a sample size n to be large when $np \geq 10$ and $n(1-p) \geq 10$



← ← ← *If the population from which the sample is selected looks like this …*

… the Central Limit Theorem tells us that a good model for the sampling distribution of the sample proportion is … $\hat{p} = \frac{x}{n}$

→ → →



Summary: Population Parameters and Sample Statistics

Population parameter	*Value*	*Sample statistic used to estimate*
p *proportion of population with a certain characteristic*	Unknown	$\hat{p}$
μ *mean value of a population variable*	Unknown	$\bar{x}$

- The value of a population parameter is a **fixed** number, it is NOT random; its value is **not known.**
- The value of a sample statistic is calculated from sample data
- The value of a sample statistic will vary from sample to sample (sampling distributions)

AMRITA VISHWA VIDYAPEETHAM

Example 1

A sample of size n=16 is drawn from a normally distributed population with E(X)=20 and SD(X)=8.

$\mu = 20, \quad \sigma = 8$

$\bar{X} = \dfrac{X_1 + X_2 + \cdots X_{16}}{16} \sim N(\mu, \sigma/\sqrt{n})$

$\boxed{X \sim N(20,8)}; \bar{X} \sim N(20, \frac{8}{\sqrt{16}})$ $\quad \frac{8}{\sqrt{n}} = 2$

$a) P(\bar{X} \geq 24) = P(\frac{\bar{X} - 20}{2} \geq \frac{24 - 20}{2})$

$= P(z \geq 2) = 1 - .9772 = .0228$

$= 1 - P(Z \leq 2)$ $\quad 2.0, 0.0$

$b) P(16 \leq \bar{X} \leq 24)$

$= P\left(\frac{16 - 20}{2} \leq z \leq \frac{24 - 20}{2}\right)$

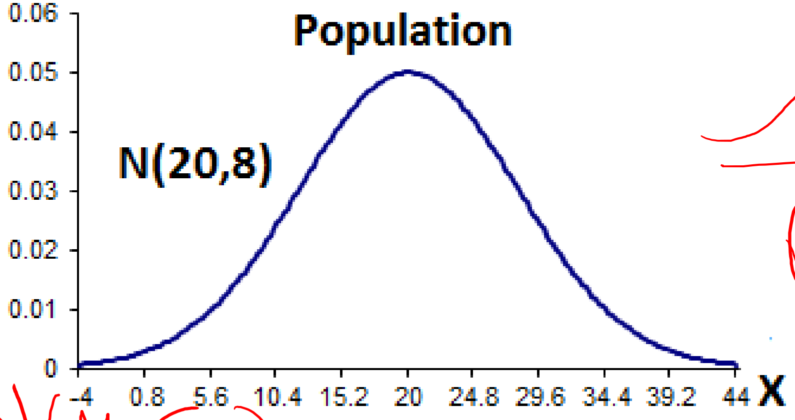
$= P(-2 \leq z \leq 2)$ $\quad = P(Z \leq 2) - P(Z \leq -2)$

$= .9772 - .0228 = .9544$

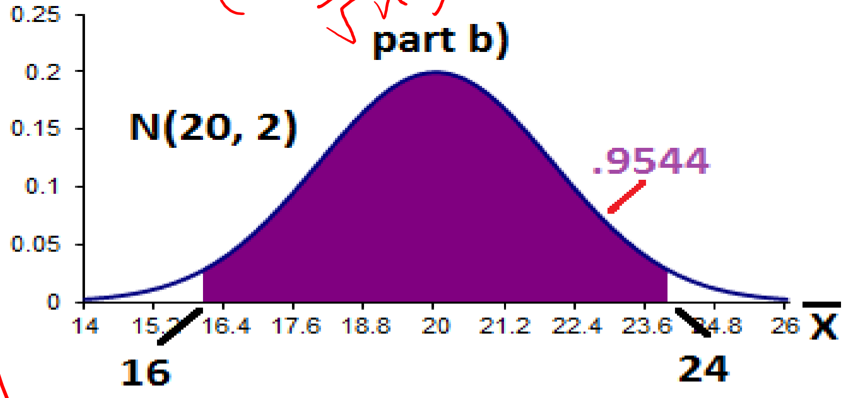
$X \sim N(\mu, \sigma)$

$Z \sim N(0,1)$

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

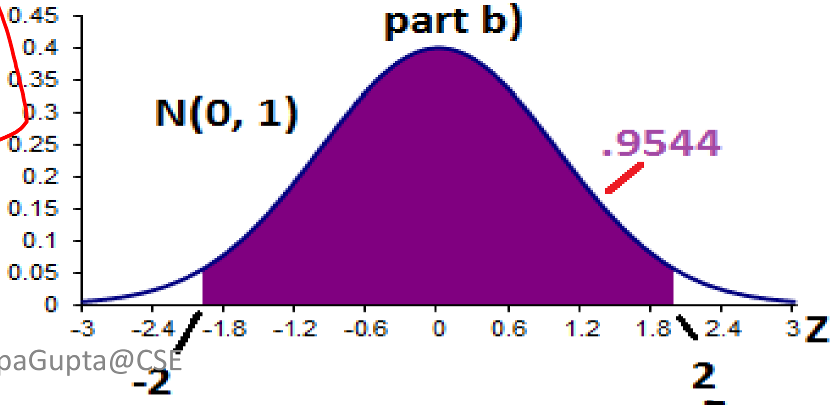


$P(Z \leq 2)$



$\bar{X} \sim N(\mu, \sigma/\sqrt{n})$

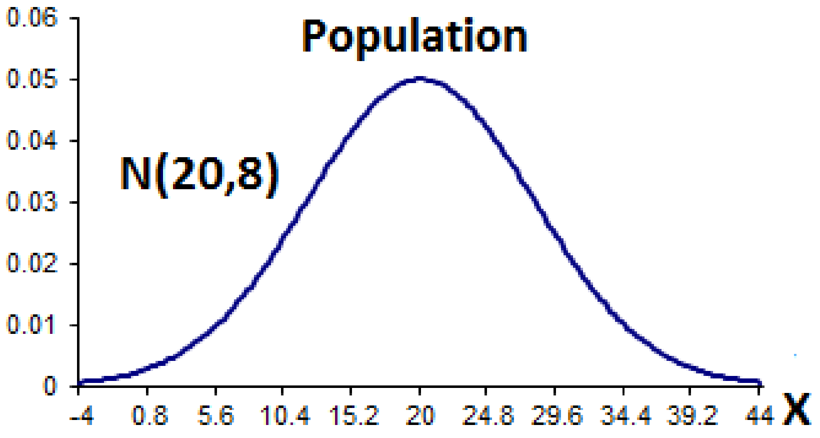
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$



Example 1 (cont.)

c. Do we need the Central Limit Theorem to solve part a or part b?

- NO. We are given that the population is normal, so the sampling distribution of the mean will also be normal for any sample size n. The CLT is not needed.



Example 2

- The probability distribution of 6-month incomes of account executives has mean \$20,000 and standard deviation \$5,000. n=64 account executives are randomly selected. What is the probability that the sample mean exceeds \$20,500?

μ Population

Population

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

64

$V(\bar{x}) = \frac{\sigma^2}{n}$

$S(\bar{x}) = \frac{\sigma}{\sqrt{n}}$

Given: E(X) = \$20,000
SD(X) = \$5,000

$$E(\bar{X}) = \$20{,}000$$

$$SD(\bar{X}) = \frac{SD(x)}{\sqrt{n}} = \frac{5{,}000}{\sqrt{64}} = 625$$

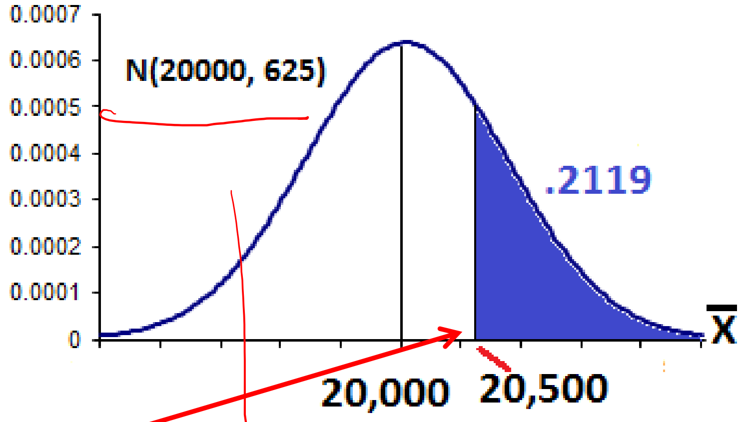
By CLT, $\bar{X} \sim N(20{,}000, 625)$

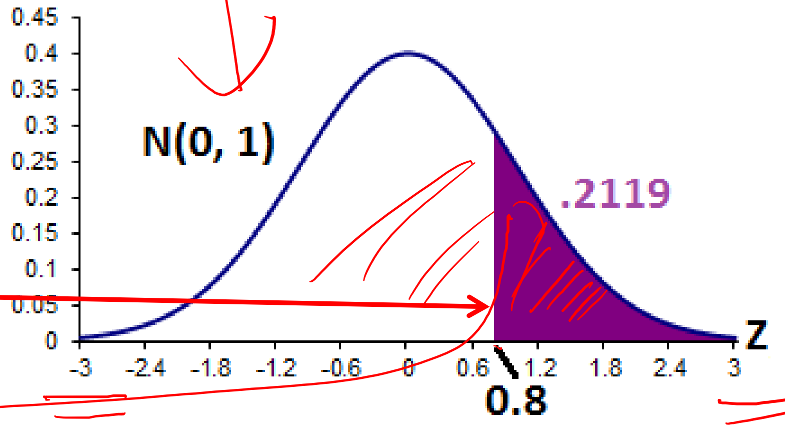
$$P(\bar{X} > 20{,}500) =$$

$$P\left(\frac{\bar{X} - 20{,}000}{625} > \frac{20{,}500 - 20{,}000}{625}\right) =$$

$1 - P(Z \le 0.8)$ → 0.80

$$P(z > .8) = 1 - .7881 = .2119$$





Example 3

- Battery life X~N(20, 10). Guarantee: average battery life in a case of 24 exceeds 16 hrs. Find the probability that a randomly selected case meets the guarantee.

Given $X \sim N(\overset{\mu}{20}, \overset{\sigma}{10})$

$\bar{X} \sim N\left(20, \frac{10}{\sqrt{24}}\right) = N(20, 2.04)$

$$E(\bar{X}) = 20; SD(\bar{X}) = \frac{10}{\sqrt{24}} = 2.04. \bar{X} \sim N(20, 2.04)$$

$$P(\bar{X} > 16) = P(\frac{\bar{X} - 20}{2.04} > \frac{16 - 20}{2.04}) = P(z > -1.96) = .1 - .0250 = .9750$$

Example 4

Population

- Cans of salmon are supposed to have a net weight of 6 oz. The canner says that the net weight is a random variable with mean μ=6.05 oz. and stand. dev. σ=.18 oz.

sample

- Suppose you take a random sample of 36 cans and calculate the sample mean weight to be 5.97 oz. Find the probability that the mean weight of the sample is less than or equal to 5.97 oz.

Solution

$P(\bar{X}_{36} \le 5.97)$

$\bar{X}$ sampling dist: E($\bar{X}$)=6.05 SD($\bar{X}$)=.18/6=.03

By the CLT, X sampling dist is approx. normal

$\boxed{P(\bar{X} \le 5.97)}$ = P(z $\le$ [5.97-6.05]/.03)

=P(z $\le$ -.08/.03)=$\boxed{P(z \le -2.67)}$= .0038

-2.00 0.67

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

$\bar{X} \sim N(6.05, 0.03)$

$P\left(\frac{\bar{X} - 6.05}{0.03}\right)$

As $n = 36 > 30$

$\Rightarrow \bar{X} \sim N(\mu, \sigma/\sqrt{n})$

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

Example 5

- 12% of students at NCSU are left-handed. What is the probability that in a sample of 100 students, the sample proportion that are left-handed is less than 11%?

$$E(\hat{p}) = p = .12; SD(\hat{p}) = \sqrt{\frac{.12 * .88}{100}} = .032$$

$$np = 100 \times .12 = 12 \geq 10;$$

$$n(1-p) = 100 \times .88 = 88 \geq 10;$$

So …

By the CLT, $\hat{p} \sim N(.12, .032)$

$\sqrt{p(1-p)/n}$

sampling dist

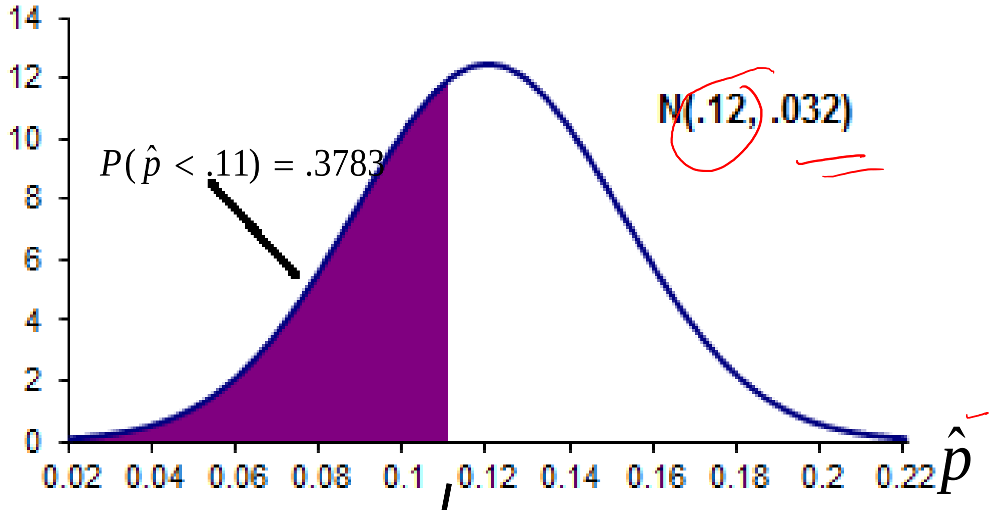
$E(\hat{p}) = p$

$\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$

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

$P(\hat{p} < 0.11)$

Example 5 (cont.)



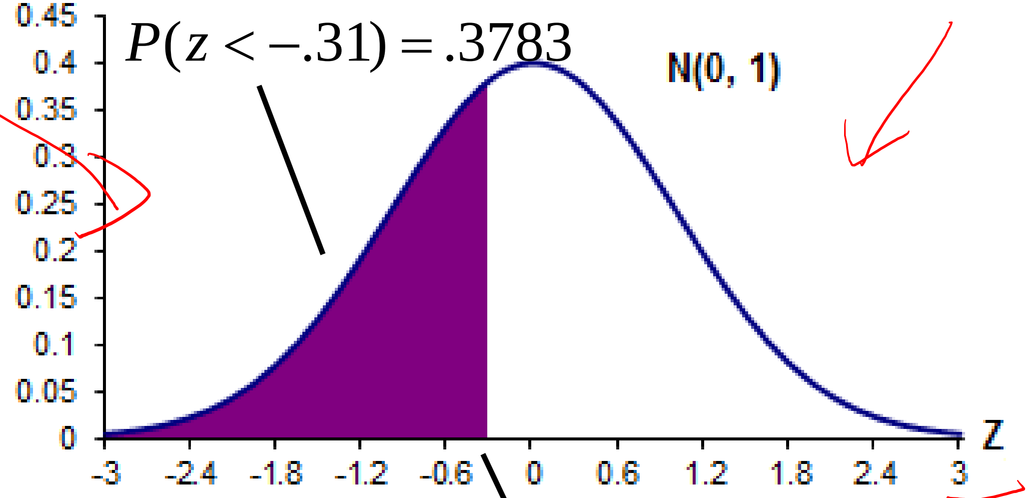
$\hat{p} = .11$

$$P(\hat{p} < .11) = P\left(\frac{\hat{p} - .12}{.032} < \frac{.11 - .12}{.032}\right)$$

$$= P(z < -.31) = .3783$$

-0, $-.31$

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

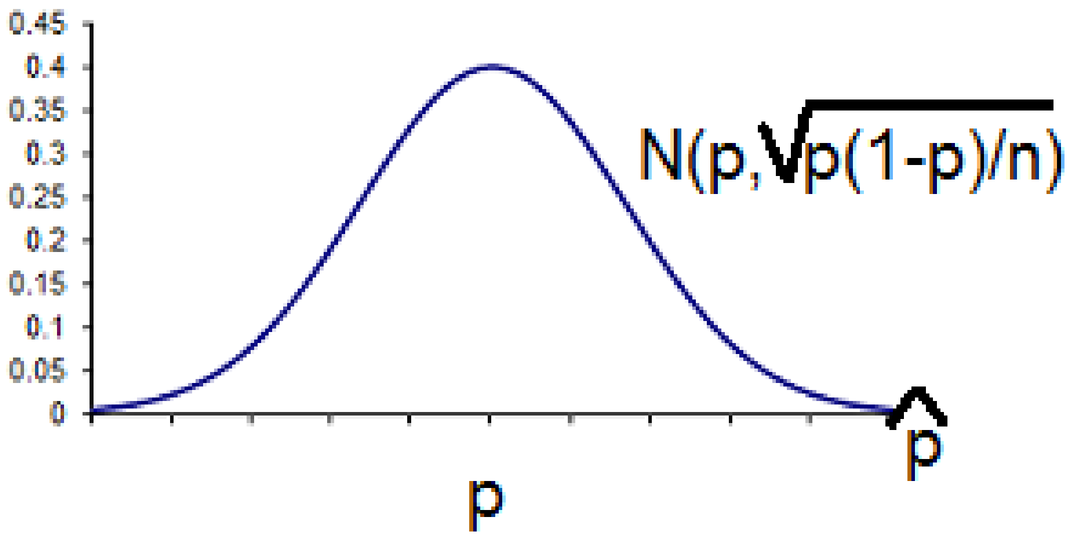


$z = -.31$

Shape of Sampling Distribution of $\hat{p}$

- The sampling distribution of $\hat{p}$ is approximately **normal** when the sample size n is large enough. n large enough means np ≥ 10 and n(1-p) ≥ 10

$\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$



Example 6 : binge drinking by college students

- Study by Harvard School of Public Health: 44% of college students binge drink.
- At a particular college 244 students were surveyed; 36% admitted to binge drinking in the past week
- Assume the value 0.44 given in the Harvard study is the proportion p of college students that binge drink; that is 0.44 is the population proportion p
- Compute the probability that in a sample of 244 students, 36% or less have engaged in binge drinking.

Example 6: binge drinking by college students (cont.)

- Let p be the proportion in a sample of 244 that engage in binge drinking.
- We want to compute

$$P(\hat{p} \leq .36)$$

- E($\hat{p}$) = p = .44; SD($\hat{p}$) = $\sqrt{\frac{pq}{n}} = \sqrt{\frac{.44 * .56}{244}} = .032$

- Since np = 244*.44 = 107.36 and nq = 244*.56 = 136.64 are both greater than 10, we can model the sampling distribution of p with a normal distribution, so …

Example 6: binge drinking by college students (cont.)

$$\hat{p} \sim N(.44, .032)$$

$$\text{So } P(\hat{p} \le .36) = P\left(\frac{\hat{p} - .44}{.032} \le \frac{.36 - .44}{.032}\right)$$

$$= P(z \le -2.5) = .0062$$

Example 7: snapchat by college students

- Recent scientifically valid survey : 77% of college students use snapchat. p

- 1136 college students surveyed; 75% reported that they use snapchat.

- Assume the value 0.77 given in the survey is the proportion p of college students that use snapchat; that is **0.77 is the population proportion p**

- Compute the probability that in a sample of 1136 students, 75% or less use snapchat.

Example 7: snapchat by college students (cont.)

- Let $\hat{p}$ be the proportion in a sample of 1136 that use snapchat.
- We want to compute $\boxed{P(\hat{p} \leq .75)}$

1 − p

- E($\hat{p}$) = p = .77; SD($\hat{p}$) = $\sqrt{\frac{pq}{n}} = \sqrt{\frac{.77*.23}{1136}} = .0125$
- Since $\boxed{np = 1136*.77 = 874.72}$ and $\boxed{nq = 1136*.23 = 261.28}$ are both greater than 10, we can model the sampling distribution of p with a normal distribution, so …

Then CLT $\hat{p} \sim N(0.77, 0.0125)$

Example 7: snapchat by college students (cont.)

$$\hat{p} \sim N(.77, .0125)$$

$$\text{So } P(\hat{p} \le .75) = P\left(\frac{\hat{p} - .75}{.0125} \le \frac{.75 - .77}{.0125}\right)$$

$$= P(z \le -1.6) = .0548$$

-1.00 , 0.60