CENSUS OF AGE AND ELECTION OF MARKETING RESEARCH COURSE AS OPTION OF STUDENTS

(1) Student number	(2) Age (X_1)	(3) Election of marketing research course 1 = yes, 0 = no (X_2)	(1) Student number	(2) Age (X_1)	(3) Election of marketing research course 1 = yes, 0 = no (X_2)
1	25	1	26	22	0
2	27	0	27	19	1
3	29	1	28	20	0
4	31	1	29	19	0
5	25	0	30	24	0
6	29	0	31	25	0
7	27	0	32	22	1
8	24	0	33	20	0
9	27	1	34	21	1
10	28	1	35	21	0
11	33	0	36	23	1
12	29	1	37	21	0
13	26	0	38	23	0
14	28	0	39	18	0
15	28	1	40	21	1
16	26	0	41	19	0
17	26	1	42	23	0
18	36	1	43	22	1
19	29	0	44	19	0
20	26	0	45	20	0
21	21	0	46	20	0
22	19	0	47	21	0
23	24	0	48	20	1
24	22	0	49	19	0
25	20	1	50	18	0
				;X_1=1184	;X_2=17

A POPULATION OF 50 EXISTS

POPULATION PARAMETERS

AGE: T = **23.7,** [^{2} **= 16.9**

Election of Mktg. Research: Ξ **= .34,** [^{2} **= .2244**

A RANDOM SAMPLE (n = 5) IS DRAWN

SAMPLE STATISTICS

AGE: 0 **= 22.6,** **S^2 = 8.3**

Election of Mktg. Research: **P = .4,** **S^2 = .3**

SAMPLE STATISTICS DESCRIBE THE POPULATION PARAMETERS

MAKING INFERENCES
about Population **PARAMETERS**
FROM SAMPLE STATISTICS

ASSUME A POPULATION OF
FIVE ELEMENTS
A, B, C, D, E

TEN SAMPLES OF n = 2 ARE POSSIBLE

1	AB
2	AC
3	AD
4	AE
5	BC
6	BD
7	BE
8	CD
9	CE
10	DE

FORMULA ⟶ C^{N}_{n}

POPULATION SIZE = 50
SAMPLE SIZE = 5

 # OF SAMPLES POSSIBLE IS

$$C^{50}_{5}$$

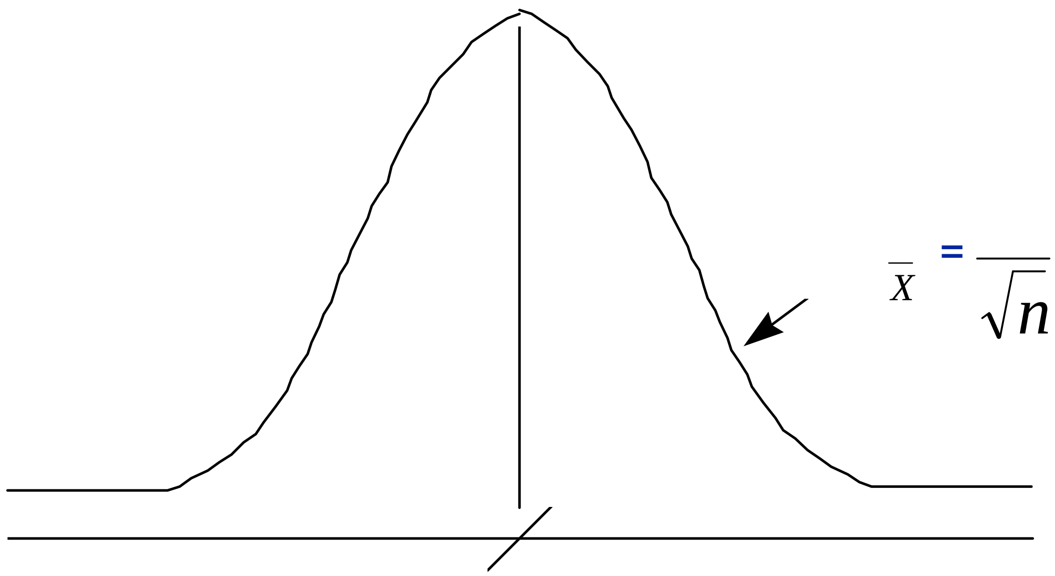
OR

$$\frac{\lfloor 50}{\lfloor 45 \quad \lfloor 5} \quad \textbf{SAMPLES}$$

$$= \frac{50 \times 49 \times 48 \times 47 \times 46}{5 \times 4 \times 3 \times 2 \times 1}$$

= 2.12 MILLION SAMPLES

THE SAMPLING DISTRIBUTION OF THE MEAN (X)



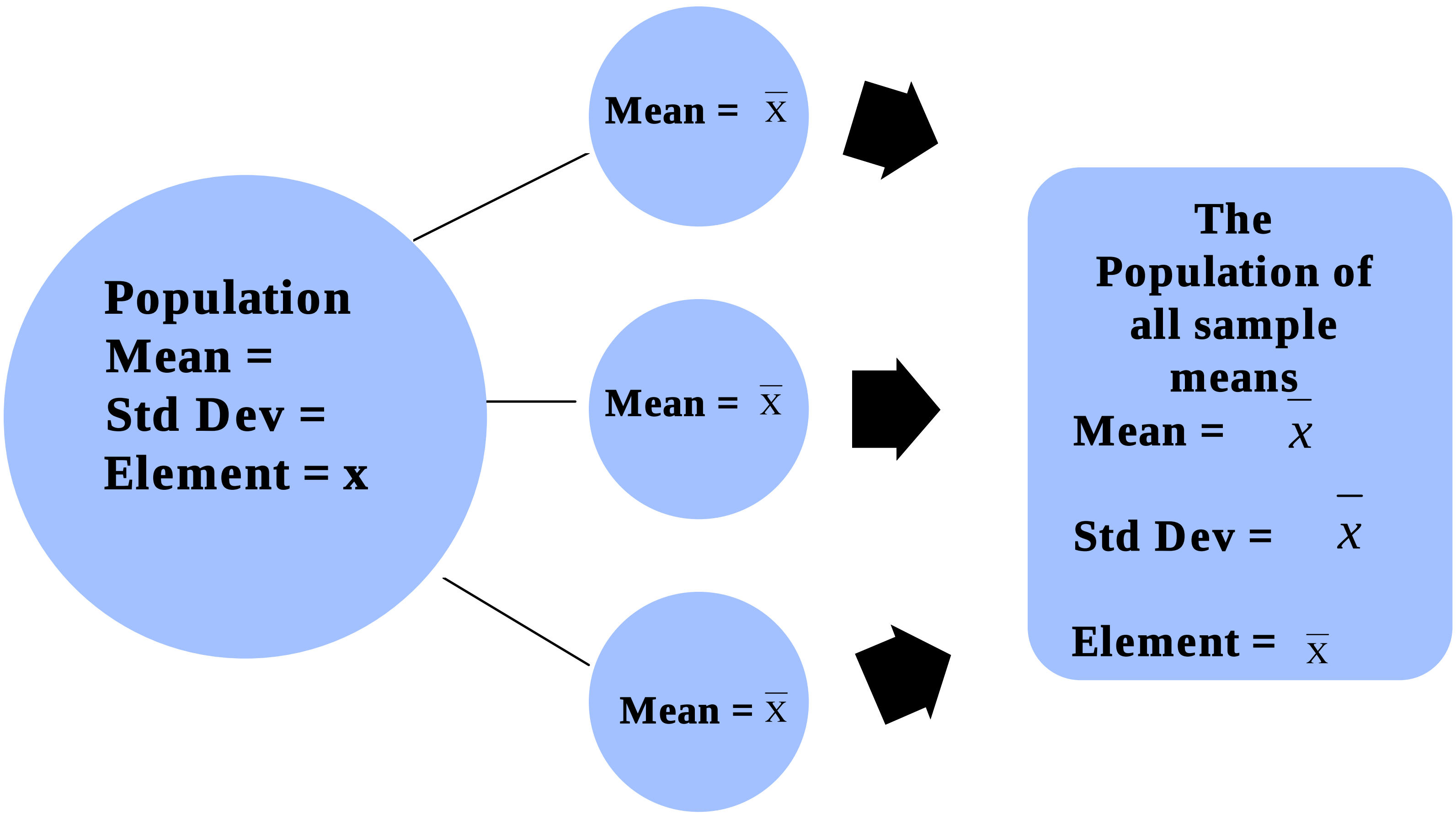
Three Important Properties

(1)
(2)
(3)

These are invoked to make inferences about the population parameters from the SS's

The Relationship Among the Three Distributions

Samples
Std Dev = s
Element = x

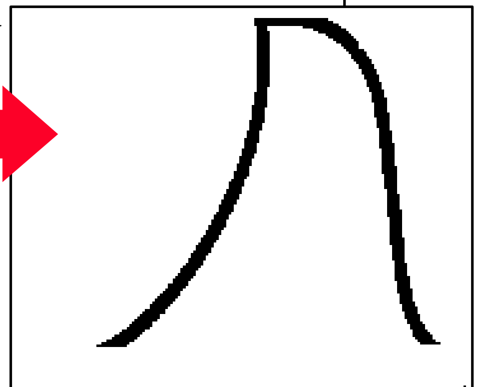


Comparing the Three Distributions

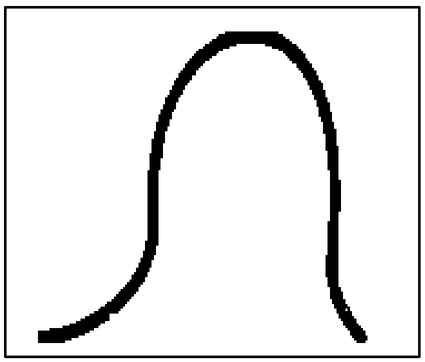
Distribution	Distribution of	Number of Distributions	Shape of the Distribution(s)
Population	all population values: x's	one	can be any shape
Sample	all values in a sample: x's	many	usually mirrors the population
Sampling	all sample means of a given sample size: $\bar{x}$'s	one for each different sample size	always normal provided sample size is large

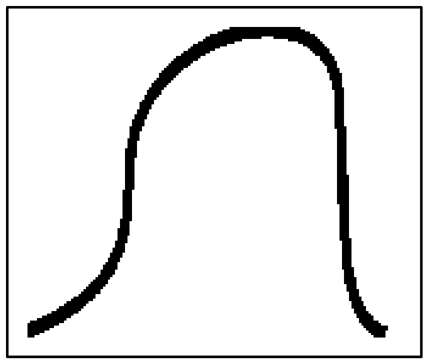
Shapes of the Three Distributions

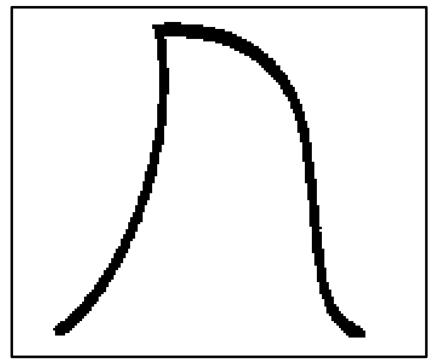
There is only one population distribution. It may have any shape. Even if it has a huge number of elements, it need not be normal.



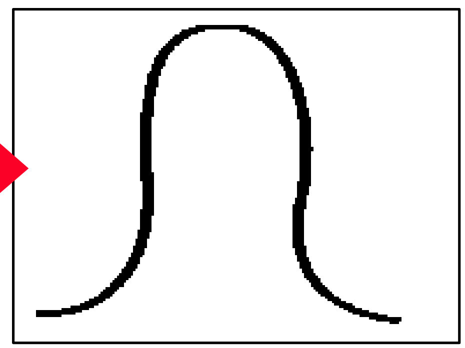
There are many sample distributions. Each will differ in shape, but they will tend to resemble the population distribution.

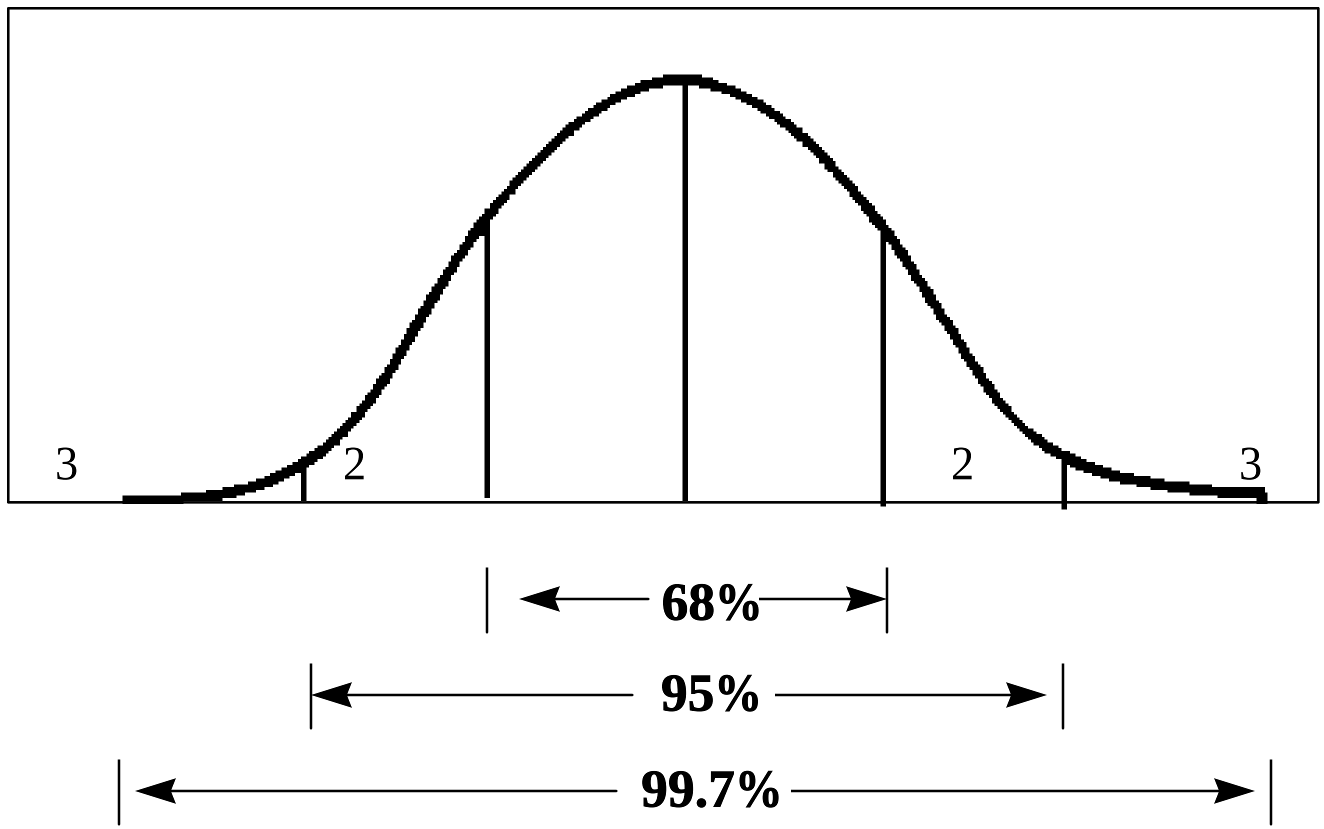






For a given "n" there is only one sampling distribution of the mean. If "n" is large, it will be normal, regardless of the shape of the population distribution.

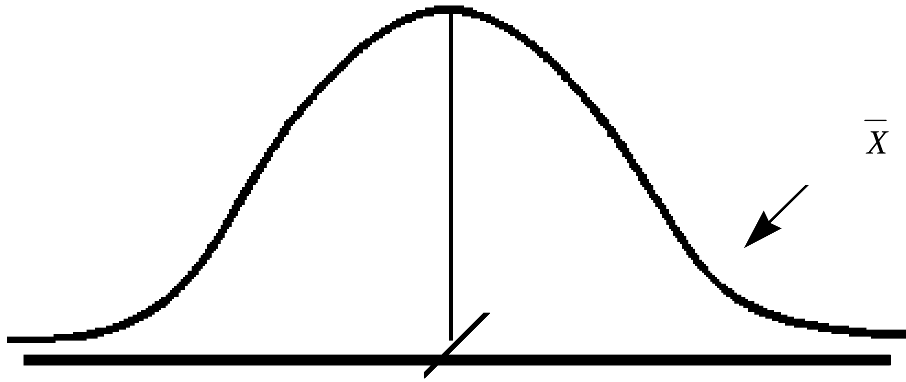




Area under the normal curve

The True Sampling Distribution

(Known <u>only</u> in this example)



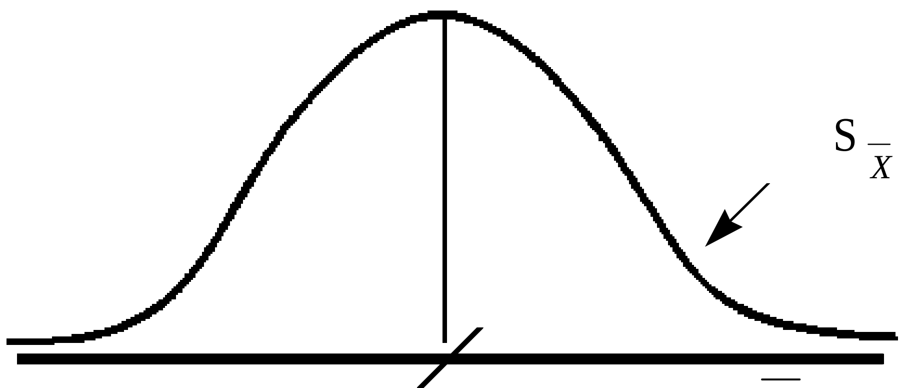
$_{\bar{X}} = \frac{\ }{\sqrt{n}}$

$= 23.7$

$_{\bar{x}} = 1.833$

The Estimated Sampling Distribution

(What we have got)



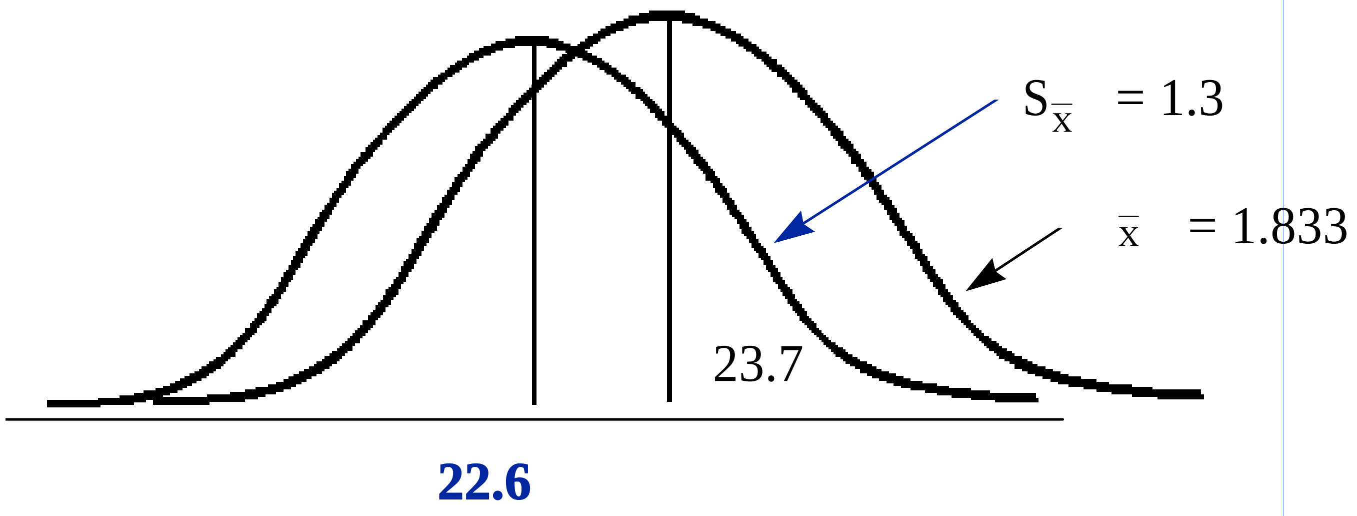
$S_{\bar{X}} \quad \frac{S}{\sqrt{n}}$

$\bar{x}$

$\bar{X} = 22.6$

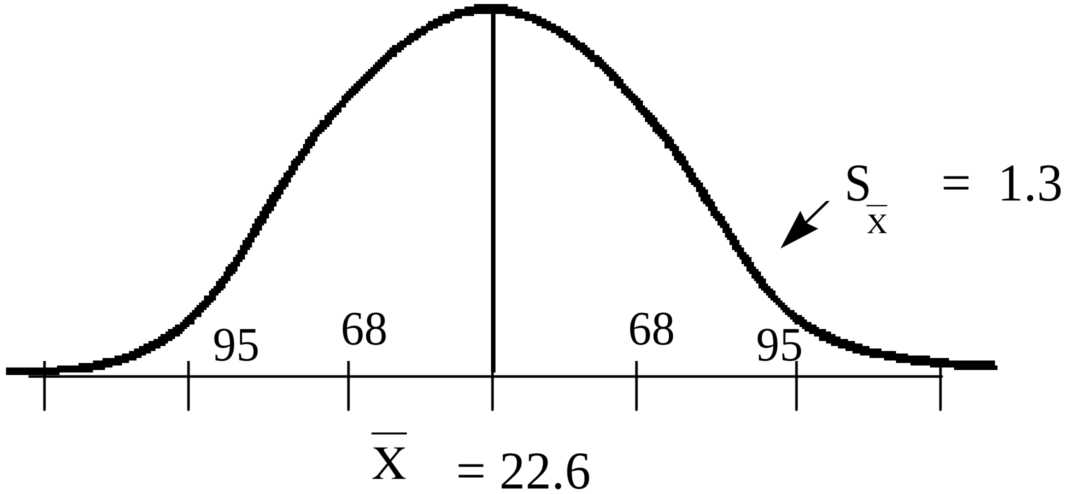
$S_{\bar{X}} = 1.3$

Making Believe That the True Sampling Distribution and the Estimated Sampling Distribution Are Close!



$$C.I._{95} = \bar{X} \pm 2 \frac{S}{\sqrt{}}$$

BEWARE OF FORMULA PLUGGING!



95 percent C.I. = 22.6 $\pm$ (2) (1.3)

68 percent C.I. = 22.6 $\pm$ (1) (1.3)

99.7 percent C.I. = 22.6 $\pm$ (3) (1.3)

The confidence interval is the measure of sampling error in probability sampling

SAMPLE SIZE DETERMINATION
(Confidence Interval Approach)

Figure 14.2 95% Confidence Interval

Random Sample of Households (n=300) to determine monthly expenses on dept. store shopping

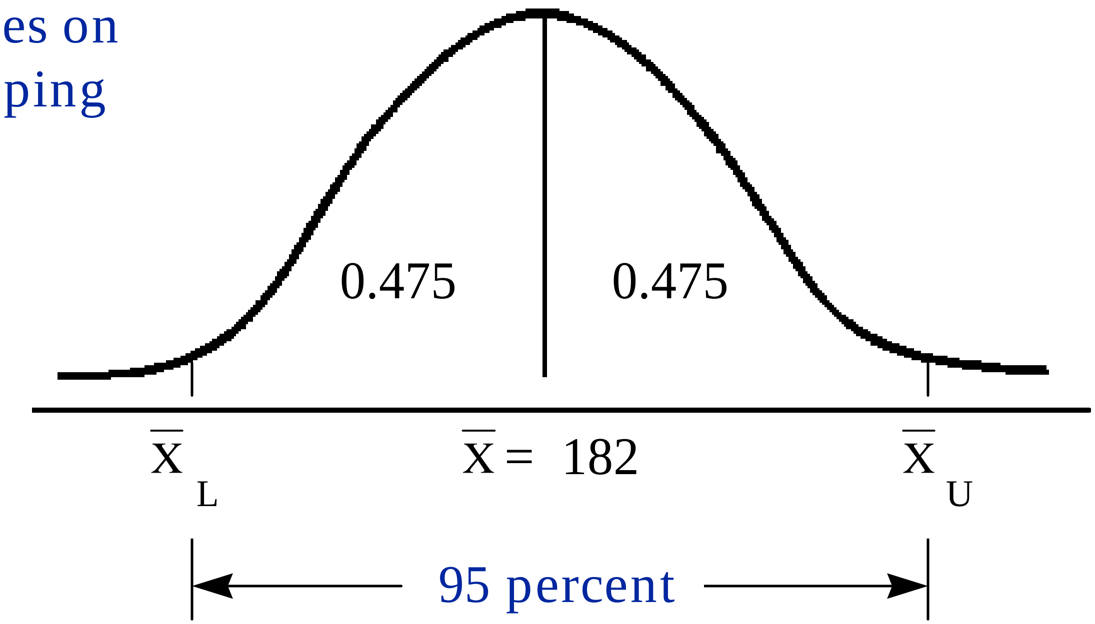
$\bar{X}$ = 182

= 55

S = 55

$$C.I._{95} = \bar{X} \pm 1.96 \frac{(\quad \text{or } S)}{\sqrt{\quad}}$$

$$= 182 \pm 6.23$$



175.77 188.23

$\bar{X}_L - \bar{X}$ = PRECISION (D) = 6.23

$\bar{X}_U - \bar{X}$ = PRECISION (D) = 6.23

POPULATION MEAN = μ = 23.7
(KNOWN ONLY IN THIS EXAMPLE)
SAMPLE MEAN = $\bar{X}$ = 22.6

95 percent C.I. = 22.6 $\pm$ (2) (1.3)

= 20 ⟷ 25.2

CORRECT INTERPRETATION: If the sampling process was repeated hundreds (or millions) of times, the population mean (μ) would lie in the above confidence interval in 95 percent of those cases.

WRONG INTERPRETATION: There is a 95 percent chance or probability that the population mean (μ) lies in the above confidence interval.

WHY WRONG INTERPRETATION IS WRONG!

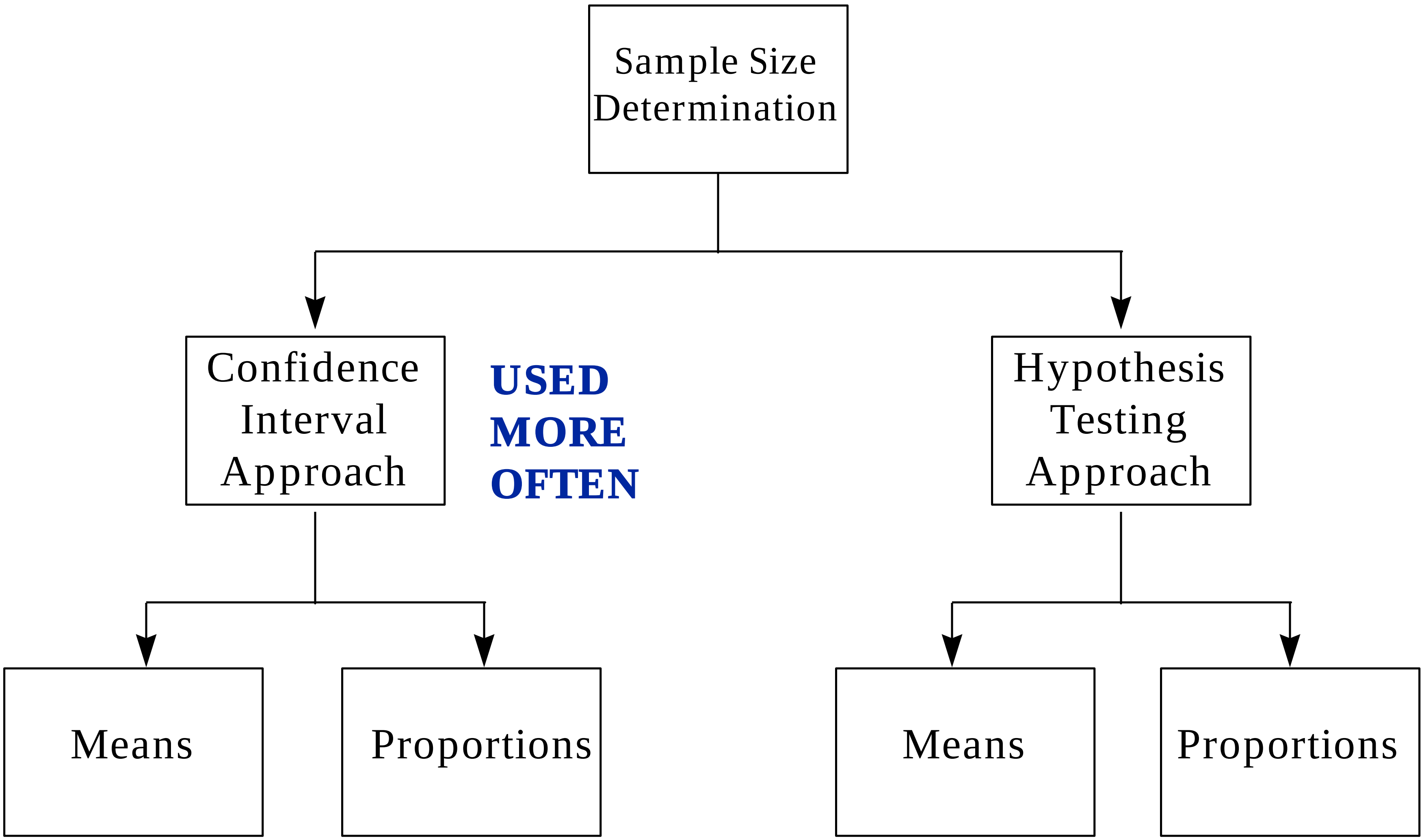
Common Misinterpretations of What “Statistically Significant” Means

- Viewing p values as if they represent the probability that the results occurred because of sampling error; e.g., $p = .05$ implies there is only a .05 probability that the resuilts were caused by chance.

- Equating statistical significance with practical significance.

- Viewing the or p levels as in some way related to the probability that the research hypothesis captured in the alternative hypothesis is true; e.g., a p value such as $p < .001$ is “highly significant” and therefore more valid than $p < .05$.

Figure 14.1 Statistical Approaches to Determining Sample Size



NOW LET US WORK THE CALCULATION BACKWARDS!

We want to estimate monthly household expense within $\pm$ 5 of the true population value. What should the sample size be?

$$\bar{X}_L - X = \text{PRECISION (D)} = 5$$

$$\bar{X}_U - \bar{X} = \text{PRECISION (D)} = 5$$

$$= 55$$

$$S = 55$$

D

$$C.I._{95} = \bar{X} \pm \frac{1.96\ (\quad \text{or } S)}{\sqrt{\quad}}$$

$$5 = \frac{1.96 \times 55}{\sqrt{\quad}}$$

$$n = \frac{(55)^2\ (1.96)^2}{5^2} = 465$$

1. (ABSOLUTE) PRECISION $= \pm 2\sqrt{\frac{pq}{n}}$

$$\pm .03 = \pm 2\sqrt{\frac{(.51)(.49)}{n}}$$

$$n = 1111 \text{ approx.}$$

NEW COKE vs. OLD COKE TASTE TEST

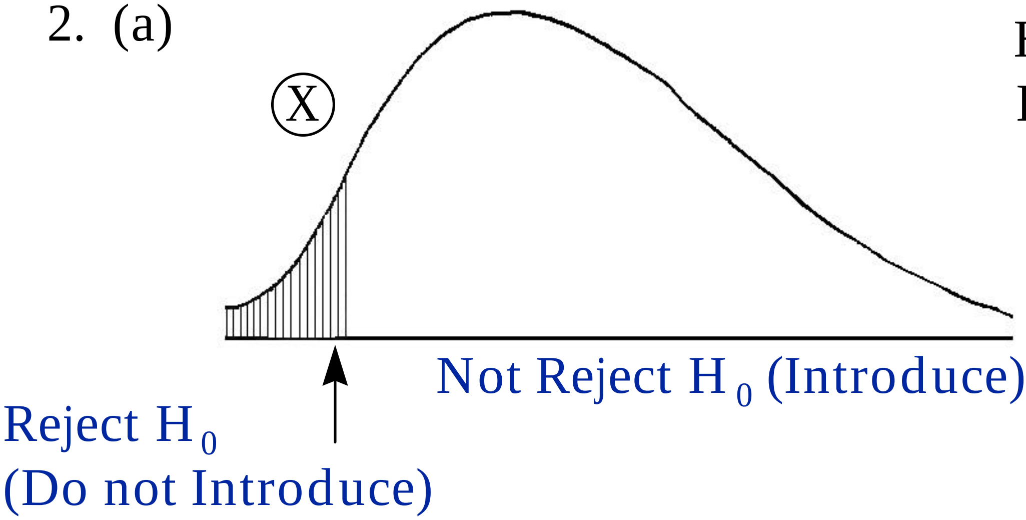
n = 200,000

P = .6 q = .4

PRECISION = ?

(90 percent confidence interval)

2. (a)

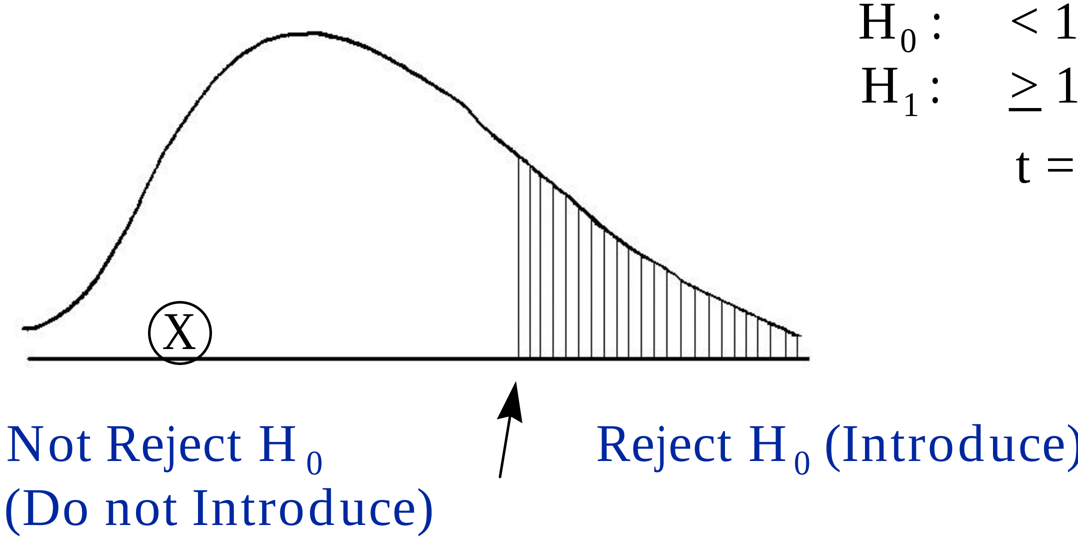


$H_0: \geq 100$

$H_1: < 100$

$t = -1.76$

(b)



$H_0: < 100$

$H_1: \geq 100$

$t = -1.76$

Table 14.1 Sample Size for Estimating Multiple Parameters

	VARIABLE		
	MEAN HOUSEHOLD MONTHLY EXPENSE ON:		
	Department Store Shopping	Clothes	Gifts
Confidence level	95%	95%	95%
z value	1.96	1.96	1.96
Precision level (*D*)	$5	$5	$4
Standard deviation of the population ( )	$55	$40	$30
Required sample size (*n*)	465	246	217

USE

DO NOT USE

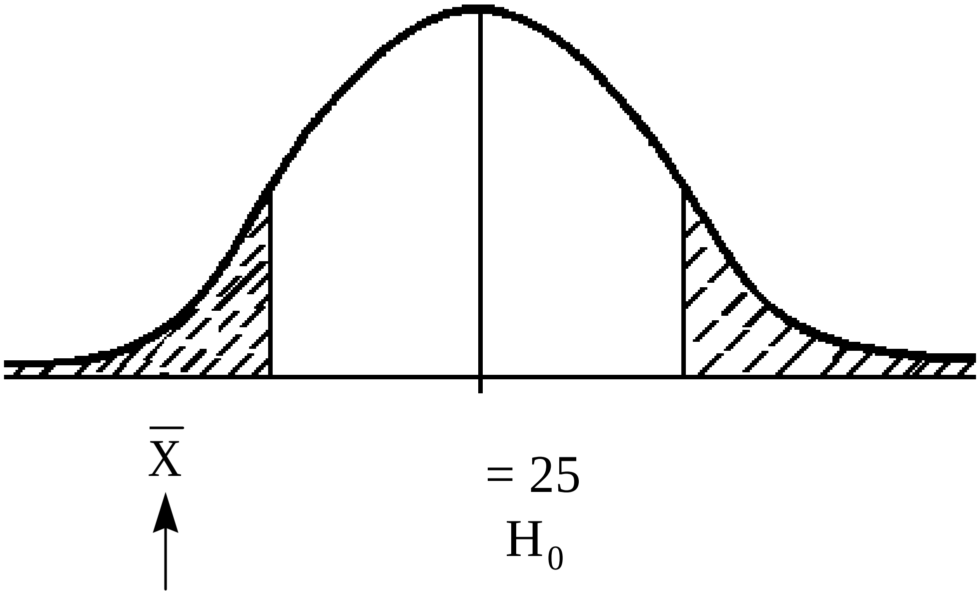
SUMMARY OF HYPOTHESIS TESTING ERRORS

	True condition	
Sample conclusions	H_0 is true	H_0 is false
Do not reject H_0	(1) Correct decision (2) Confidence level (3) Probability = 1 -	(1) Type II error (2) Probability =
Reject H_0	(1) Type I error (2) Significance level (3) Probability =	(1) Correct decision (2) Power of the test (3) Probability = 1 -

TRADEOFF

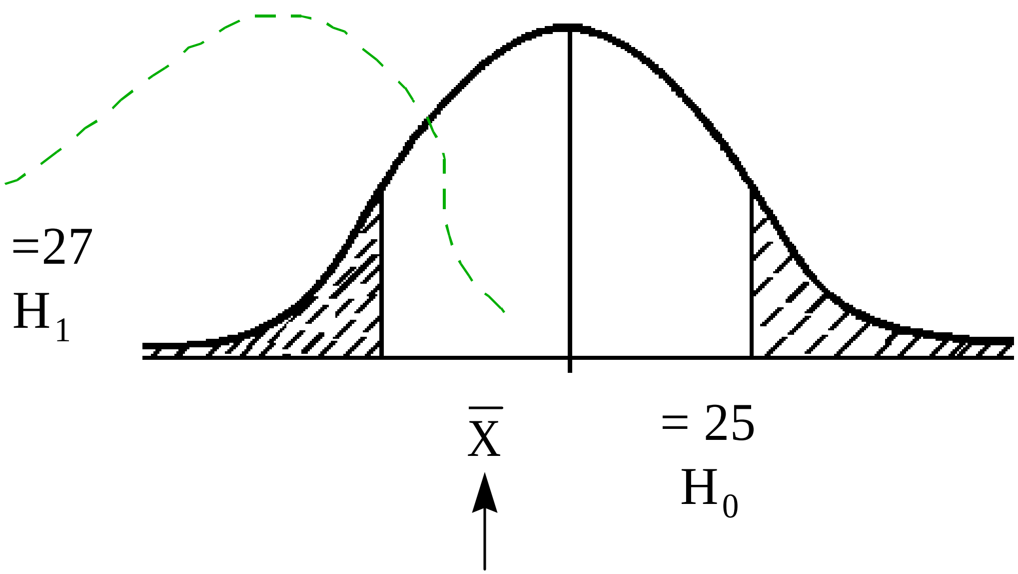
TYPE I ERROR is PRESPECIFIED
i.e., the chances of rejecting a null hypothesis when it is in fact true

Illustrating Type I & Type II Errors



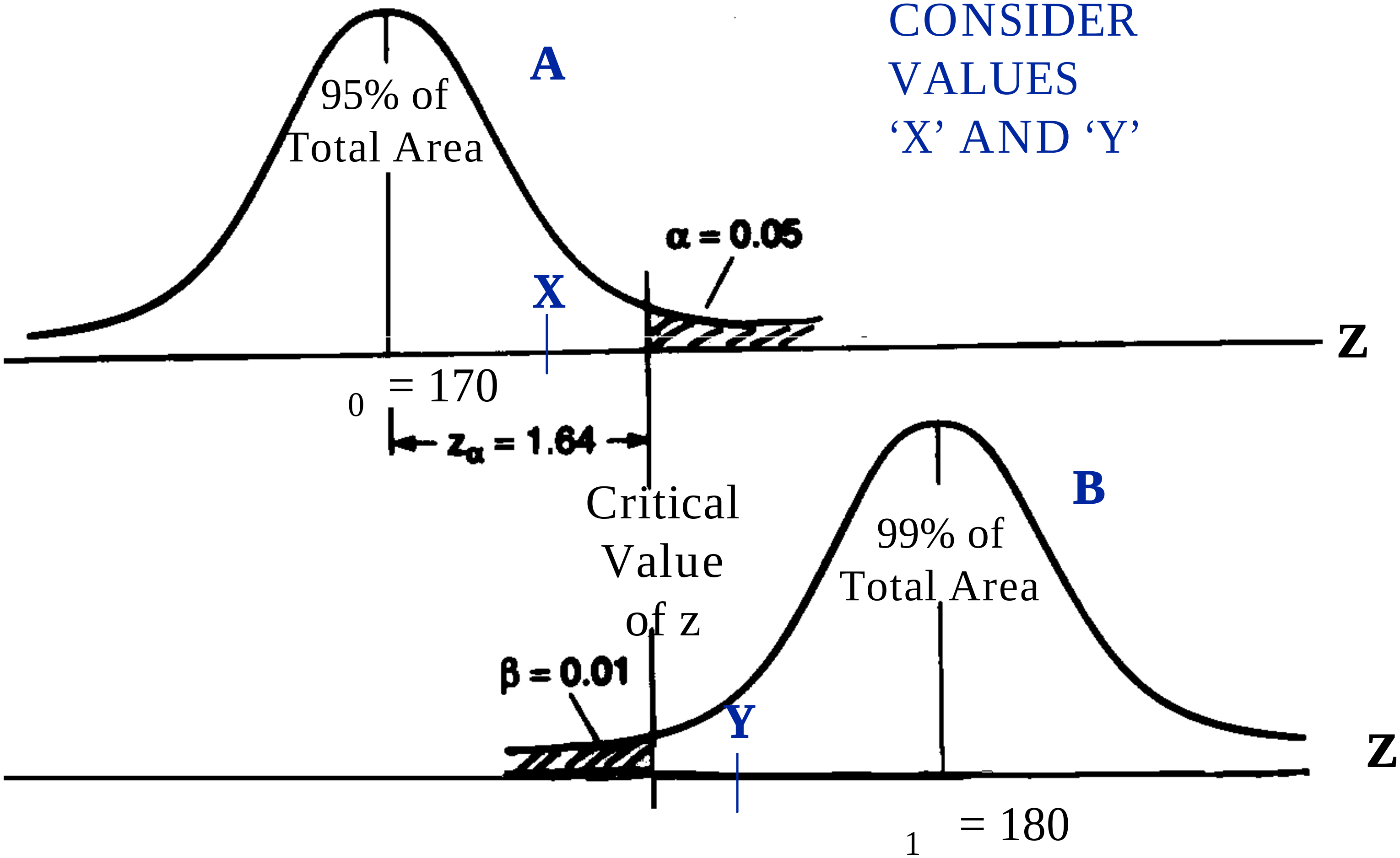
If X (or any sample statistic being tested) falls in the shaded areas, we will just pretend that it belongs to some other sampling distribution

(Hence the size of the shaded areas relative to the size of the sampling distribution specifies the Type I error)



We think $\bar{X}$ lies in the unshaded region. (In reality it belongs to some other “ghost” distribution and a Type II error has been made)

Figure 14.3 Type I error ( ) and Type II Error ( )



CONSIDER VALUES 'X' AND 'Y'

IF X BELONGS TO A ⟶ NO PROBLEM
IF X BELONGS TO B ⟶ TYPE II ERROR
IF Y BELONGS TO A ⟶ TYPE I ERROR
IF Y BELONGS TO B ⟶ NO PROBLEM

UNDERSTANDING TYPE I AND II ERRORS IN MARKETING RESEARCH

TYPE 1	→	Rejecting something that is true
TYPE 2	→	"Accepting" something that is not true

SCENARIO:

T **H_0: A = B**

H_1: A IS BETTER THAN B

New product concept "A" is just as good as new product concept "B" ... but we make a Type I error and assume concept "A" is better than "B" (reject H_0 when H_0 is true)

H_0: A = B

T **H_1: A IS BETTER THAN B**

Concept "A" is better than "B" ... but we make a Type II error and assume that the two new product concepts are the same

---

SCENARIO:

Executing an innocent person is an example of Type ? error

Judicial Analogy Illustrating Decision Errors

	TRUE SITUATION: DEFENDANT IS	
Verdict	**Innocent**	**Guilty**
Innocent	Correct Decision Probability: 1-	Error Probability: TYPE 2
Guilty	Error Probability: TYPE 1	Correct Decision Probability: 1-

TRADE-OFF

YOU WANT = 0.0000 AND HENCE WILL ACCEPT HIGH LEVELS OF