AD8403

DATA ANALYTICS

UNIT I

INFERENTIAL STATISTICS I

Lesson Plan

Planned Hour	Description of Portion to be Covered	Relevant CO Nos	Highest Cognitive level**
1	Populations – Samples – Random Sampling	CO1	K1
1	Probability And Statistics	CO1	K1
1	Sampling Distribution – Creating A Sampling Distribution	CO1	K1
1	Mean Of All Sample Means	CO2	K1
1	Standard Error Of The Mean – Other Sampling Distributions	CO2	K1
1	Hypothesis Testing	CO2	K1
1	Z-test – Z-test Procedure - Statement Of The Problem	CO2	K1
1	Null Hypothesis – Alternate Hypotheses	CO2	K1
1	Decision Rule – Calculations – Decisions - Interpretations	CO2	K1

CO 1	To recognize the different types of data and gain insight into the data science process
CO 2	Understand the flow and learn the steps in a data science process
T1	David Cielen, Arno D. B. Meysman, and Mohamed Ali, “Introducing Data Science”, Manning Publications, 2016

- Descriptive statistics describes data (for example, a chart or graph) and **inferential statistics** allows you to make predictions ("inferences") from that data.

- With inferential statistics, you take data from samples and make generalizations about a population.

Statistics

Descriptive

1. Organizing and summarizing data using numbers & graphs

2. Data Summary:
Bar Graphs, Histograms, Pie Charts, etc.
Shape of graph & skewness

3. Measures of Central Tendency:
Mean, Median, & Mode

4. Measures of Variability:
Range, variance, & Standard deviation

Inferential

1. Using sample data to make an inference or draw a conclusion of the population.

2. Uses probability to determine how confident we can be that the conclusions we make are correct.
(Confidence Intervals & Margins of Error)

Populations, Samples and Probability

POPULATIONS

- *Any complete set of observations (or potential observations) may be characterized as a* **population**

- **Real Populations**
  - A *real* population is one in which all potential observations are accessible at the time of sampling.

- **Hypothetical Populations**
  - A *hypothetical* population is one in which all potential observations are not accessible at the time of sampling.
  - **The population in which whose unit is not available in solid form**

SAMPLES

- *Any subset of observations from a population may be characterized as a* **sample**

- **Optimal Sample Size**
  - There is no simple rule of thumb for determining the best or optimal sample size for any particular situation.
  - Often sample sizes are in the hundreds or even the thousands for surveys, but they are less than 100 for most experiments

- For each of the following pairs, indicate with a Yes or No whether the relationship between the first and second expressions could describe that between a sample and its population, respectively.

(a) students in the last row; students in class

(b) citizens of Wyoming; citizens of New York

(c) 20 lab rats in an experiment; all lab rats, similar to those used, that could undergo the same experiment

(d) all U.S. presidents; all registered Republicans

(e) two tosses of a coin; all possible tosses of a coin

- Identify all of the expressions from the above which involve a hypothetical population.

RANDOM SAMPLING

- *Random sampling* occurs if, at each stage of sampling, the selection process guarantees that all potential observations in the population have an equal chance of being included in the sample.

- A casual or haphazard sample doesn't qualify as a random sample.

- Indicate whether each of the following statements is True or False. A random selection of 10 playing cards from a deck of 52 cards implies that

(a) the random sample of 10 cards accurately represents the important features of the whole deck.

(b) each card in the deck has an equal chance of being selected.

(c) it is impossible to get 10 cards from the same suit (for example, 10 hearts).

(d) any outcome, however unlikely, is possible.

TABLES OF RANDOM NUMBERS

- Used to obtain a random sample
- The size of the population determines whether you deal with numbers having one, two, three, or more digits.
- For example, if you were attempting to take a random sample from a population consisting of 679 students at some college, you could use the 1000 three-digit numbers ranging from 000 to 999

- Use the upper-left-hand corner of the specimen page (table H, appendix C)

- Read in a consistent direction—for instance, from left to right.

- Then as each row is used up, shift down to the start of the next row and repeat the entire process.

- As a given number between 001 and 720 is encountered, the person identified with that number is included in the random sample

- Let’s assume that we have a population of 720 students and each student has been assigned a number from 1 to 720. Suppose we wish to sample 10 students
- Since we have a population of 720 and 720 is a three digit number, we need to use the first three digits of the numbers listed on the chart.
- Since the first number on the specimen page in Table H is 100 (disregard the fourth and fifth digits in each five-digit number),.
- The person identified with that number is included in the sample. The next three-digit number, 325, identifies the second person. Ignore the next number, 765, since none of the numbers between 721 and 999 is identified with any names in the student directory.
- Also, ignore repeat appearances of any number between 001 and 720. The next three-digit number, 135, identifies the third person. Continue this process until the specified sample size has been achieved

- For example, a six-digit random number, such as 239421,
- Identifies the name on page 239 (the first three digits) and line 421 (the last three digits).
- This process is repeated for a series of six-digit random numbers until the required number of names has been sampled.

- Describe how you would use the table of random numbers to take

(a) a random sample of five statistics students in a classroom where each of nine rows consists of nine seats.

(b) A random sample of size 40 from a large directory consisting of 3041 pages, with 480 lines per page

- Many pollsters use random digit dialing in an effort to give each telephone number—whether landline or wireless—
  - In the United States an equal chance of being called for an interview.

- Essentially, the first six digits of a 10-digit phone number, including the area code, are randomly selected from tens of thousands of telephone exchanges, while the final four digits are taken directly from random numbers.

RANDOM ASSIGNMENT OF SUBJECTS

- Experiments evaluate an independent variable by focusing on a **treatment group and a control group**
- Here the subjects in experiments can't be selected randomly from any real population, they can be **assigned randomly**, that is, with equal likelihood, to these two groups

- Random assignment refers to the use of chance procedures in psychology experiments to ensure that each participant has the same opportunity to be assigned to any given group.
- Study participants are randomly assigned to different groups, such as the **treatment group or control group.**

How to Assign Subjects

- Random assignment might involve tactics such as
  - flipping a coin,
  - drawing names out of a hat,
  - rolling dice, or
  - assigning random numbers to participants.

- To determine if changes in one variable lead to changes in another variable, psychologists must perform an experiment
- For instance, as each new subject arrives to participate in the experiment,
  - a flip of a coin can decide whether that subject should be assigned to the treatment group (if heads turns up) or the control group (if tails turn up).

- **Creating Equal Groups**
  - Equal numbers of subjects should be assigned to the treatment and control groups for a variety of reasons, including the increased likelihood of detecting any difference between the two groups.

- Assume that 12 subjects arrive, one at a time, to participate in an experiment. Use random numbers to assign these subjects in equal numbers to group A and group B. Specifically, random numbers should be used to identify the first subject as either A or B, the second subject as either A or B, and so forth, until all subjects have been identified. There should be six subjects identified with A and six with B.

(a) Formulate an acceptable rule for single-digit random numbers. Incorporate into this rule a procedure that will ensure equal numbers of subjects in the two groups. Check your answer in Appendix B before proceeding.

(b) Reading from left to right in the top row of the random number page (Table H, Appendix C), use the random digits of each random number in conjunction with your assignment rule to determine whether the first subject is A or B, and so forth. List the assignment for each subject.

PROBABILITY

- The *probability* that it will rain this weekend (20 percent, or one in five) or that you'll win a state lottery (one in many millions, unfortunately).

- ***Probability*** **refers to the proportion or fraction of times that a particular event is likely to occur.**

- Coin is tossed, and therefore, the probability of heads should equal .50, or 1/2.

$$P(A) = \frac{\text{Number of favorable outcomes to A}}{\text{Total number of possible outcomes}}$$

- Find the probability of getting a number less than 5 when a dice is rolled by using the probability formula

- What is the probability of getting a sum of 9 when two dice are thrown?

Table 8.1
PROBABILITY DISTRIBUTION FOR HEIGHTS OF AMERICAN MEN

HEIGHT (INCHES)	RELATIVE FREQUENCY
76 or taller	.02
75	.02
74	.03
73	.05
72	.08
71	.11
70	.12
69	.14
68	.12
67	.11
66	.08
65	.05
64	.03
63	.02
62 or shorter	.02
Total	1.00

ADDITION RULE

- What's the probability that a randomly selected man will be at least 73 inches tall?
- "What's the probability that a man will stand 73 inches tall *or* taller?"

- The probability that a man, X, will stand 73 or more inches tall, symbolized as Pr($X \geq 73$),

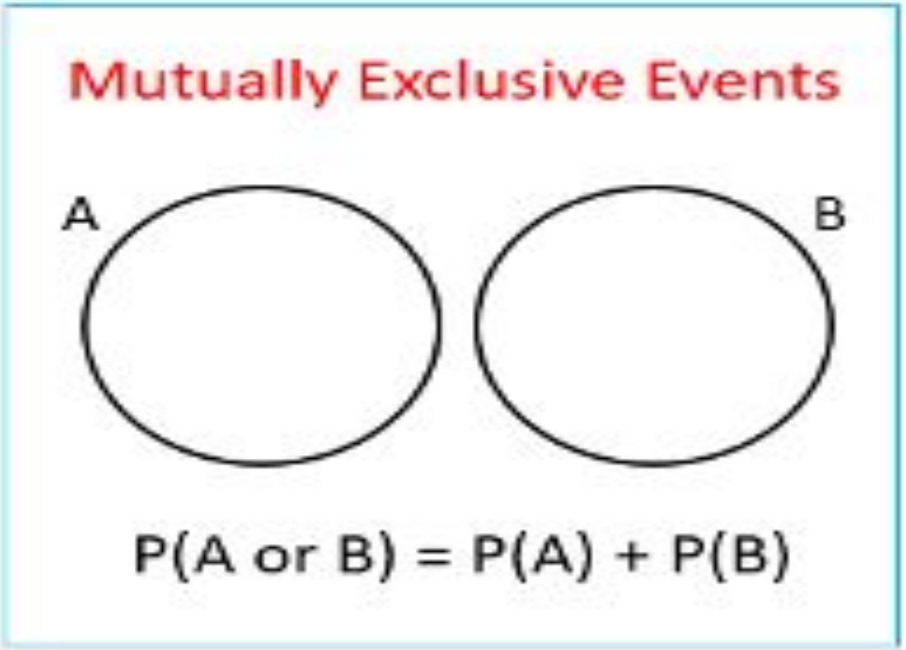
- equals the sum of the probabilities in Table that a man will stand 73 or 74 or 75 or 76 inches or taller, that is,

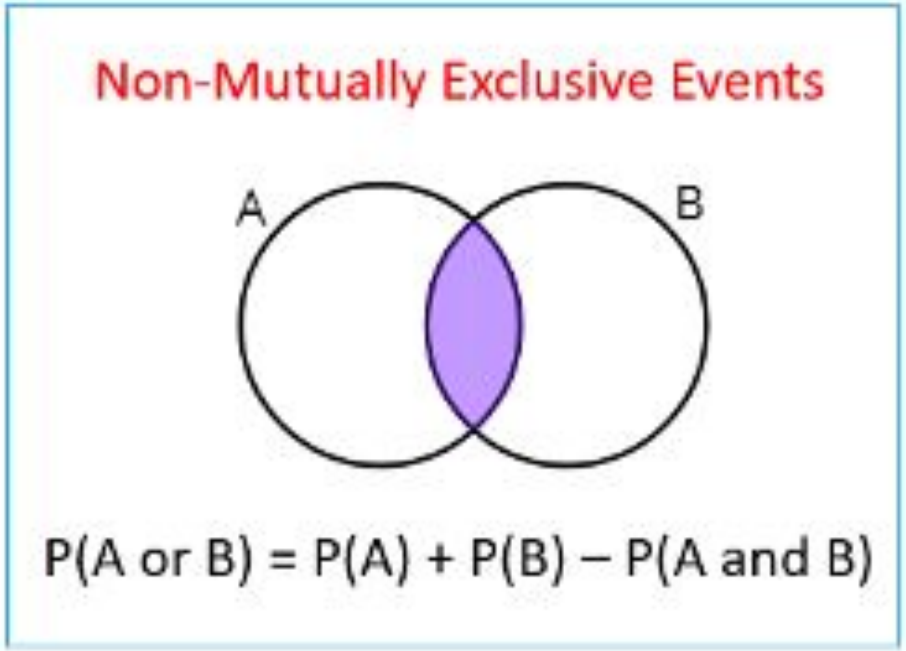
  Pr(x >= 73) = Pr(73)+Pr(74)+Pr(75)+Pr(76)

  = .05 +.03 +.02 +.02 = .12

Mutually Exclusive

- Two sets are known to be **mutually exclusive** when they have no common elements.





- Two sets are **non-mutually exclusive** if they share common elements.

- Whenever events can't occur together—that is, more technically, when there are mutually exclusive events—the probability that any one of these several events will occur is given by the addition rule

ADDITION RULE FOR MUTUALLY EXCLUSIVE EVENTS

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B)$$

- *Probabilities are added - in order to find the probability that any one of these events will occur*

- If the events A and B are not mutually exclusive, the probability is: **(A or B) = p(A) + p(B) – p(A and B)**.

- A box contains 2 apples, 4 guava, 5 berry and 3 mangoes. If a single random fruit is chosen from the box, what is the probability that it is an apples or guava marble?

- In a university of 300 students, 170 are boys and 130 are girls. On a workshop conducted by the university, 40 boys and 50 girls attended the workshop. If a student is chosen at random from the university, what is the probability of choosing a girl or an student who attended the workshop?

- Assuming that people are equally likely to be born during any one of the months, what is the probability of Jack being born during

(a) June?

(b) any month other than June?

(c) either May or June?

MULTIPLICATION RULE

- Whenever you must find the probability for two or more sets of independent events connected by the word *and,* use the multiplication rule

- What is the probability that two randomly selected men will be at least 73 inches tall?

- "What is the probability that the first man will stand at least 73 inches tall *and* that the second man will stand at least 73 inches tall?"

- The probability that both men will stand at least 73 inches tall equals the product of the probabilities in Table that the first man, X1, will stand at least 73 inches tall *and* that the second man, X2, will stand at least 73 inches tall, that is,

$\Pr(X1 >= 73 \; and \; X2 >= 73) = [\Pr(X1) >= 73][\Pr(X2\ ) >=73] = (.12)(.12)$
.0144

- Whenever one event has no effect on the other—that is, more technically, when there are **independent events**

- The **multiplication rule** *tells us to multiply together the separate probabilities of several independent events in order to find the probability that these events will occur together*

MULTIPLICATION RULE FOR INDEPENDENT EVENT

$$\Pr(A \text{ and } B) = [\Pr(A)][\Pr(B)]$$

- Assuming that people are equally likely to be born during any of the months, and also assuming (possibly over the objections of astrology fans) that the birthdays of married couples are independent, what's the probability of

(a) the husband being born during January and the wife being born during February?

(b) both husband and wife being born during December?

(c) both husband and wife being born during the spring (April or May)? (**Hint:** First, find the probability of just one person being born during April or May.)

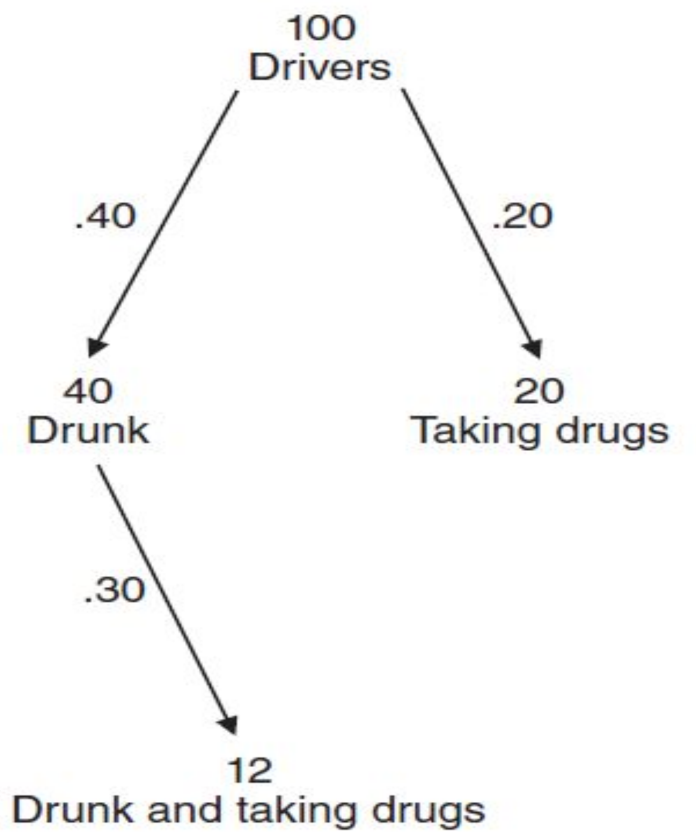
Conditional Probability

- *Conditional probability is a measure of the probability of an event occurring given that another event has occurred.*
- *If the event of interest is A and the event B is known or assumed to have occurred*
- Two dependent events occur together, the probability of the second event must be adjusted to reflect its dependency on the prior occurrence of the first event

- A math teacher gave her class two tests. 25% of the class passed both tests and 42% of the class passed the first test. What percent of those who passed the first test also passed the second test?
    - $P(A \cup B) / P(A)$

- Examples of conditional probabilities are
  - The probability that you will earn a grade of A in a course, given that you have already gotten an A on the midterm,
  - The probability that you'll graduate from college, given that you've already completed the first two years

Alternative Approach to Conditional Probabilities



- "What is the conditional probability of being drunk, given that the driver takes illegal drugs?"

- “What is the conditional probability of being drunk, given that the driver takes illegal drugs?”

- Referring to Figure 8.1, divide the number of drivers who are drunk and take drugs, 12, by the number of drivers who take drugs, 20, that is, 12/20 = .60. (This conditional probability of .60, given drivers who take drugs

- is twice that of .30, given drunk drivers, because the fixed number of drivers who are drunk and take drugs, 12, represents proportionately more (.60) among the relatively small number of drivers who take drugs, 20, and proportionately less (.30) among the relatively large number of drunk drivers, 40.)

- You are off to soccer, and want to be the Goalkeeper, but that depends who is the Coach today:
  - with Coach Sam the probability of being Goalkeeper is **0.5**
  - with Coach Alex the probability of being Goalkeeper is **0.3**
- Sam is Coach more often ... about 6 out of every 10 games (a probability of **0.6**).
- So, what is the probability you will be a Goalkeeper today?

- Among 100 couples who had undergone marital counseling, 60 couples described their relationships as improved, and among this latter group, 45 couples had children. The remaining couples described their relationships as unimproved, and among this group, 5 couples had children. (Hint: Using a frequency analysis, begin with the 100 couples, first branch into the number of couples with improved and unimproved relationships, then under each of these numbers, branch into the number of couples with children and without children. Enter a number at each point of the diagram before proceeding.)

(a) What is the probability of randomly selecting a couple who described their relationship as improved?

(b) What is the probability of randomly selecting a couple with children?

(c) What is the conditional probability of randomly selecting a couple with children, given that their relationship was described as improved?

(d) What is the conditional probability of randomly selecting a couple without children, given that their relationship was described as not improved?

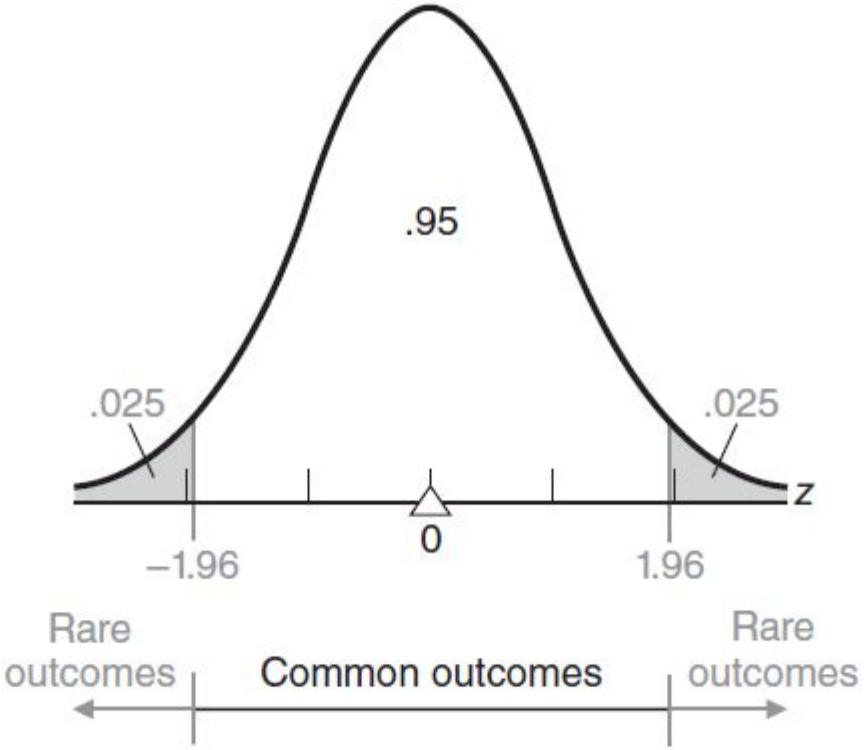
(e) What is the conditional probability of an improved relationship, given that a couple has children?

PROBABILITY AND STATISTICS

- Probability assumes a key role in inferential statistics including, for instance, the important area known as *hypothesis testing*

- **Common Outcomes**

- **Rare Outcomes**



- **The *sampling distribution of the mean* refers to the probability distribution of means for all possible random samples of a given size from some population.**

- The sampling distribution of a statistic is the distribution of all possible values taken by the statistic when all possible samples of a fixed size n are taken from the population. It is a theoretical idea—we do not actually build it.

- The sampling distribution of a statistic is the probability distribution of that statistic

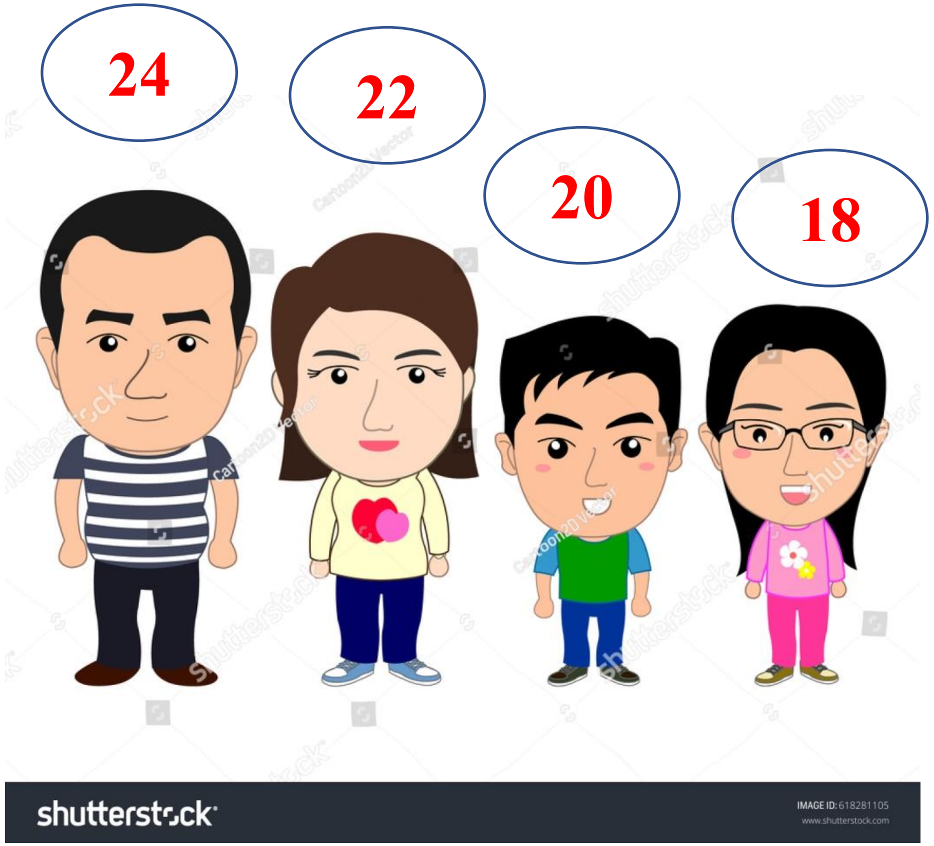
- For example suppose you sample 50 students from your class regarding their mean GPA. If you obtained many different samples of 50 , you would compute a different mean for each sample. We are interested in the distribution of the mean GPA from all possible samples of 50 students.

Sampling Distribution
of the Mean

- Sampling distribution is a distribution of all possible values of a sample statistics for a given size sample selected from a population.

Creating A Sampling Distribution

- Assume there is a population….
  - Population size N = 4, each belongs to one category
  - Random variable x , is age of individuals
  - Values of x is 18,20,22,24 (years)

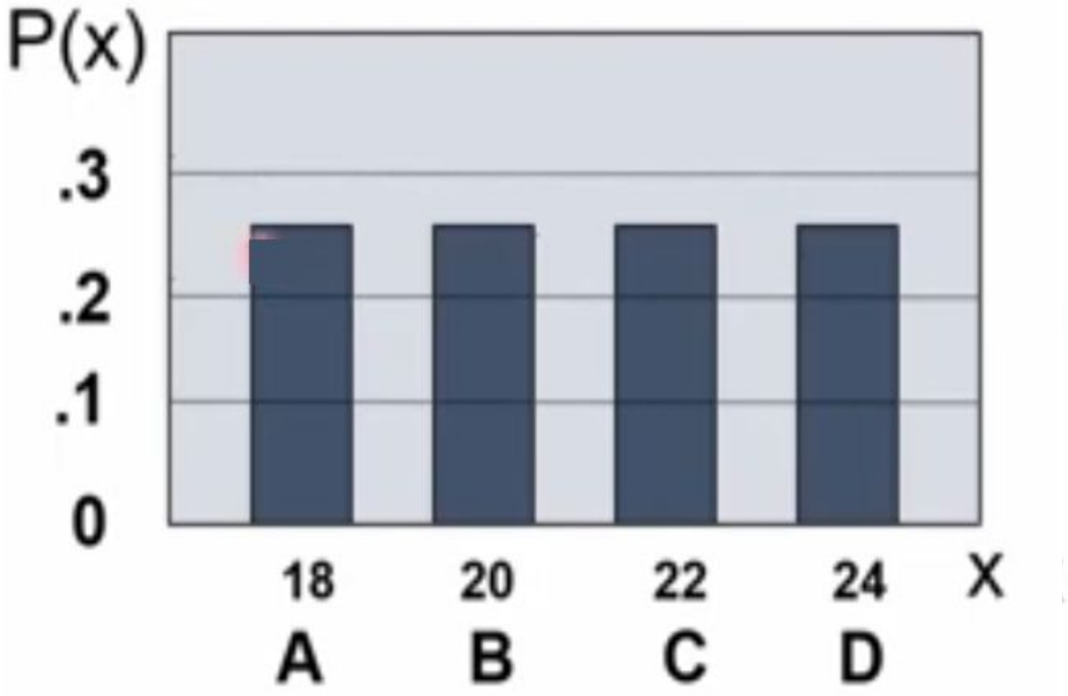


Summary measures of population distribution

$$\mu = \frac{\sum X_i}{N}$$

If you consider only 1 from each category then the sample size n = 1

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}} = \boxed{}$$



- Now consider the sample size of n = 2

1st Obs	2nd Observation			
	18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

16 Possible samples (sampling with replacement)

1st Obs	2nd Observation			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

- Then calculate the mean value of each of these samples sizes

16 sample means

Construct a Relative frequency table showing the sampling distribution of the mean.

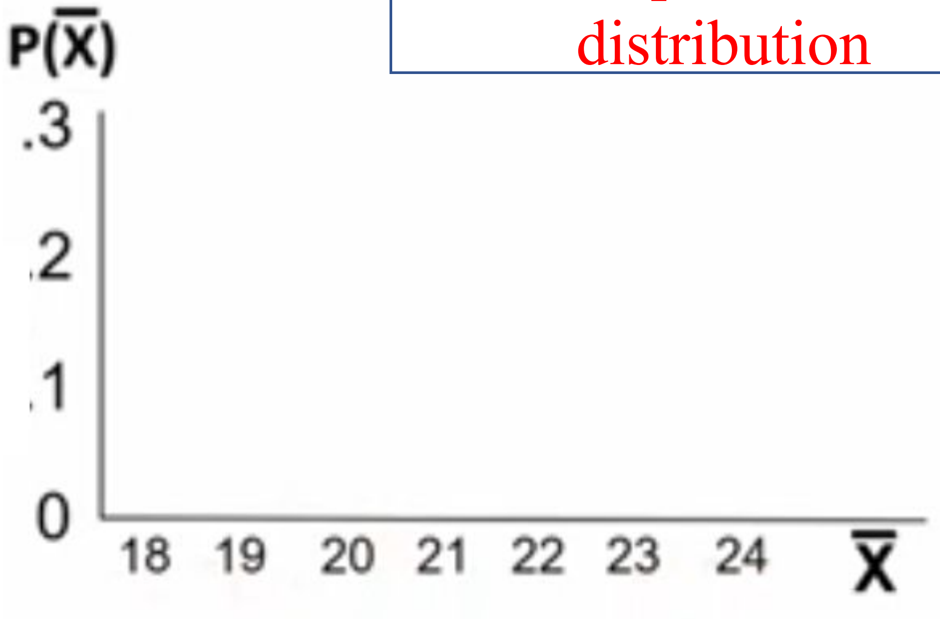
All possible samples	Mean($\bar{X}$)	Probability
18,18	18	1/16
18,20	19	1/16
18,22	20	1/16
18,24	21	1/16
20,18	19	1/16
20,20	20	1/16
20,22	21	1/16
20,24	22	1/16

Mean($\bar{X}$)	Probability
18	1/16
19	2/16
20	3/16
21	4/16
22	3/16
23	2/16
24	1/16

16 sample means

1st Obs	2nd Observation			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Sample means distribution

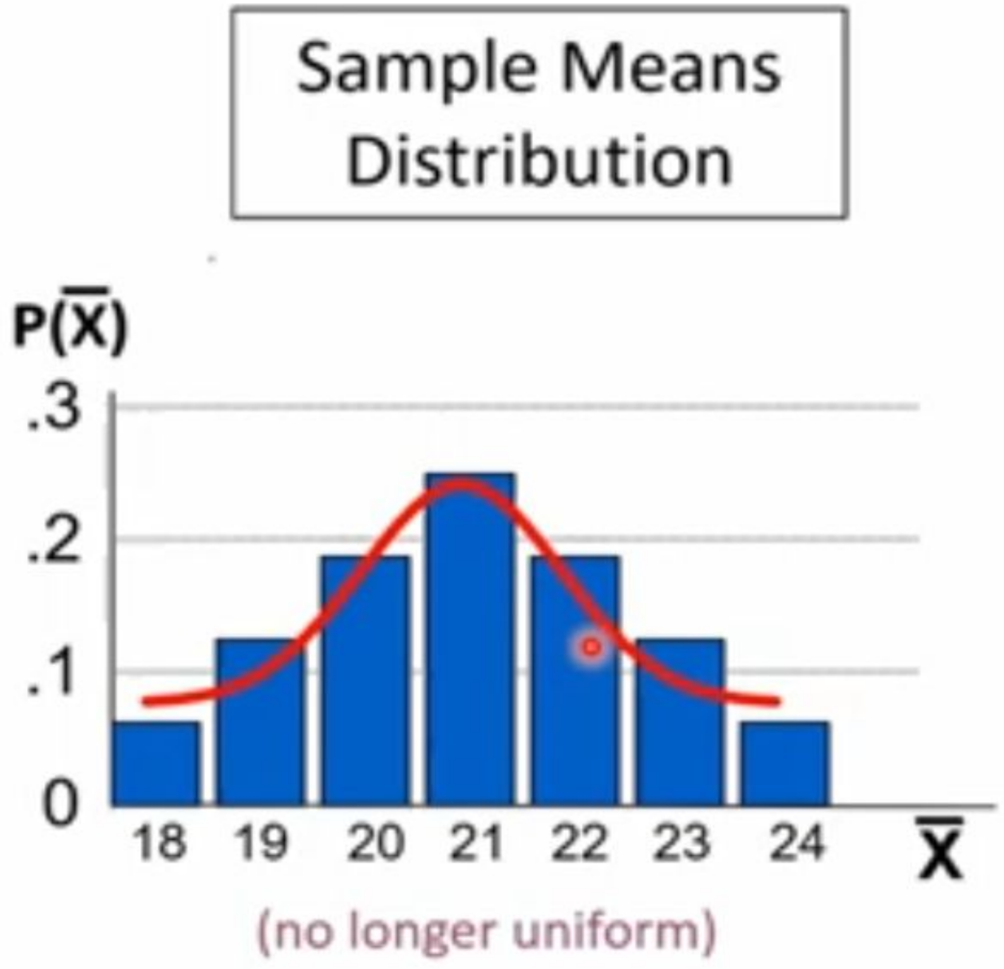


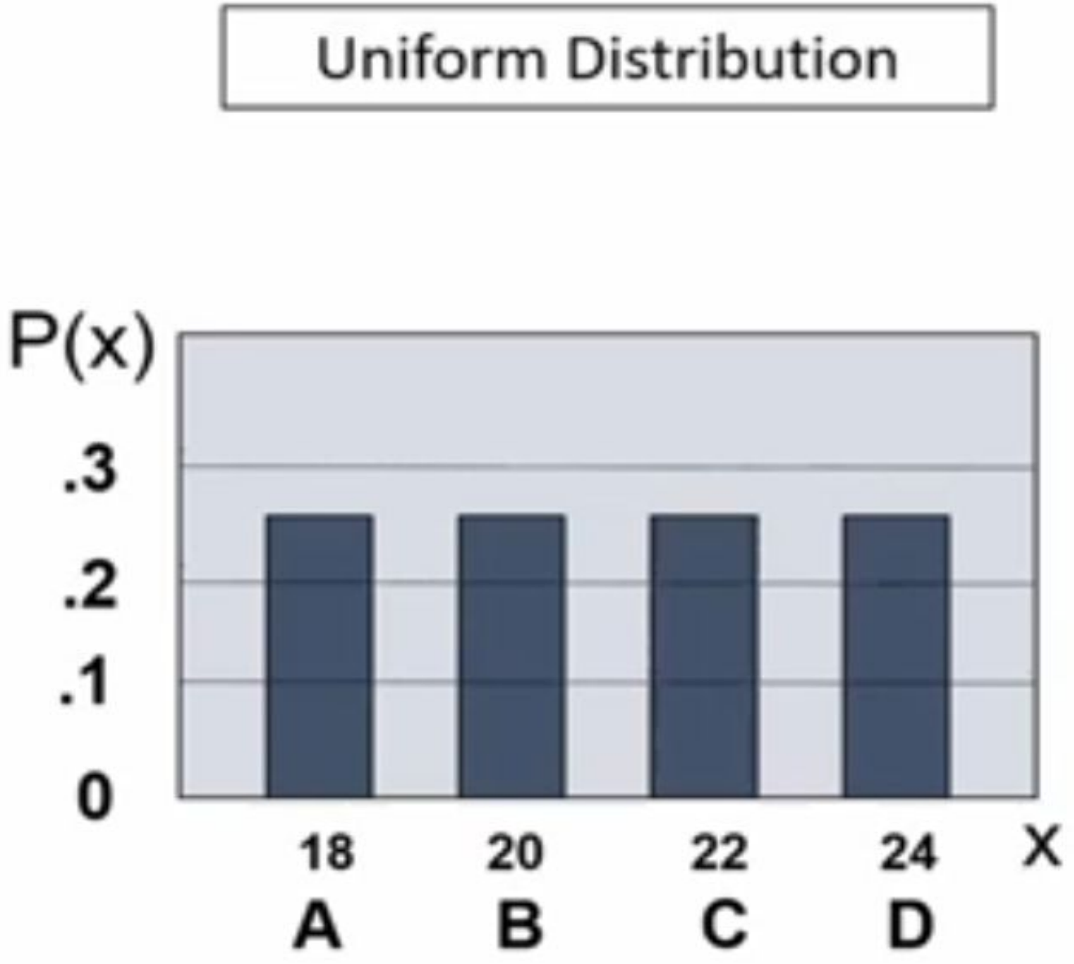
- Plot the probability vs mean and explore the distribution
- The probability of the original population is uniform whereas the probability of the sample means is not uniform and it forms normal distribution

- Imagine a very simple population consisting of only five observations:
    - 2, 4, 6, 8, 10.

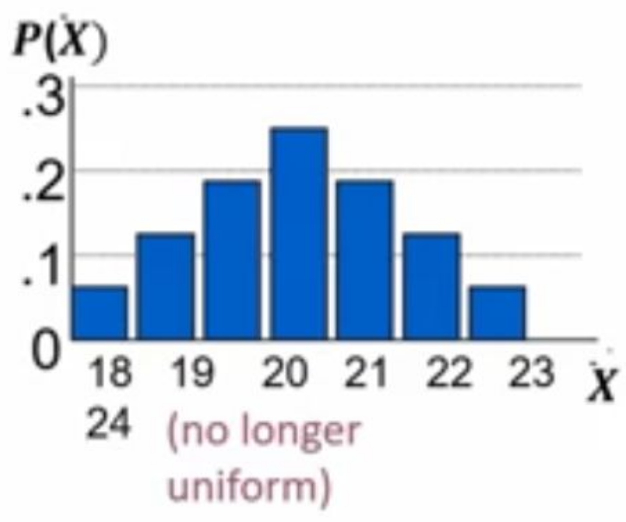
(a) List all possible samples of size two.

(b) Construct a relative frequency table showing the sampling distribution of the mean.


Sample Means Distribution
P(X̄)
.3
.2
.1
0
18
19
20
21
22
23
24
X̄
(no longer uniform)


Uniform Distribution
P(x)
.3
.2
.1
0
18
20
22
24
x
A
B
C
D

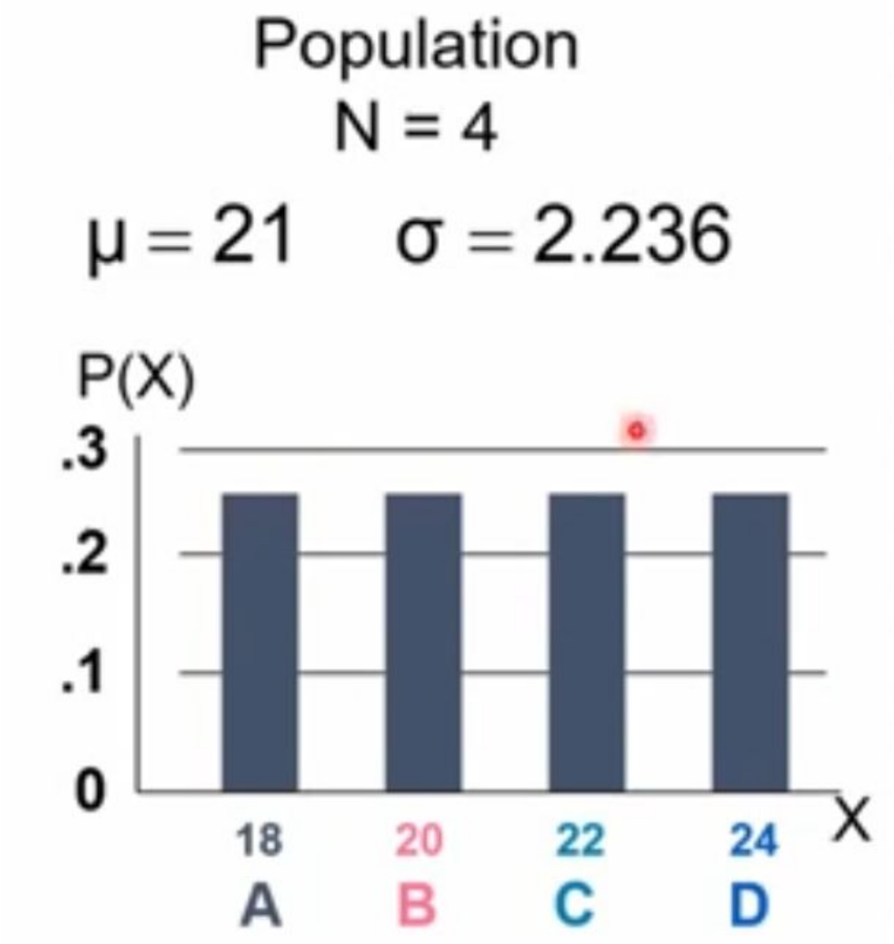
Summary measures of this Sampling Distribution

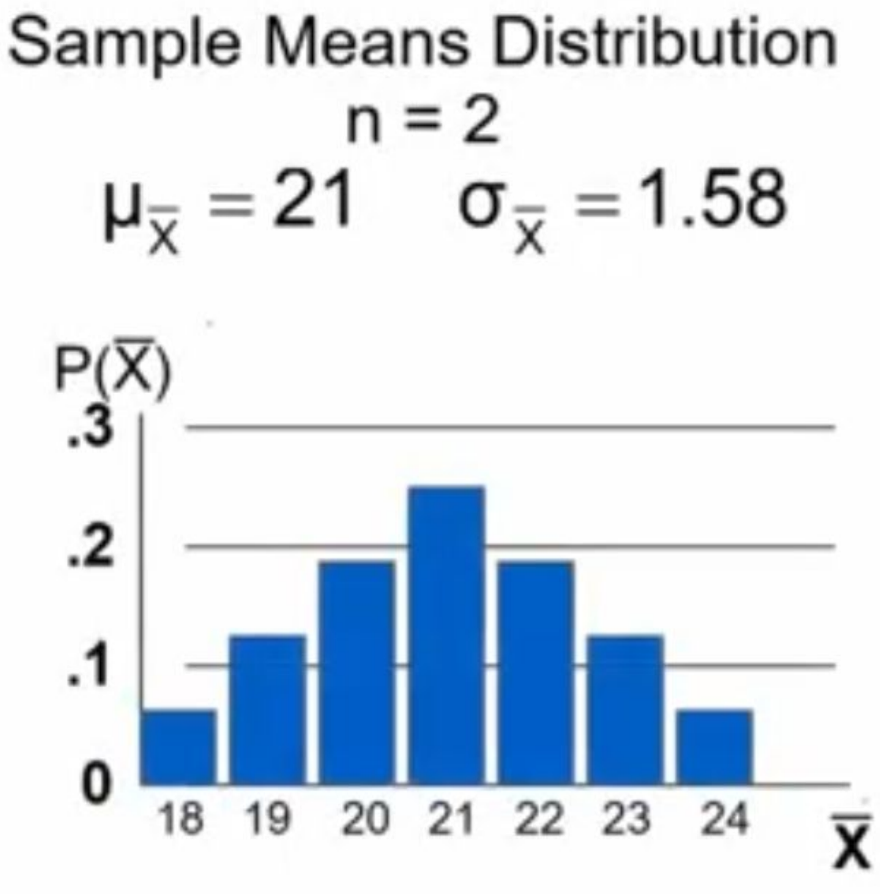


$$\mu_{\bar{X}} = \frac{18+19+19+\cdots+24}{16} = 21$$

$$\boxed{\sigma_{\bar{X}} = \sqrt{\frac{(18-21)^2+(19-21)^2+\cdots+(24-21)^2}{16}} = 1.58}$$

Comparing the population mean to the sample means distribution





Sampling distribution of the mean : Standard error of the mean

- Different samples of the same size from the same population will yield different sample means.
- It roughly measures the average amount by which sample means deviate from the population mean.
- This difference(Variability) is measured by standard error of the n

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

- Standard error of the mean decreases as the sample size increases from the given formula(square root (n))

Sampling Distribution of the Mean: If the population is Normal

- If the population is normal with mean μ and standard deviation σ, the sampling distribution of $\bar{X}$ is also normally distributed

$$\boxed{\mu_{\bar{X}} = \mu} \quad \text{and} \quad \boxed{\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}}$$

- Indicate whether the following statements are True or False. The standard error of the mean, X, . . .

(a) roughly measures the average amount by which sample means deviate from the population mean.

(b) measures variability in a particular sample.

(c) increases in value with larger sample sizes.

(d) equals 5, given that $\sigma = 40$ and $n = 64$.

SHAPE OF THE SAMPLING DISTRIBUTION

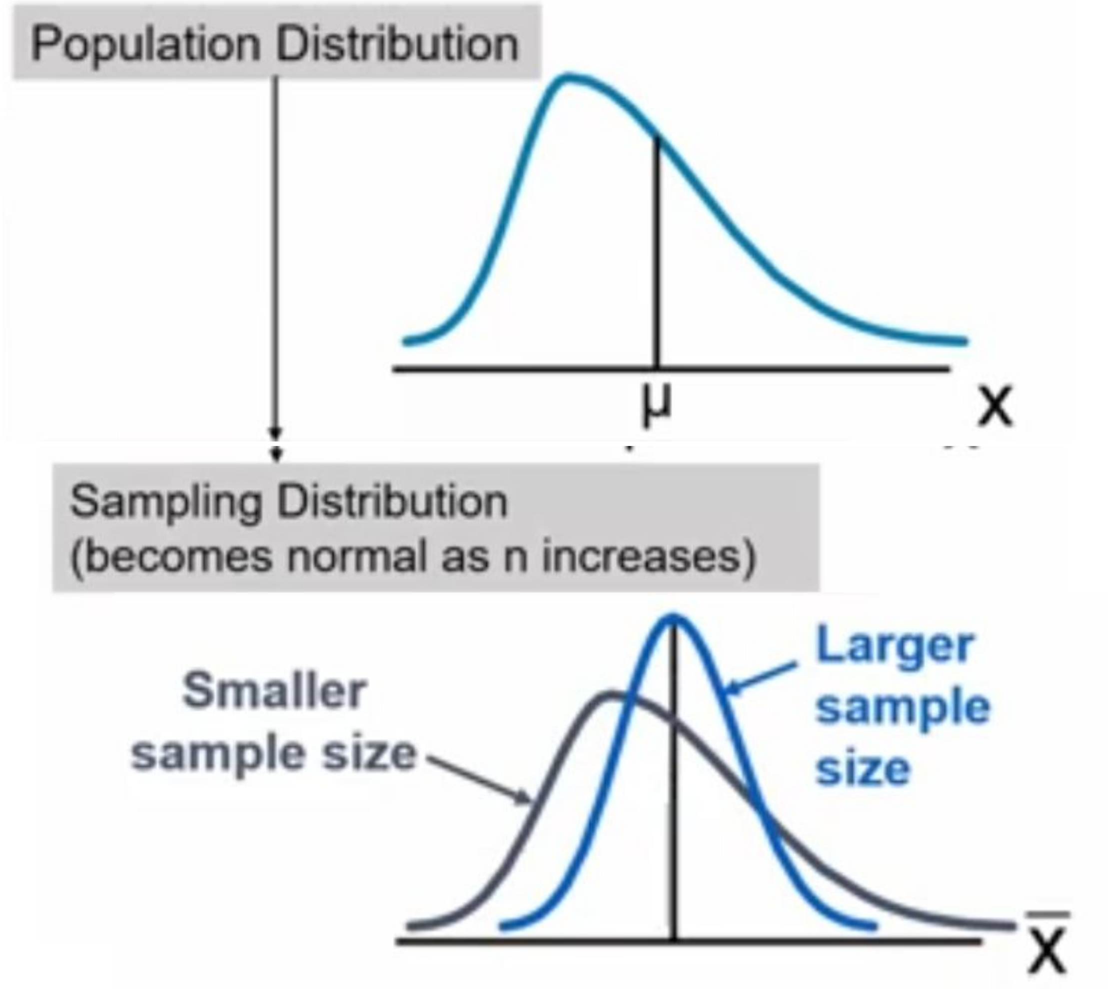
- The *central limit theorem* states that, regardless of the shape of the population, the shape of the sampling distribution of the mean approximates a normal curve *if the sample size is sufficiently large*.

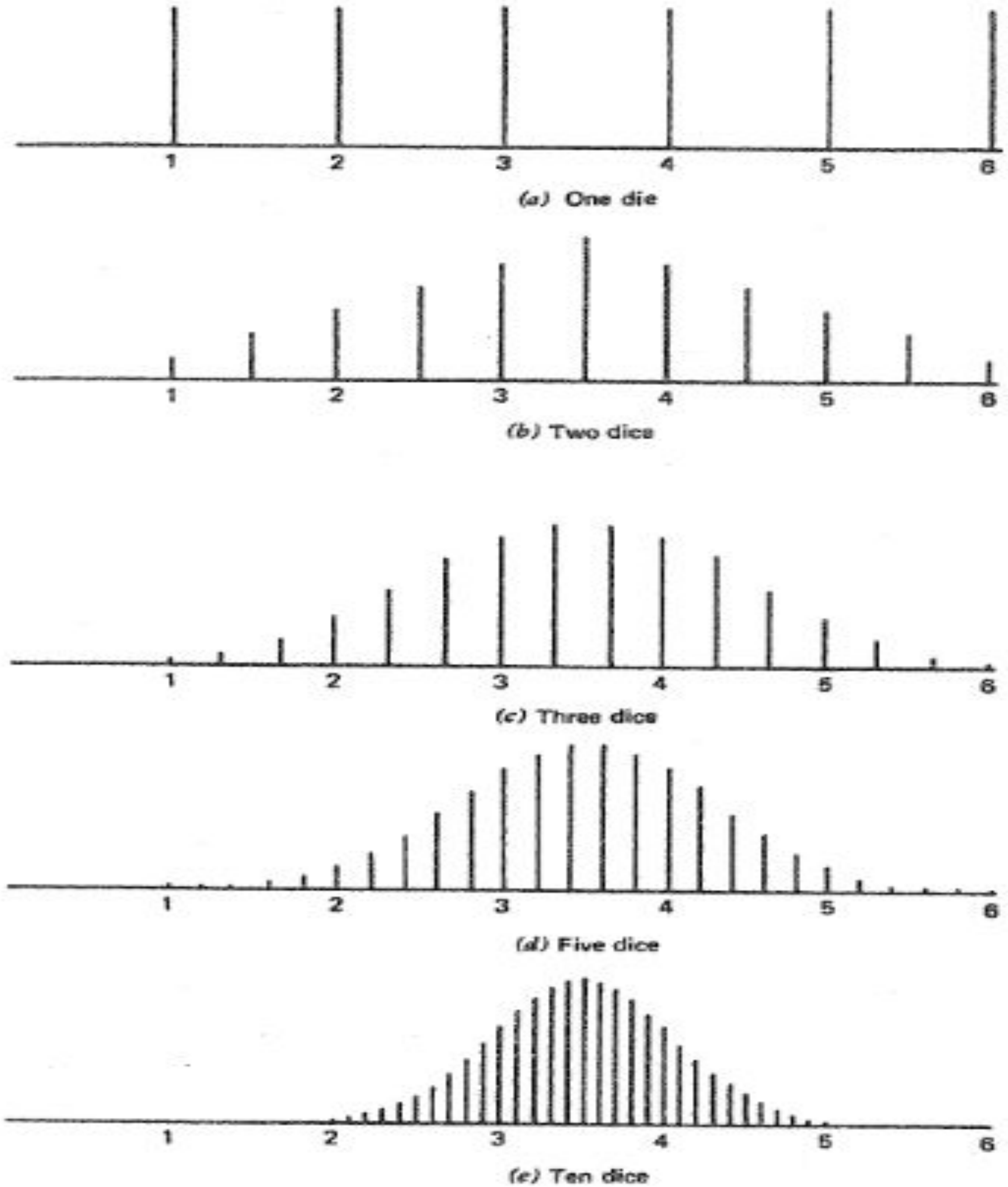
- Even if the population is not normal

- Sample means from the population will be approximately normal as long as sample size is large enough

$$\boxed{\mu_{\bar{x}} = \mu} \quad \text{and} \quad \boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

Sample mean sampling distribution : if the population is not normal





(a) One die

(b) Two dice

(c) Three dice

(d) Five dice

(e) Ten dice

Introduction to Hypothesis Testing: The *z* Test

Definitions

- Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter.

- A hypothesis is **an assumption, an idea that is proposed for the sake of argument so that it can be tested to see if it might be true**

- Hypothesis testing is **a form of statistical inference that uses data from a sample to draw conclusions about a population parameter or a population probability distribution**

- Assume that the SAT scores for all college-bound students during a recent year were distributed around a mean of 500 with a standard deviation of 110

- An investigator at a university wishes to test the claim that, on the average, the SAT math scores for local freshmen equals the national average of 500

- Assume that it is not possible to obtain scores for the entire freshman class.

- Instead, SAT math scores are obtained for a random sample of 100 students from the local population of freshmen, and the mean score for this sample equals 533.

Hypothesized Sampling Distribution

- A null hypothesis is a type of hypothesis in statistics that proposes that there is no difference between certain characteristics of a population (or data-generating process).
  - Abbreviated as *H*0, the null hypothesis

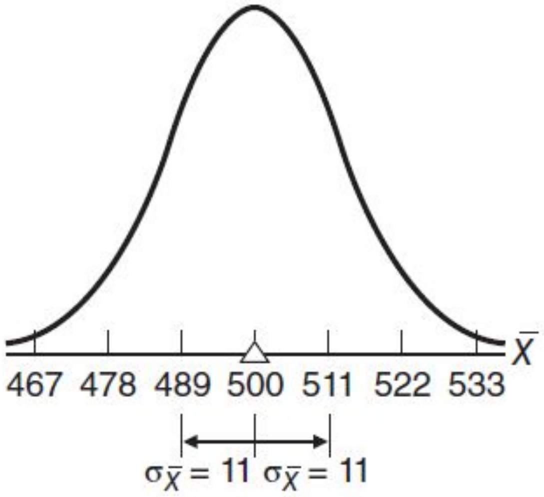
- The **null hypothesis (*H*0)** *is a statistical hypothesis that usually asserts that nothing special is happening with respect to some characteristic of the underlying population.*

- The alternative hypothesis proposes that there is a difference
    - Abbreviated as *H*a, or H1

- *The* **alternative hypothesis (*H*1)** *asserts the opposite of the null hypothesis.*

Retain or Reject

- If the null hypothesis is true, then the distribution of sample means—that is, the sampling distribution of the mean for all possible random samples, each of size 100, from the local population of freshmen—will be centered about the national average of 500.



- *Hypothesized sampling distribution of the mean centered about a hypothesized population mean of 500.*

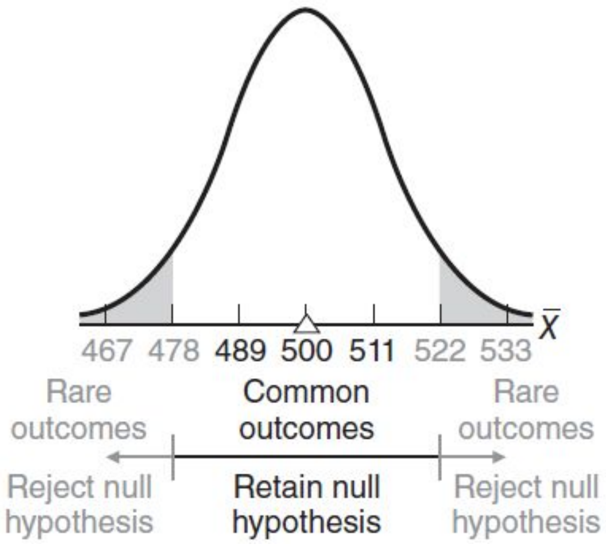
- In Figure 10.1, vertical lines appear, at intervals of size 11, on either side of the hypothesized population mean of 500.
- Substitute 110 for the population standard deviation, σ, and 100 for the sample size, n, in Formula 9.2 to obtain

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{110}{\sqrt{100}} = \frac{110}{10} = 11$$

- Notice that the shape of the hypothesized sampling distribution in Figure approximates a normal curve, since the sample size of 100 is large enough to satisfy the requirements of the central limit theorem

- Common Outcomes
  - An observed sample mean qualifies as a common outcome if the difference between its value and that of the hypothesized population mean **is small enough** to be viewed as a probable outcome under the null hypothesis.
- Rare Outcomes
  - An observed sample mean qualifies as a rare outcome if the difference between its value and the hypothesized population mean **is too large** to be reasonably viewed as a probable outcome under the null hypothesis

Boundaries for Common and Rare Outcomes



- *One possible set of common and rare outcomes (values of X).*
- ***Conclusion***
  - The mean math score for the local population of freshmen probably differs from (exceeds) the national average and hence the null hypothesis is rejected.

z TEST FOR A POPULATION MEAN

- The sampling distribution of z represents the distribution of z values that would be obtained if a value of z were calculated for each sample mean for all possible random samples of a given size from some population

- The conversion from $\bar{X}$ to z yields a distribution that approximates the standard normal curve in Table A of Appendix C

- the original hypothesized population mean (500) emerges as a z score of 0 and the original standard error of the mean (11) emerges as a z score of 1

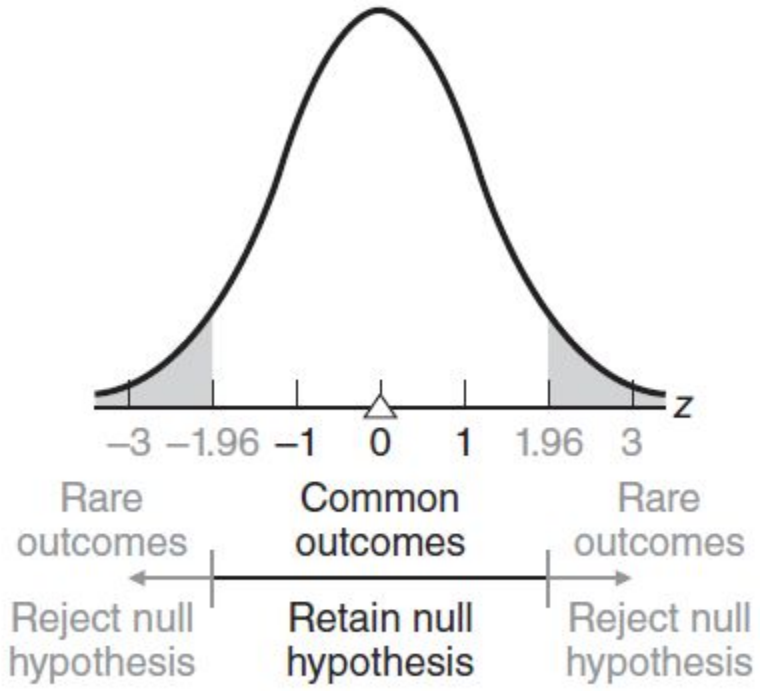


FIGURE 10.3
Common and rare outcomes (values of z).

Reminder: Converting a Raw Score to *z*

$$\text{Standard score} = \frac{\text{raw score} - \text{mean}}{\text{standard deviation}}$$

- **Converting a Sample Mean to *z***

z RATIO FOR A SINGLE POPULATION MEAN

$$z = \frac{\bar{X} - \mu_{\text{hyp}}}{\sigma_{\bar{x}}}$$

- Calculate the value of the z test for each of the following situations:

(a) $X = 566$; $\sigma = 30$; $n = 36$; hyp $= 560$

(b) $X = 24$; $\sigma = 4$; $n = 64$; hyp $= 25$

(c) $X = 82$; $\sigma = 14$; $n = 49$; hyp $= 75$

(d) $X = 136$; $\sigma = 15$; $n = 25$; hyp $= 146$

STEP-BY-STEP PROCEDURE

HYPOTHESIS TEST SUMMARY: z TEST FOR A POPULATION MEAN (SAT SCORES)

Research Problem

Does the mean SAT math score for all local freshmen differ from the national average of 500?

Statistical Hypotheses

$$H_0 : \mu = 500$$
$$H_1 : \mu \neq 500$$

Decision Rule

Reject H_0 at the .05 level of significance if $z \geq 1.96$ or if $z \leq -1.96$.

Calculations

Given

$$\overline{X} = 533;\ \mu_{\text{hyp}} = 500;\ \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{110}{\sqrt{100}} = 11$$

$$z = \frac{533 - 500}{11} = 3$$

Decision

Reject H_0 at the .05 level of significance because $z = 3$ exceeds 1.96.

Interpretation

The mean SAT math score for all local freshmen does not equal—it exceeds—the national average of 500.

Progress Check *10.2 Indicate what's wrong with each of the following statistical hypotheses:

(a) $H_0: \mu = 155$
$H_1: \mu \neq 160$

(b) $H_0: \bar{X} = 241$
$H_1: \bar{X} \neq 241$

Progress Check *10.3 First using words, then symbols, identify the null hypothesis for each of the following situations. (Don't concern yourself about the precise form of the alternative hypothesis at this point.)

(a) A school administrator wishes to determine whether sixth-grade boys in her school district differ, on average, from the national norms of 10.2 pushups for sixth-grade boys.

(b) A consumer group investigates whether, on average, the true weights of packages of ground beef sold by a large supermarket chain differ from the specified 16 ounces.

(c) A marriage counselor wishes to determine whether, during a standard conflict-resolution session, his clients differ, on average, from the 11 verbal interruptions reported for "well-adjusted couples."

DECISION RULE

- *A* **decision rule** *specifies precisely when H0 should be rejected* (because the observed *z* qualifies as a rare outcome).

- **Level of Significance (α)**
  - The **level of significance** (*α*) *indicates the degree of rarity required of an observed outcome in order to reject the null hypothesis* (*H*0).

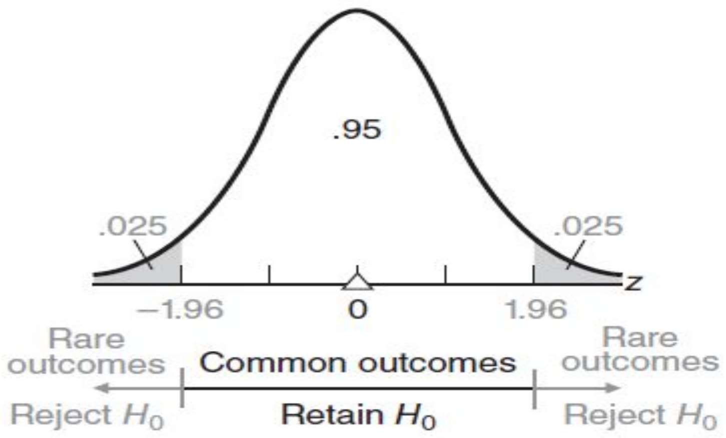


FIGURE 10.4
Proportions of area associated with common and rare outcomes (α = 05).

- For each of the following situations, indicate whether $H0$ should be retained or rejected and justify your answer by specifying the precise relationship between observed and critical z scores. Should $H0$ be retained or rejected, given a hypothesis test with critical z scores of $\pm$ 1.96 and

(a) $z = 1.74$

(b) $z = 0.13$

(c) $z = -2.51$

- A botanist claims that a particular strain of tomatoes developed by her lab is more disease resistant and is hence more durable.
- She conducts several trial studies with different batches of tomatoes and assigns a disease-resistance-quality score to each batch. A random sample of 30 batches of tomatoes is taken, and is found to have a mean quality score 168.
- Is there sufficient evidence to support the researcher's claim? The mean quality score of the entire tomato yield is standardized to 100, with a standard deviation of 22. Assume a normal distribution.
- Describe the process of proving your hypothesis in detail and derive relevant proof with valid reasoning.