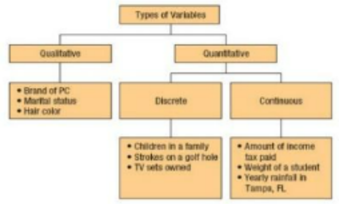
CHAPTER 1 : What is Statistic?

1. **Statistic** is one of a tools used to make decisions in business.
2. **Why study statistics?** Used as a basic knowledge for making a decision.
3. **Types of statistics :**
   a. **Descriptive Statistics**
      Organize, summarize and present data
   b. **Inferential Statistics**
      Analyze sample data and the results are applied to the population
4. **Population VS Sample**
   a. **Population** is a collection of all possible object
   b. **Sample** is a portion or part of the population
5. **Types of Variables :**
   a. **Qualitative** (the information is non-numeric)
   b. **Quantitative** (the information is numerically)



6. **Four level of measurement :**
   a. **Nominal** (the data have no order)
   b. **Ordinal** (data arranged in some order)
   c. **Interval Level** (there's no natural zero point)
   d. **Ratio** (zero as a starting point)

CHAPTER 2 : Describing Data – Freq Table, Freq Distribution

1. **Qualitative Data**
   a. Frequency Table
   b. Relative Frequency Table
   c. Graphic presentation : bar chart, pie chart
2. **Quantitative Data**
   a. **Frequency Distribution** (Class Interval, Class Frequency, Class Midpoint)
   b. **Relative Frequency Distribution** (To convert freq dist to relative freq, each of class divided by the sum or total no of observations)
   c. **Graphic presentation** : Histogram, Polygon
   d. **Cumulative Distribution** (determine how many observation lie above or below certain values)

CHAPTER 3 : Describing Data – Numerical Measure

A. Measure of Location :

1. **Mean (Ungrouped Data)**
   a. **Arithmetic Mean**
      - **Population Mean**
        $\mu = \frac{\sum x}{N}$ ($\sum x$ = total particular value) (N = no of values in the population)

        $$\mu = \frac{\Sigma X}{N} = \frac{11+4+10+\cdots+1}{42} = \frac{192}{42} = 4.57$$

      - **Sample Mean**
        $\bar{x} = \frac{\sum x}{n}$ ($\sum x$ = total particular value) (n = no of values in the sample)

        $$\text{Sample mean} = \frac{\text{Sum of all values in the sample}}{\text{Number of values in the sample}}$$
        $$\bar{X} = \frac{\Sigma X}{n} = \frac{90+77+\cdots+83}{12} = \frac{1170}{12} = 97.5$$

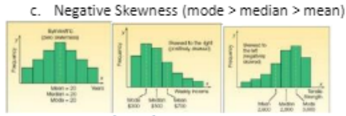
      - **Weighted Mean**
        $$\bar{X}_w = \frac{w_1X_1 + w_2X_2 + \dots + w_nX_n}{w_3 + w_3 + \dots + w_n}$$
        $$\bar{X}_w = \frac{14(\$16.50)+10(\$19.00)+2(\$25.00)}{14+10+2} = \frac{\$471.00}{26} = \$18.1154$$

   b. **Geometric Mean** (always less than or equal to arithmetic mean)
      $GM = \sqrt{(X_1)(X_2)\dots(X_n)}$

      EXAMPLE
      The return on investment earned by Atkins Construction Company for four successive years was: 30 percent, 20 percent, -40 percent, and 200 percent. What is the geometric mean rate of return on investment?

      $$GM = \sqrt[n]{(X_1)(X_2)\cdots(X_n)} = \sqrt[4]{(1.3)(1.2)(0.6)(3.0)} = \sqrt[4]{2.808} = 1.294$$

2. **Median** (the midpoint of the value)
3. **Mode** (The value that appears most frequently)
4. **Relative Positions of Mean, Median Mode**
   a. Zero Skewness (mode = median = mean)
   b. Positive Skewness (mode < median < mean)
   c. Negative Skewness (mode > median > mean)



B. Measure of Dispersion :

1. **Range**
   Range = Largest value – Smallest value
2. **Mean Deviation**
   $MD = \frac{\sum |X - \bar{X}|}{n}$ (x = value of each observation) ($\bar{x}$ = arithmetic mean of x values) (n = number of sample)

Number of Cappuccinos Sold Daily	$(X-\bar{X})$	Absolute Deviation
20	(20 − 50) = −30	30
40	(40 − 50) = −10	10
50	(50 − 50) = 0	0
60	(60 − 50) = 10	10
80	(80 − 50) = 30	30
	Total	80

$$MD = \frac{\Sigma|X-\bar{X}|}{n} = \frac{80}{5} = 16$$

3. **Variance**
   a. **Population Variance**

$$\sigma^2 = \frac{\Sigma(X-\mu)^2}{N}$$
$$\mu = \frac{\sum x}{N} = \frac{19+17+\dots+34+10}{12} = \frac{348}{12} = 29$$

Month	Citations (X)	$X-\mu$	$(X-\mu)^2$
January	19	−10	100
February	17	−12	144
March	22	−7	49
April	18	−11	121
May	28	−1	1
June	34	5	25
July	45	16	256
August	39	10	100
September	38	9	81
October	44	15	225
November	34	5	25
December	10	−19	361
Total	348	0	1,488

   b. **Sample Variance**

$$s^2 = \frac{\Sigma(X-\bar{X})^2}{n-1}$$

4. **Standard Deviation**
   a. **Population Standard Deviation**

$$\sigma = \sqrt{\frac{\Sigma(X-\mu)^2}{N}}$$

   b. **Sample Standard Deviation**

$$s = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n-1}}$$

C. Grouped Data :

1. **The Mean of Grouped Data**

$$\bar{X} = \frac{\Sigma fM}{n}$$

Profit	Frequency (f)	Midpoint (M)	fM
$ 200 up to $ 600	8	$ 400	$ 3,200
600 up to 1,000	11	800	8,800
1,000 up to 1,400	23	1,200	27,600
1,400 up to 1,800	38	1,600	60,800
1,800 up to 2,200	45	2,000	90,000
2,200 up to 2,600	32	2,400	76,800
2,600 up to 3,000	19	2,800	53,200
3,000 up to 3,400	4	3,200	12,800
Total	180		$333,200

Solving for the arithmetic mean using formula 3-12, we get:

$$\bar{X} = \frac{\Sigma fM}{n} = \frac{\$333,200}{180} = \$1,851$$

2. **Standard Deviation of Grouped Data**

Profit	Frequency (f)	Midpoint (M)	fM	$(M-\bar{X})$	$(M-\bar{X})^2$	$f(M-\bar{X})^2$
$ 200 up to $ 600	8	400	3,200	−1,451	2,105,401	16,843,208
600 up to 1,000	11	800	8,800	−1,051	1,104,601	12,150,611
1,000 up to 1,400	23	1,200	27,600	−651	423,801	9,747,423
1,400 up to 1,800	38	1,600	60,800	−251	63,001	2,394,038
1,800 up to 2,200	45	2,000	90,000	149	22,201	999,045
2,200 up to 2,600	32	2,400	76,800	549	301,401	9,644,832
2,600 up to 3,000	19	2,800	53,200	949	900,601	17,111,419
3,000 up to 3,400	4	3,200	12,800	1,349	1,819,801	7,279,204
Total	180		333,200			76,169,780

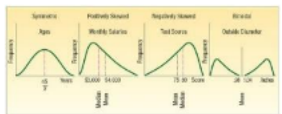
$$s = \sqrt{\frac{\Sigma f(M-\bar{X})^2}{n-1}} = \sqrt{\frac{76,169,780}{180-1}} = 652.33$$

CHAPTER 4 : Describing Data – Displaying Exploring

1. **Displaying data** (Dot Plots, Box Plots, Steam-and-Leaf, Scatter Plots, Contingency Tables)
2. **Measure of Pos** (Quartiles, Deciles, Percentiles)

$$L_p = (n+1)\frac{P}{100}$$

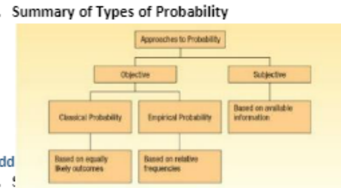
3. **Skewness**



CHAPTER 5 : Probability Concepts

A. Assigning Probability

1. **Classical Probability** (assumption the outcomes are equally likely)
   Classical = $\frac{\text{Number of favorable outcomes}}{\text{Total number of possible outcomes}}$
   a. **Mutually Exclusive** if the occurrence of any one event means that none of the others can occur at the same time
   b. **Independent** if one event doesn't affect the occurrence of another
   c. **Collectively Exhaustive** if at least one of the events occur when an experiment conducted
2. **Empirical Probability** (based on what happened in the past)
   Empirical = $\frac{\text{No of time the event occurs}}{\text{Total no of observation}}$
3. **Subjective Concept of Probability** (based on available information)
   e.g : estimating the likelihood you will be married before age 30
4. **Summary of Types of Probability**



B. Add

1. **S**
   P(A or B) = P(A) + P(B)

Weight	Event	Number of Packages	Probability of Occurrence
Underweight	A	100	.025
Satisfactory	B	3,600	.900
Overweight	C	300	.075
		4,000	1.000

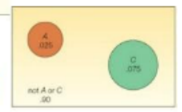
← 100/4,000

P(A or C) = P(A) + P(C) = .025 + .075 = .10

2. **General Addition (Not Mutually Exclusive)**
   P(A or B) = P(A) + P(B) – P(A and B)
   P(Disney or Busch) = P(Disney) + P(Busch) - P(both Disney and Busch) = 120/200 + 100/200 – 60/200 = .60 + .50 – .80
3. **The Complement of Addition**
   P(A) = 1-P(~A)

Weight	Event	Number of Packages	Probability of Occurrence
Underweight	A	100	.025
Satisfactory	B	3,600	.900
Overweight	C	300	.075
		4,000	1.000

P(B) = 1 - P(~B)
= 1 – P(A or C)
= 1 – [P(A) + P(C)]
= 1 – [.025 + .075]
= 1 - .10
= .90



C. Multification to Calculate Probability

1. **Special Multification**
   P(A and B) = P(A)P(B)
2. **General Multification**
   P(A and B) = P(A)P(B|A)

D. Classify Sample Observation

1. **Contingency Table** (for 2 or more characteristics)
2. **Tree Diagram** (for conditional/joint probabilities)

E. Bayes Theorem (is a method for revising a probability given additional information)

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)}$$

F. Counting Rules

1. **Permutation** (order of arrangement)

$$_nP_r = \frac{n!}{(n-r)!}$$
n = total no of objects
r = no of object selected

2. **Combination** (without regard to order)

$$_nC_r = \frac{n!}{r!(n-r)!}$$
n = total no of objects
r = no of object selected

CHAPTER 6 : Discrete Probability Distributions

1. **Probability Distribution** is a listing of all the outcomes of an experiment and the probability associated with each outcomes
2. **Characteristics** : outcome between 0 and 1 inclusive, outcomes are mutually exclusive, list is exhaustive. So the sum of event is equal 1.
3. **Random Variables** (result from experiment that assume diff values)
   a. **Discrete Random Variables** can assume only certain separated values. Result of counting something. e.g : no of students in class
   b. **Continuous Random Variables** an infinite no of values within range. Result of measurement. e.g : the weight of each student
4. **Mean of Probability Distribution**

$$\mu = \Sigma[xP(x)]$$
$\mu = \Sigma[xP(x)] = 0(.10) + 1(.20) + 2(.30) + 3(.30) + 4(.10) = 2.1$

Number of Cars Sold, x	Probability, P(x)	x · P(x)
0	.10	0.00
1	.20	0.20
2	.30	0.60
3	.30	0.90
4	.10	0.40
Total	1.00	μ = 2.10

5. **Variance of Probability Distribution**

$$\sigma^2 = \Sigma[(x-\mu)^2P(x)]$$

Number of Cars Sold, x	Probability, P(x)	$(x-\mu)$	$(x-\mu)^2$	$(x-\mu)^2P(x)$
0	.10	0 − 2.1	4.41	0.441
1	.20	1 − 2.1	1.21	0.242
2	.30	2 − 2.1	0.01	0.003
3	.30	3 − 2.1	0.81	0.243
4	.10	4 − 2.1	3.61	0.361
				σ^2 = 1.290

6. **Standard Deviation of Probability Distribution**

$$\sigma = \sqrt{\sigma^2} = \sqrt{1.290} = 1.136$$

7. **Binomial Distribution**
   Characters : only two possible outcomes, the outcomes are mutually exclusive (success or failure), the random variable is the result of counts and each trial is independent of any other trial

$$P(x) = {}_nC_x\,\pi^x(1-\pi)^{n-x}$$

C denotes a combination.
n is the number of trials.
x is the random variable defined as the number of successes.
π is the probability of a success on each trial.

Example :

a. **Mean of Binomial Distribution**

For the example regarding the number of late flights, recall that π = .20 and n = 5.

What is the average number of late flights?
$\mu = n\pi = (5)(0.20) = 1.0$

There are five flights daily from Pittsburgh via US Airways into the Bradford Regional Airport in PA. Suppose the probability that any flight arrives late is .20. What is the probability that none of the flights are late today?

$$P(x=0) = {}_nC_x\pi^x(1-\pi)^{n-x} = {}_5C_0(.20)^0(1-.20)^{5-0} = (1)(1)(.3277) = 0.3277$$

b. **Variance of Binomial Distribution**

What is the variance of the number of late flights?

$$\sigma^2 = n\pi(1-\pi) = (5)(0.20)(1-0.20) = (5)(0.20)(0.80) = 0.80$$

8. **Hypergeometric Probab**
   Characters : one or two mutually exclusive (success or failure), the probability of success or failure changes from trial to trial, the trial not are not independent

$$P(x) = \frac{({}_SC_x)({}_{N-S}C_{n-x})}{{}_NC_n}$$

N is the size of the population.
S is the number of successes in the population.
x is the number of successes in the sample. It may be 0, 1, 2, 3, ...
n is the size of the sample or the number of trials.
C is the symbol for a combination.

Example :

PlayTime Toys, Inc., employs 50 people in the Assembly Department. Forty of the employees belong to a union and ten do not. Five employees are selected at random to form a committee to meet with management regarding shift starting times.

What is the probability that four of the five selected for the committee belong to a union?

Here's what's given:
N = 50 (number of employees)
S = 40 (number of union employees)
x = 4 (number of union employees selected)
n = 5 (number of employees selected)

$$P(x) = \frac{({}_SC_x)({}_{N-S}C_{n-x})}{{}_NC_n}$$
$$P(4) = \frac{({}_{40}C_4)({}_{50-40}C_{5-4})}{{}_{50}C_5} = \frac{\left(\frac{40!}{4!36!}\right)\left(\frac{10!}{1!9!}\right)}{\frac{50!}{5!45!}} = \frac{(91,390)(10)}{2,118,760} = .431$$

9. **Poisson Probability Distribution**
   Characters : event occurs during a specified interval (time, distance, area, volume), the interval are independent, the probability is proportional to the length of the interval.
   e.g : the no of vehicles sold per day, the no of calls per hour

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

μ (mu) is the mean number of occurrences (successes) in a particular interval.
e is the constant 2.71828 (base of the Napierian logarithmic system).
x is the number of occurrences (successes).
P(x) is the probability for a specified value of x.

Example :

Assume baggage is rarely lost by Northwest Airlines. Suppose a random sample of 1,000 flights shows a total of 300 bags were lost. Thus, the arithmetic mean number of lost bags per flight is 0.3 (300/1,000). If the number of lost bags per flight follows a Poisson distribution with U = 0.3, find the probability of not losing any bags.

$$P(0) = \frac{\mu^x e^{-\mu}}{x!} = \frac{0.3^0 e^{-3}}{0!} = .7408$$

$\mu = n\pi$

b. **Variance of Poisson Distribution**
   $\delta = n\pi$

CHAPTER 7 : Continuous Probability Distribution

Total area within a continuous probability distribution is equal to 1

1. **Uniform Probability Distribution**
   a. **Uniform Distribution**

$$P(x) = \frac{1}{b-a}$$

   b. **Mean of the Uniform Distribution**

$$\mu = \frac{a+b}{2}$$

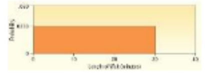
   c. **Standard Deviation of the Uniform Distribution**

$$\sigma = \sqrt{\frac{(b-a)^2}{12}}$$

Example :

Southwest Arizona State University provides bus service to students while they are on campus. A bus arrives at the North Main Street and College Drive stop every 30 minutes between 6 A.M. and 11 P.M. during weekdays. Students arrive at the bus stop at random times. The time that a student waits is uniformly distributed from 0 to 30 minutes.

- **Graph of this distribution**
  Px = $\frac{1}{b-a}$ = 1/30-0 = 0.333



- **Show the area of this distribution is 1.00**

$$\text{Area} = (\text{height})(\text{base}) = \frac{1}{(30-0)}(30-0) = 1.00$$

- **Mean waiting time**

$$\mu = \frac{a+b}{2} = \frac{0+30}{2} = 15$$

- **Standard deviation of waiting time**

$$\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(30-0)^2}{12}} = 8.66$$

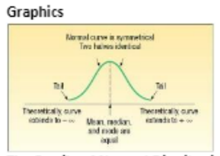
- **Probability the student will wait > 25 mins**

$$P(25 < \text{Wait Time} < 30) = (\text{height})(\text{base}) = \frac{1}{(30-0)}(5) = 0.1667$$

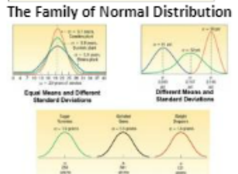
- **Probability the student will wait between 10 and 20 mins**

$$P(10 < \text{Wait Time} < 20) = (\text{height})(\text{base}) = \frac{1}{(30-0)}(10) = 0.3333$$

2. **Normal Probability Distribution**
   Characters : bell-shaped, symmetrical, asymptotic (the curve gets closer to the X-axis but never touched it, mean median mode are equal, total area under the curve is 1.00,
   a. **Graphics**



   b. **The Family of Normal Distribution**



   c. **Normal Distribution**

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

   d. **Standard Normal Distribution**
      mean of 0, standard deviation of 1x. its called z distribution.

$$z = \frac{X-\mu}{\sigma}$$
z = particular observation
μ = mean of the distribution
δ = standard deviation

Example 1 :

The weekly incomes of shift foreman in the glass industry follow the normal probability distribution with a mean of $1,000 and a standard deviation of $100.

What is the z value for the income, let's call it X, of a foreman who earns $1,100 per week? For a foreman who earns $900 per week?

For X = $1,100: $z = \frac{X-\mu}{\sigma} = \frac{\$1,100-\$1,000}{\$100} = 1.00$

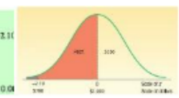
For X = $900: $z = \frac{X-\mu}{\sigma} = \frac{\$900-\$1,000}{\$100} = -1.00$

What is the likelihood of selectir a foreman whose weekly incom is between $1,000 and $1,100?

Example 2 :

What is the probability of selecting a shift foreman in the glass industry whose income is:
Between $790 and $1,000?

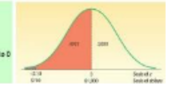
For X = $790: $z = \frac{X-\mu}{\sigma} = \frac{\$790-\$1,000}{\$100} = -2.10$
For X = $1,000: $z = \frac{X-\mu}{\sigma} = \frac{\$1,000-\$1,000}{\$100} = 0.00$



Example 3 :

What is the probability of selecting a shift foreman in the glass industry whose income is:
Less than $790?

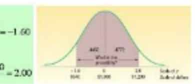
Find Z for X = $790: $z = \frac{X-\mu}{\sigma} = \frac{\$790-\$1,000}{\$100} = -2.10$
To find the area below -2.10, subtract from 0.50 the area from -2.10 to 0
= 0.50-0.4821
= 0.0179



Example 4 :

What is the probability of selecting a shift foreman in the glass industry whose income is:
Between $840 and $1,200?

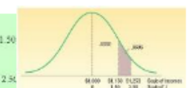
For X = $840: $z = \frac{X-\mu}{\sigma} = \frac{\$840-\$1,000}{\$100} = -1.60$
For X = $1,200: $z = \frac{X-\mu}{\sigma} = \frac{\$1,200-\$1,000}{\$100} = 2.00$



Example 5 :

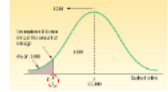
What is the probability of selecting a shift foreman in the glass industry whose income is:
Between $1,150 and $1,250

For X = $1,150: $z = \frac{X-\mu}{\sigma} = \frac{\$1,150-\$1,000}{\$100} = 1.50$
For X = $1,250: $z = \frac{X-\mu}{\sigma} = \frac{\$1,250-\$1,000}{\$100} = 2.50$



Curve area = 0.498 – 0.4332 = 0.0606

Example 6 : Finding X Given Area



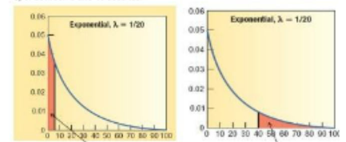
Solve X using the formula:
$z = \frac{x-\mu}{\sigma} = \frac{x-67,900}{2,050}$
$-1.75 = \frac{x-67,900}{2,050}$
$-1.75(2,050) = x - 67,900$
$x = 67,900 - 1.75(2,050)$
$x = 64,312$

3. **Normal Approximation to the Binomial**
   The normal distribution is generally good approximation of the binomial distribution for large values of n (when nπ and n(1-π) are both greater than 5
4. **Correction Error**
   - At least X occurs, use the area above (X-,5)
   - More than X occurs, use the area above (X+.5)
   - X or fewer occurs, use the area below (X-.5)
   - Fewer than X occurs, use the area below (X+.5)
5. **Exponential Probability Distribution**
   Characters : positively skewed, not symmetric, usually describes inter-interval situations
   e.g : the time until the next phone call arrives in CS
   Formula : $P(X) = \lambda e^{-\lambda x}$
   $P(\text{arrival time} < x) = 1-e^{-\lambda x}$

   Example :

   Orders for prescriptions arrive at a pharmacy management website according to an exponential probability distribution at a mean of one every twenty seconds.

   Find the probability the next order arrives in:
   1) in less than 5 seconds,
   2) in more than 40 seconds,
   3) or between 5 and 40 seconds.



$P(\text{Arrival time} < x) = 1 - e^{-\lambda x}$
$P(\text{Arrival} < 5) = 1 - (1 - e^{-\frac{1}{20}5}) = 1 - 0.7788 = 0.2212$

$P(\text{Arrival} > 40) = 1 - P(\text{Arrival} < 40) = 1 - (1 - e^{-\frac{1}{20}40}) = 1 - 0.8647 = 0.1353$

CHAPTER 8:Sampling Method Central Limit Theorem

1. **Most Probability Sampling :**
   - **Simple Random**(sample selected so each item has the same chance of being included)
   - **Systematic Random Sampling** (the items of population are arranged in some order)
   - **Stratified Random Sampling** (population divided in group based on some characteristics)
   - **Cluster** (population is divided into cluster)
2. **Sampling Error** (difference between sample statistic & its corresponding populating parameter)
3. **Sampling Distribution of the Sample Mean** (is a probability consisting of all possible mean)

Example 1 :

Tartus Industries has seven production employees (considered the population). The hourly earnings of each employee are given in the table below.

Employee	Hourly Earnings	Employee	Hourly Earnings
Joe	$7	Jan	$7
Sam	7	Art	8
Sue	8	Ted	9
Bob	8		

1. What is the population mean?
2. What is the sampling distribution of the sample mean for samples of size 2?
3. What is the mean of the sampling distribution?
4. What observations can be made about the population and the sampling distribution?

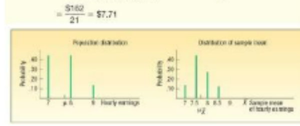
1. The population mean is $7.71, found by:

$$\mu = \frac{\Sigma X}{N} = \frac{\$7+\$7+\$8+\$8+\$7+\$8+\$9}{7} = \$7.71$$

2. To arrive at the sampling distribution of the sample mean, we need to select all possible samples of 2 without replacement from the population, then compute the mean of each sample. There are 21 possible samples.

$$_NC_n = \frac{N!}{n!(N-n)!} = \frac{7!}{2!(7-2)!} = 21$$

$$\mu_{\bar{X}} = \frac{\text{Sum of all sample means}}{\text{Total number of samples}} = \frac{\$7.00+\$7.50+\cdots+\$8.50}{21} = \frac{\$162}{21} = \$7.71$$



4. **Central Limit Theorem** (if all samples are selected from any population, the sampling distribution mean only a normal distribution)
5. **Standard Error of the Mean (Known Sigma)**
   The mean of the distribution sample will exactly equal to population mean

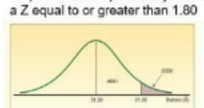
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

The mean amount per bottle is 31.2 ounces and the population standard deviation is 0.4 ounces.
At 8 A.M. today the quality technician randomly selected 16 bottles from the filling line. The mean amount of cola contained in the bottles is 31.38 ounces.
Is this an unlikely result? Is it likely the process is putting too much soda in the bottles? To put it another way, is the sampling error of 0.18 ounces unusual?

Step 1: Find the z-value corresponding to the sample mean of 31.38

$$z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{31.38-31.20}{\$0.4/\sqrt{16}} = 1.80$$

Step 2: Find the probability of observing a Z equal to or greater than 1.80



6. **Standard Error of the Mean (Unknown Sigma)**

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

CHAPTER 9 : Estimation and Confidence Interval

1. **Point Estimate for Population Mean** (sample value that estimates a population parameter)
2. **Confidence Interval for Population Mean**
   C.I = point estimate +- margin of error
3. **Factors Affecting Confidence Interval** (the sample size n, variability in the population usually δ estimated by s, the desired level of confidence)
4. **Obtain Z Value for a Given Confidence Level**
   Confidence 90% = Z value 1.65 = 0.4505
   Confidence 95% = Z value 1.96 = 0.4750
   Confidence 98% = Z value 2.33 = 0.4901
   Confidence 99% = Z value 2.58 = 0.4951
5. **Confidence Intervals Population Mean– δ Known**

$$\bar{X} \pm z\frac{\sigma}{\sqrt{n}}$$

$\bar{x}$ – sample mean
z – z-value for a particular confidence level
σ – the population standard deviation
n – the number of observations in the sample

Example :

The American Management Association wishes to have information on the mean income of middle managers in the retail industry. A random sample of 256 managers reveals a sample mean of $45,420. The standard deviation of this population is $2,050. The association would like answers to the following questions:

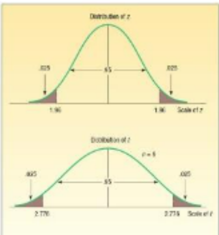
What is a reasonable range of values for the population mean?

Suppose the association decides to use the 95 percent level of confidence:

$$\bar{X} \pm z\frac{\sigma}{\sqrt{n}} = \$45{,}420 \pm 1.96\frac{\$2{,}050}{\sqrt{256}} = \$45{,}420 \pm \$251$$

The confidence limit are $45,169 and $45,671
The ±$251 is referred to as the margin of error

6. **The Interval Estimates – Interpretation**
   If the confidence level is 95%, the interval include the population mean
7. **Confidence Intervals – δ Unknown**
   Characters : continuous distribution, bell-shaped, symmetrical, not one but a family of t distribution, more spread out than Z Value
   Comparing z and t distribution with 95% C.I



Example :

A tire manufacturer wishes to investigate the tread life of its tires. A sample of 10 tires driven 50,000 miles revealed a sample mean of 0.32 inch of tread remaining with a standard deviation of 0.09 inch.

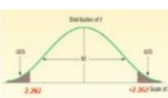
Construct a 95 percent confidence interval for the population mean. Would it be reasonable for the manufacturer to conclude that after 50,000 miles the population mean amount of tread remaining is 0.30 inches?

$n = 10$
$\bar{x} = 0.32$
$s = 0.09$

Compute the C.I. using the t-dist. (since σ is unknown)

$$\bar{X} \pm t_{\alpha/2, n-1}\frac{s}{\sqrt{n}}$$
$$= \bar{X} \pm t_{.05/2, 10-1}\frac{s}{\sqrt{n}}$$
$$= 0.32 \pm t_{.025, 9}\frac{0.09}{\sqrt{10}}$$
$$= 0.32 \pm 2.262\frac{0.09}{\sqrt{10}}$$
$$= 0.32 \pm 0.064$$
$$= (0.256, 0.384)$$

Conclude: the manufacturer can be reasonably sure (95% confident) that the mean remaining tread depth is between 0.256 and 0.384 inches



8. **When to Use z or t Distribution for C.I ?**

Use Z-distribution	Use t-distribution
If the population standard deviation is known or the sample is greater than 30.	If the population standard deviation is unknown and the sample is less than 30.
$\bar{X} \pm z\frac{\sigma}{\sqrt{n}}$	$\bar{X} \pm t\frac{s}{\sqrt{n}}$

9. **Confidence Interval for Proportion**
   a. **Sample Proportion**

$$p = \frac{X}{n}$$

   b. **Population Proportion**

$$p \pm z\sqrt{\frac{p(1-p)}{n}}$$

   c. **Example**

The union representing the Bottle Blowers of America (BBA) is considering a proposal to merge with the Teamsters Union. According to BBA union bylaws, at least three-fourths of the union membership must approve any merger.

A random sample of 2,000 current BBA members reveals 1,600 plan to vote for the merger proposal. What is the estimate of the population proportion?

Develop a 95 percent confidence interval for the population proportion. Basing your decision on this sample information, can you conclude that the necessary proportion of BBA members favor the merger? Why?

First, compute the sample proportion :

$$p = \frac{x}{n} = \frac{1{,}600}{2000} = 0.80$$

Compute the 95% C.I.

$$C.I. = p \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$
$$= 0.80 \pm 1.96\sqrt{\frac{.80(1-.80)}{2{,}000}} = .80 \pm .018$$
$$= (0.782, 0.818)$$

Conclude : The merger proposal will likely pass because the interval estimate includes values greater than 75 percent of the union membership.

10. **Sample Size for Estimating the Population Mean**

$$n = \left(\frac{z \cdot \sigma}{E}\right)^2$$

where:
n is the size of the sample
z is the standard normal value corresponding to the desired level of confidence
σ is the population standard deviation
E is the maximum allowable error

Example :

A student in public administration wants to determine the mean amount members of city councils in large cities earn per month as remuneration for being a council member.

The error in estimating the mean is to be less than $100 with a 95 percent level of confidence. The student found a report by the Department of Labor that estimated the standard deviation to be $1,000. What is the required sample size?

$$n = \left(\frac{z \cdot \sigma}{E}\right)^2 = \left(\frac{(1.96)(\$1{,}000)}{\$100}\right)^2 = (19.6)^2 = 384.16 = 385$$

11. **Sample Size for Estimating Pop. Proportion**

$$n = \pi(1-\pi)\left(\frac{Z}{E}\right)^2$$

where:
n is the size of the sample
z is the standard normal value corresponding to the desired level of confidence
E is the maximum allowable error

Example :

The American Kennel Club wanted to estimate the proportion of children that have a dog as a pet. If the club wanted the estimate to be within 3% of the population proportion, how many children would they need to contact? Assume a 95% level of confidence and that the club estimated that 30% of the children have a dog as a pet.

$$n = \pi(1-\pi)\left(\frac{Z}{E}\right)^2 \qquad n = (.30)(.70)\left(\frac{1.96}{.03}\right)^2 = 897$$

12. **Finite Population Correction Factor**

Standard Error of the Mean

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Standard Error of the Proportion

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$$

13. **C.I For Estimating Means & Proportion with FPC**

C.I. for the Mean (μ)

$$\bar{X} \pm z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

C.I. for the Mean (μ)

$$\bar{X} \pm t\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

C.I. for the Proportion (π)

$$p \pm z\sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$$

Example :

There are 250 families in Scandia, Pennsylvania. A random sample of 40 of these families revealed the mean annual church contribution was $450 and the standard deviation of this was $75.

Could the population mean be $445 or $425?

1. What is the population mean? What is the best estimate of the population mean?
2. Discuss why the finite-population correction factor should be used.

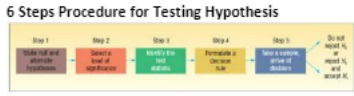
Given in Problem:
N – 250
n – 40
s - $75

$$\bar{X} \pm t\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Since n/N = 40/250 = 0.16, the finite population correction factor must be used.
The population standard deviation is not known therefore use the t-distribution (may use the z-dist since n>30)
Use the formula below to compute the confidence interval:

CHAPTER 10: One Sample Test of Hypothesis

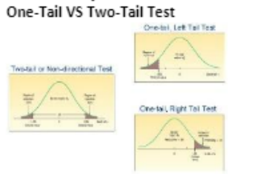
1. **Hypothesis** (a statement about population parameter subject to verification)
   a. **Hypothesis Testing** (a procedure based on sample evidence to determin reasonable statement)
   b. **Null Hypothesis** (statement about the value of population parameter develope for the purpose of testing numerical evidence)
   c. **Alternate Hypothesis** (statement that is accepted if the sample data provid sufficient evidence that the null hypothesis is false)
2. **6 Steps Procedure for Testing Hypothesis**



3. **Important Things about H_0 and H_1**
   a. H_0 : Null Hypothesis, H_1 : Alternate Hypothesis
   b. H_0 and H_1 are mutually exclusive & collectively exhaustive
   c. H_0 is always presumed to be true
   d. H_1 has the burden of proof
   e. Equality is always part of H_0 (e.g : = , ≥ , ≤)
   f. ≠ , < , > always part of H_1
4. **Decisions and Errors in Hypothesis testing**

Null Hypothesis	Researcher: Does Not Reject H_0	Researcher: Rejects H_0
H_0 is true	Correct decision	Type I error
H_0 is false	Type II error	Correct decision

5. **Type of Errors in Hypothesis Testing**
   a. **Type I Error :** Rejecting H_0 when it is true
   b. **Type II Error :** Rejecting H_0 when it is false
6. **Select the Test Statistic :**
   a. **Test Statistic :** used to determine reject H_0
   b. **Critical Value :** point between the region where H0 is rejected and region wher it's not rejected
7. **One-Tail VS Two-Tail Test**



8. **Testing a Mean**

H_0: μ = value H_1: μ ≠ value	H_0: μ ≥ value H_1: μ < value	H_0: μ ≤ value H_1: μ > value
Reject H_0 if: $\lvert Z\rvert > Z_{\alpha/2}$ $\lvert t\rvert > t_{\alpha/2, n-1}$	Reject H_0 if: $Z < -Z_\alpha$ $t < -t_{\alpha, n-1}$	Reject H_0 if: $Z > Z_\alpha$ $t > t_{\alpha, n-1}$

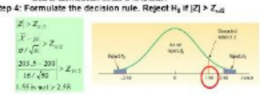
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \qquad z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

   a. **Testing a Population Mean, δ Known**

Example 1:

Jamestown Steel Company manufactures and assembles desks and other office equipment . The weekly production of the Model A325 desk at the Fredonia Plant follows the normal probability distribution with a mean of 200 and a standard deviation of 16. Recently, new production methods have been introduced and new employees hired. The VP of manufacturing would like to investigate whether there has been a change in the weekly production of the Model A325 desk.

Step 1: State the null hypothesis and the alternate hypothesis.
H_0: μ = 200
H_1: μ ≠ 200 (note: keyword in the problem "has changed")
Step 2: Select the level of significance.
α = 0.01 as stated in the problem
Step 3: Select the test statistic.
Use Z-distribution since σ is known
Step 4: Formulate the decision rule. Reject H_0 if $|Z| > Z_{\alpha/2}$



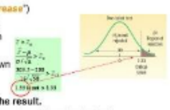
Step 5: Make a decision and interpret the result.
Because 1.55 does not fall in the rejection region, H_0 is not rejected.

Example 2:

Suppose in the previous problem the vice president wants to know whether there has been an increase in the number of units assembled. To put it another way, can we conclude, because of the improved production methods, that the mean number of desks assembled in the last 50 weeks was more than 200?

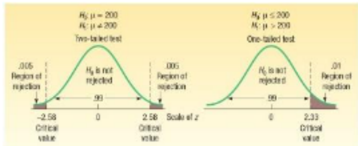
Recall: σ=16, n=200, α=.01

Step 1: State the null hypothesis and the alternate hypothesis.
H_0: μ ≤ 200
H_1: μ > 200
(note: keyword in the problem "an increase")
Step 2: Select the level of significance.
α = 0.01 as stated in the problem
Step 3: Select the test statistic.
Use Z-distribution since σ is known
Step 4: Formulate the decision rule.
Reject H_0 if $Z > Z_\alpha$



Step 5: Make a decision and interpret the result.
Because 1.55 does not fall in the rejection region, H_0 is not rejected.
We conclude that the average number of desks assembled in the last 50 weeks is not more than 200

One-Tailed VS Two-Tailed



   b. **Testing Population Mean – δ Unknown**

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

with $n - 1$ degrees of freedom, where:
$\bar{X}$ is the sample mean.
μ is the hypothesized population mean.
s is the sample standard deviation.
n is the number of observations in the sample.

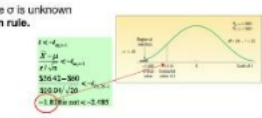
Example :

The McFarland Insurance Company Claims Department reports the mean cost to process a claim is $60. An industry comparison showed this amount to be larger than most other insurance companies, so the company instituted cost-cutting measures.

To evaluate the effect of the cost-cutting measures, the Supervisor of the Claims Department selected a random sample of 26 claims processed last month. The sample information is reported below.
At the .01 significance level is it reasonable a claim is now less than $60?

$45	$49	$62	$40	$43	$61
48	53	67	63	78	64
48	54	51	56	63	69
58	51	58	59	56	57
38	76				

Step 1: State the null hypothesis and the alternate hypothesis.
H_0: μ ≥ $60
H_1: μ < $60
(note: keyword in the problem "now less than")
Step 2: Select the level of significance.
α = 0.01 as stated in the problem
Step 3: Select the test statistic.
Use t-distribution since σ is unknown
Step 4: Formulate the decision rule.
Reject H_0 if $t < -t_{\alpha, n-1}$



Step 5: Make a decision and interpret the result.
Because -1.818 does not fall in the rejection region, H_0 is not rejected at the .01 significance level. We have not demonstrated that the cost-cutting measures reduced the mean cost per claim to less than $60.

t-distribution

	Confidence Level									
	60%	70%	80%	85%	90%	95%	98%	99%	99.8%	99.9%
	Level of Significance									
2 Tailed	0.40	0.30	0.20	0.15	0.10	0.05	0.02	0.01	0.002	0.001
1 Tailed	0.20	0.15	0.10	0.075	0.05	0.025	0.01	0.005	0.001	0.0005
df										
1	1.376	1.963	3.133	4.195	6.320	12.69	31.81	63.67	—	—
2	1.060	1.385	1.883	2.278	2.912	4.271	6.816	9.520	19.65	26.30
3	0.978	1.250	1.637	1.924	2.352	3.179	4.525	5.797	9.937	12.39
4	0.941	1.190	1.533	1.778	2.132	2.776	3.744	4.596	7.115	8.499
5	0.919	1.156	1.476	1.699	2.015	2.570	3.365	4.030	5.876	6.835
6	0.906	1.134	1.440	1.650	1.943	2.447	3.143	3.707	5.201	5.946
7	0.896	1.119	1.415	1.617	1.895	2.365	2.999	3.500	4.783	5.403
8	0.889	1.108	1.397	1.592	1.860	2.306	2.897	3.356	4.500	5.039
9	0.883	1.100	1.383	1.574	1.833	2.262	2.822	3.250	4.297	4.780
10	0.879	1.093	1.372	1.559	1.813	2.228	2.764	3.170	4.144	4.586
11	0.875	1.088	1.363	1.548	1.796	2.201	2.719	3.106	4.025	4.437
12	0.873	1.083	1.356	1.538	1.782	2.179	2.682	3.055	3.930	4.318
13	0.870	1.079	1.350	1.530	1.771	2.160	2.651	3.013	3.852	4.221
14	0.868	1.076	1.345	1.523	1.761	2.145	2.625	2.977	3.788	4.141
15	0.866	1.074	1.341	1.517	1.753	2.131	2.603	2.947	3.733	4.073
16	0.865	1.071	1.337	1.512	1.746	2.120	2.584	2.921	3.687	4.015
17	0.863	1.069	1.333	1.508	1.740	2.110	2.567	2.899	3.646	3.965
18	0.862	1.067	1.330	1.504	1.734	2.101	2.553	2.879	3.611	3.922
19	0.861	1.066	1.328	1.500	1.729	2.093	2.540	2.861	3.580	3.884
20	0.860	1.064	1.325	1.497	1.725	2.086	2.529	2.846	3.552	3.850
21	0.859	1.063	1.323	1.494	1.721	2.080	2.518	2.832	3.528	3.820
22	0.858	1.061	1.321	1.492	1.717	2.074	2.509	2.819	3.505	3.792
23	0.857	1.060	1.319	1.489	1.714	2.069	2.500	2.808	3.485	3.768
24	0.857	1.059	1.318	1.487	1.711	2.064	2.493	2.797	3.467	3.746
25	0.856	1.058	1.316	1.485	1.708	2.060	2.486	2.788	3.451	3.725
26	0.856	1.058	1.315	1.483	1.706	2.056	2.479	2.779	3.435	3.707
27	0.855	1.057	1.314	1.482	1.703	2.052	2.473	2.771	3.421	3.690
28	0.855	1.056	1.313	1.480	1.701	2.048	2.468	2.764	3.409	3.674
29	0.854	1.055	1.311	1.479	1.699	2.045	2.463	2.757	3.397	3.660
30	0.854	1.055	1.310	1.477	1.697	2.042	2.458	2.750	3.386	3.646
40	0.851	1.050	1.303	1.468	1.684	2.021	2.424	2.705	3.307	3.551
50	0.849	1.047	1.299	1.462	1.676	2.009	2.404	2.678	3.262	3.496
60	0.848	1.045	1.296	1.458	1.671	2.000	2.391	2.661	3.232	3.460
70	0.847	1.044	1.294	1.456	1.667	1.994	2.381	2.648	3.211	3.435
80	0.846	1.043	1.292	1.453	1.664	1.990	2.374	2.639	3.196	3.417
90	0.846	1.042	1.291	1.452	1.662	1.987	2.369	2.632	3.184	3.402
100	0.845	1.042	1.290	1.451	1.660	1.984	2.365	2.626	3.174	3.391
∞	0.842	1.036	1.282	1.440	1.645	1.960	2.327	2.576	3.091	3.291

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998