Objective:

So far, we have seen various ways of identifying certain probability distributions with special emphasis on normal distribution, central limit theorem and random sampling. In this session, we shall proceed with statistical way of thinking and solving some problems related to Inferential statistics.

Key takeaways:

a. **Statistical hypothesis**, or simply a hypothesis, is an assumption about a population parameter.
b. **Hypothesis testing** is the procedure whereby we decide to “reject” or “fail to reject” a hypothesis.
c. **Null hypothesis H0**: This is the hypothesis (assumption) under investigation or the statement being tested. The null hypothesis is a statement that “there is no effect,” “there is no difference,” or “there is no change.” The possible outcomes in testing a null hypothesis are ‘reject’ or ‘fail to reject.’
d. **Alternate hypothesis H1**: This is a statement you will adopt if there is strong evidence (sample data) against the null hypothesis. A statistical test is designed to assess the strength of the evidence (data) against the null hypothesis.
e. **Fail to Reject H0**: We never say we “accept H0” - we can only say we “fail to reject” it. Failing to reject H0 means there is NOT enough evidence in the data and in the test to justify rejecting H0. So, we retain the H0 knowing we have not proven it true beyond all doubt.
f. **Rejecting H0:** This means there IS significant evidence in the data and in the test to justify rejecting H0. When H0 is rejected the data is said to be statistically significant. We adopt H1 knowing we will occasionally be wrong.
g. **Margin of error:** Margin of error is the **maximum expected difference** between the true population parameter and a sample estimate of that parameter
h. `Confidence Intervals
i. ***p*-value** is the probability of getting a value up to and including the one in the sample in the direction of the critical region
j. Degrees of freedom
k. t – distribution
l. Remember:

Level of Confidence	Value of z
90%	1.64
95%	1.96
99%	2.58

Activity Sheet:

1. A population of 29 year-old males has a mean salary of $29,321 with a standard deviation of $2,120. If a sample of 100 men is taken, what is the probability their mean salaries will be less than $29,000?
2. A random sample of 100 items is taken, producing a sample mean of 49. The population SD is 4.49. Construct a 90% confidence interval to estimate the population mean.
3. A random sample of 35 items is taken, producing a sample mean of 2.364 with a sample variance of 0.81. Assume x is normally distributed and construct a 90% confidence interval for the population mean.
4. State the null and alternative hypotheses to be used in testing the following claims and determine generally where the critical region is located: (a) The mean snowfall at Lake George during the month of February is 21.8 centimeters. (b) No more than 20% of the faculty at the local university contributed to the annual giving fund.

5. Suppose a car manufacturer claims a model give 25 mpg. A consumer group asks 40 owners of this model to calculate their mpg and the mean value was 22 with a standard deviation of 1.5. Is the manufacturer's claim supported?

6. The CEO of a large electric utility claims that 80 percent of his 1,000,000 customers are very satisfied with the service they receive. To test this claim, the local newspaper surveyed 100 customers, using simple random sampling. Among the sampled customers, 73 percent say they are very satisfied. Based on these findings, can we reject the CEO's hypothesis that 80% of the customers are very satisfied? Use a 0.05 level of significance.

7. A large freight elevator can transport a maximum of 9,800 pounds. Suppose a load of cargo containing 49 boxes must be transported via the elevator. Experience has shown that the weight of boxes of this type of cargo follows a distribution with mean $\mu = 205$ pounds and standard deviation $\sigma = 15$ pounds. Based on this information, what is the probability that all 49 boxes can be safely loaded onto the freight elevator and transported?

8. A marketing director of a large department store wants to estimate the average number of customers who enter the store every five minutes. She randomly selects five-minute intervals and counts the number of arrivals at the store. She obtains the figures 68, 42, 51, 57, 56, 80, 45, 39, 36 and 79. The analyst assumes the number of arrivals is normally distributed. Using

this data, the analyst computes a 95% confidence interval to estimate the mean value for all five-minute intervals. What interval value does she get?

t-value @ 95% confidence, and 9 dof = t (0.05/2,9) = 2.262

9. A student, to test his luck, went to an examination unprepared. It was a MCQ type examination with two choices for each questions. There are 50 questions of which at least 20 are to be answered correctly to pass the test. What is the probability that he clears the exam? If each question has 4 choices instead of two, What is the probability that he clears the exam?

10. Write the Null, and alternate Hypotheses and identify type I and type II errors in the following scenarios:
    a) An innocent person is sent to jail

    b) A manager sees some evidence that stealing is occurring but lacks enough confidence to conclude the theft, and he decides not to fire the employee

11. A car manufacturer advertises a car that gets 47 mpg. Let μ be the mean mileage for this model. You assume that the dealer will not underrate the mileage, but suspect he may overrate the mileage a. What can be used for H0? b. What can be used for H1?

12. A company that manufactures ball bearings claims the average diameter is 6 mm. To check that the average diameter is correct, the company decides to formulate a statistical test. a. What can be used for H0? b. What can be used for H1?

Population Parameter	Population Distribution	Conditions	Confidence Interval
μ	Normal	You know σ^2 n is large or small $\bar{X}$ is the sample mean	$(\bar{X} - z\frac{\sigma}{\sqrt{n}}, \bar{X} + z\frac{\sigma}{\sqrt{n}})$
μ	Non-normal	You know σ^2 n is large (> 30) $\bar{X}$ is the sample mean	$(\bar{X} - z\frac{\sigma}{\sqrt{n}}, \bar{X} + z\frac{\sigma}{\sqrt{n}})$
μ	Normal or Non-normal	You don't know σ^2 n is large (> 30) $\bar{X}$ is the sample mean s^2 is the sample variance	$(\bar{X} - z\frac{s}{\sqrt{n}}, \bar{X} + z\frac{s}{\sqrt{n}})$
p	Binomial	n is large p_s is the sample proportion q_s is $1 - p_s$	$(p_s - z\sqrt{\frac{p_s q_s}{n}}, p_s + z\sqrt{\frac{p_s q_s}{n}})$

	Sigma known	Sigma unknown
n≥30	$\bar{x} \pm z^*\sigma/\sqrt{n}$	$\bar{x} \pm z^*s/\sqrt{n}$
n<30	$\bar{x} \pm z^*\sigma/\sqrt{n}$	$\bar{x} \pm t^*s/\sqrt{n}$