1)What is Estimation Statistics? Explain point estimate and interval estimate.

Ans- Estimation statistics, or simply estimation, is a data analysis framework that uses a combination of effect sizes, confidence intervals, precision planning, and meta-analysis to plan experiments, analyze data and interpret results.

A point estimate is a single value estimate of a parameter. For instance, a sample mean is a point estimate of a population mean. An interval estimate gives you a range of values where the parameter is expected to lie. A confidence interval is the most common type of interval estimate.

2) Write a Python function to estimate the population mean using a sample mean and standard deviation.

Ans- import statistics

data = [7,5,4,9,12,45]

print("Standard Deviation of the sample is % s "% (statistics.stdev(data)))

print("Mean of the sample is % s " % (statistics.mean(data)))

3) What is Hypothesis testing? Why is it used? State the importance of Hypothesis testing.

Ans- Hypothesis testing is used to assess the plausibility of a hypothesis by using sample data. The test provides evidence concerning the plausibility of the hypothesis, given the data. Statistical analysts test a hypothesis by measuring and examining a random sample of the population being analyzed.

Hypotheses are used **to support scientific research and create breakthroughs in knowledge**. These brief statements are what form the basis of entire research experiments. Thus, a flaw in the formulation of a hypothesis may cause a flaw in the design of an entire experiment.

5) Write a Python script to conduct a hypothesis test on the difference between two population means,given a sample from each population.

Ans-

import numpy as np

import scipy.stats as stats

alpha = 0.1

male\_errors = np.array([13, 130, 39, 33, 68, 10, 13, 18, 3, 11, 9, 70, 30, 19, 26, 15, 22, 5, 60, 23, 38, 59, 58, 67, 8, 20, 167, 3, 86, 5])

female\_errors = np.array([14, 138, 3, 78, 27, 8, 3, 111, 69, 8, 20, 36, 109, 80, 3, 18, 31, 12, 15, 128, 35, 35, 27, 122, 68, 32, 111, 91, 176, 66])

# possible types "right-tailed, left-tailed, two-tailed"

tail\_hypothesis\_type = "two-tailed"

if tail\_hypothesis\_type == "two-tailed":

alpha /= 2

# calculate number of sample for both female and male

n1 = len(female\_errors)

n2 = len(male\_errors)

# calculate sample mean

x1\_bar = np.mean(female\_errors)

x2\_bar = np.mean(male\_errors)

# calculate sample standard deviation => ddof = 1, population std => ddof = 0

s1 = np.std(female\_errors, ddof=1)

s2 = np.std(male\_errors, ddof=1)

print("Basic Statistics from sample data")

print("Number of female sample:", n1, "Number of male sample:", n2)

print("Female pointing error mean:", x1\_bar, "Male pointing error mean:", x2\_bar)

print("Female pointing error std:", s1, "Male pointing error std:", s2)

# Test statistics, on upper critical value

if s1 > s2:

F\_score = s1\*\*2 / s2\*\*2

critical\_value = stats.f.ppf(1-alpha, n1-1, n2-1)

else:

F\_score = s2\*\*2 / s1\*\*2

critical\_value = stats.f.ppf(1-alpha, n2-1, n1-1)

print("\n----------------------------TEST STATISTICS---------------------------------------------")

print("F-score is:", F\_score, " and critical value is:", critical\_value)

conclusion = "Failed to reject the null hypothesis."

t\_test\_type = "pooled\_variance t test"

if F\_score > critical\_value:

conclusion = "Null Hypothesis is rejected."

t\_test\_type = "separate-variances t test"

print(conclusion)

6) What is a null and alternative hypothesis? Give some examples.

Ans- Null Hypothesis: On the average, the dosage sold under this brand is 50 mg (population mean dosage = 50 mg). Alternative Hypothesis: On the average, the dosage sold under this brand is not 50 mg (population mean dosage ≠ 50 mg).

**he null hypothesis is the one to be tested and the alternative is everything else**. In our example: The null hypothesis would be: The mean data scientist salary is 113,000 dollars. While the alternative: The mean data scientist salary is not 113,000 dollars.

7) Write down the steps involved in hypothesis testing.

Ans-

Step 1: State your null and alternate hypothesis. ...

Step 2: Collect data. ...

Step 3: Perform a statistical test. ...

Step 4: Decide whether to reject or fail to reject your null hypothesis. ...

Step 5: Present your findings.

8) Define p-value and explain its significance in hypothesis testing.

Ans- The p value is a number, calculated from a statistical test, that describes how likely you are to have found a particular set of observations if the null hypothesis were true. P values are used in hypothesis testing to help decide whether to reject the null hypothesis.

9) Generate a Student's t-distribution plot using Python's matplotlib library, with the degrees of freedom parameter set to 10.

Ans- import numpy as np

import matplotlib.pyplot as plt

distribution = np.linspace(0, np.minimum(rv.dist.b, 3))

print("Distribution : \n", distribution)

plot = plt.plot(distribution, rv.pdf(distribution))

10) Write a Python program to calculate the two-sample t-test for independent samples, given two random samples of equal size and a null hypothesis that the population means are equal.

Ans- import scipy.stats as stats

# Creating data groups

data\_group1 = np.array([14, 15, 15, 16, 13, 8, 14,

17, 16, 14, 19, 20, 21, 15,

15, 16, 16, 13, 14, 12])

data\_group2 = np.array([15, 17, 14, 17, 14, 8, 12,

19, 19, 14, 17, 22, 24, 16,

13, 16, 13, 18, 15, 13])

# Perform the two sample t-test with equal variances

stats.ttest\_ind(a=data\_group1, b=data\_group2, equal\_var=True)

11) What is Student’s t distribution? When to use the t-Distribution.

Ans- The t-distribution, also known as the Student's t-distribution, is a type of probability distribution that is similar to the normal distribution with its bell shape but has heavier tails. It is used for estimating population parameters for small sample sizes or unknown variances.

12) What is t-statistic? State the formula for t-statistic.

Ans- You can calculate a t-value using a common t-test with the formula: t = (X‾ - μ0) / (s / √n), where X‾ is the sample mean, μ0 represents the population mean, s is the standard deviation of the sample and n stands for the size of the sample.

13) A coffee shop owner wants to estimate the average daily revenue for their shop. They take a random sample of 50 days and find the sample mean revenue to be $500 with a standard deviation of $50. Estimate the population mean revenue with a 95% confidence interval.

Ans- 3.73

14) A researcher hypothesizes that a new drug will decrease blood pressure by 10 mmHg. They conduct a clinical trial with 100 patients and find that the sample mean decrease in blood pressure is 8 mmHg with a standard deviation of 3 mmHg. Test the hypothesis with a significance level of 0.05.

Ans-

Accept the null hypothesis.

15) An electronics company produces a certain type of product with a mean weight of 5 pounds and a standard deviation of 0.5 pounds. A random sample of 25 products is taken, and the sample mean weight is found to be 4.8 pounds. Test the hypothesis that the true mean weight of the products is less than 5pounds with a significance level of 0.01.

Ans- Accept the null hypothesis.

16) Two groups of students are given different study materials to prepare for a test. The first group (n1 =

30) has a mean score of 80 with a standard deviation of 10, and the second group (n2 = 40) has a mean

score of 75 with a standard deviation of 8. Test the hypothesis that the population means for the two

groups are equal with a significance level of 0.01.

Ans- Accept the null hypothesis.

17) A marketing company wants to estimate the average number of ads watched by viewers during a TV

program. They take a random sample of 50 viewers and find that the sample mean is 4 with a standard

deviation of 1.5. Estimate the population mean with a 99% confidence interval.

Ans- 5.75