Q1 statistics estimation, in statistics, any of numerous procedures used to calculate the value of some property of a population from observations of a sample drawn from the population.

point estimation, in statistics, the process of finding an approximate value of some parameter—such as the mean (average)—of a population from random samples of the population.

In statistics, interval estimation is the use of sample data to estimate an interval of possible values of a parameter of interest. This is in contrast to point estimation, which gives a single value.

In [3]:

```
#Q2
import numpy as np


# Original array
array = np.arange(10)
print(array)

r1 = np.mean(array)
print("\nMean: ", r1)

r2 = np.std(array)
print("\nstd: ", r2)
```

```
[0 1 2 3 4 5 6 7 8 9]

Mean:  4.5

std:  2.8722813232690143
```

Q3 Hypothesis testing is a statistical method that is used in making a statistical decision using experimental data. Hypothesis testing is basically an assumption that we make about a population parameter. It evaluates two mutually exclusive statements about a population to determine which statement is best supported by the sample data.

Hypothesis testing is an important procedure in statistics. Hypothesis testing evaluates two mutually exclusive population statements to determine which statement is most supported by sample data. When we say that the findings are statistically significant, it is thanks to hypothesis testing.

It helps to assume the probability of research failure and progress. It helps to provide link to the underlying theory and specific research question. It helps in data analysis and measure the validity and reliability of the research. It provides a basis or evidence to prove the validity of the research.

Q4

Average weight of male college students is greater than the average of female students then male students perform well in the exams. But average of Female students in the college greater than the average of the male students in the college then female students perform well.

Q5 Step 1: Determine the hypotheses. Step 2: Collect the data As usual, how we collect the data determines whether we can use it in the inference procedure. We have our usual two requirements for data collection.

Samples must be random to remove or minimize bias. Samples must be representative of the populations in question. Step 3: Assess the evidence. Step 4: State a conclusion. If the P-value ≤ α, we reject the null hypothesis in favor of the alternative hypothesis. If the P-value > α, we fail to reject the null hypothesis. We do not have enough evidence to support the alternative hypothesis. As always, we state our conclusion in context, usually by referring to the alternative hypothesis.

Q6 An alternative hypothesis is an opposing theory to the null hypothesis. For example, if the null hypothesis predicts something to be true, the alternative hypothesis predicts it to be false. The alternative hypothesis often is the statement you test when attempting to disprove the null hypothesis.

null hypothesis is a type of conjecture in statistics that proposes that there is no difference between certain characteristics of a population or data-generating process. The alternative hypothesis proposes that there is a difference

Q7

Hypothesis to Be Tested: Definition and 4 Steps for Testing :- The four steps of hypothesis testing include stating the hypotheses, formulating an analysis plan, analyzing the sample data, and analyzing the result.

Q8

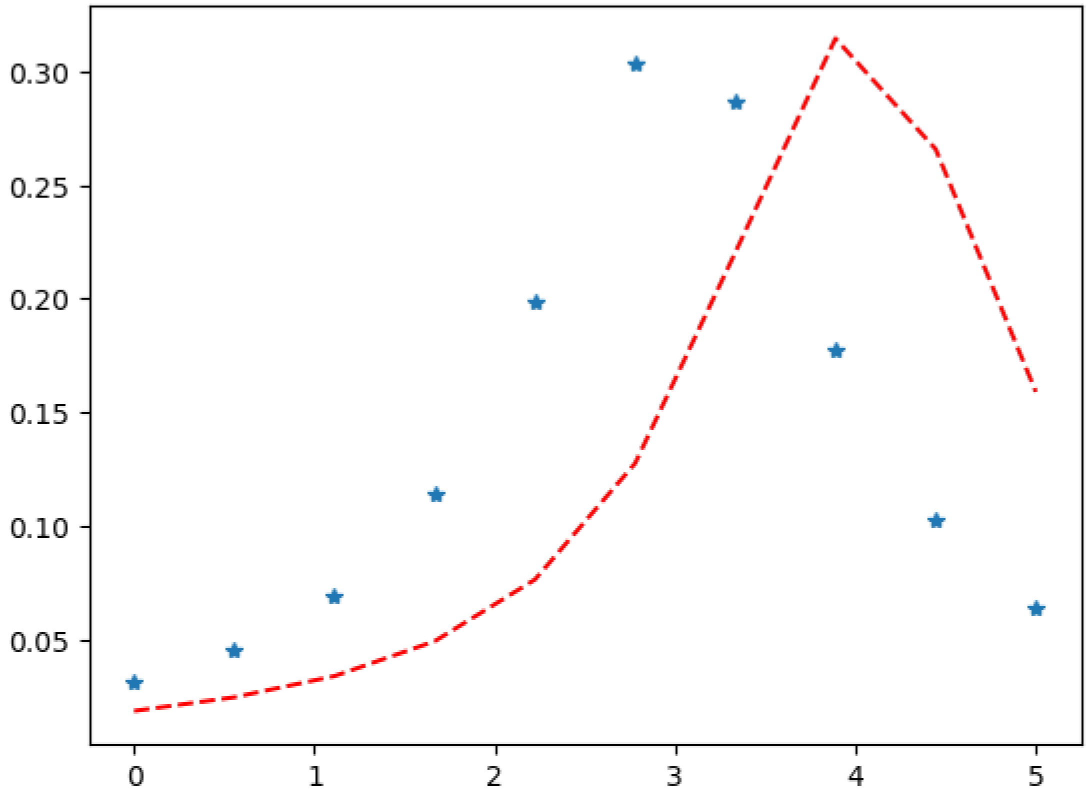
The t-distribution, also known as the Student’s t-distribution, is a probability distribution that is used in inferential statistics when the sample size is small and the population standard deviation is unknown. It is a variation of the normal distribution with heavier tails, which makes it more appropriate for estimating the mean of a population when the sample size is small or when there is uncertainty about the population standard deviation.

In [13]:
```
#Q9
import matplotlib.pyplot as plt
import numpy as np

x = np.linspace(0, 5, 10)

# Varying positional arguments
y1 = t.pdf(x, 1, 3)
```

```
y2 = t.pdf(x, 1, 4)
plt.plot(x, y1, "*", x, y2, "r--")
plt.show()
```



In [16]:

```
#Q10

# Python program to demonstrate how to
# perform two sample T-test

# Import the library
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)
```

Out[16]:

```
Ttest_indResult(statistic=-0.6337397070250238, pvalue=0.5300471010405257)
```

Q11 Student's t-distribution, also known as the t-distribution, is a probability distribution that is used in statistics for making inferences about the population mean when the sample size is small or when the population standard deviation is unknown.

Q12 You can calculate a t-value using a common t-test with the formula: t = (X̄ - μ0) / (s / √n), where X̄ 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

In [ ]:
```
Q13
For a random variable with the standard Normal distribution, Z=dN(0,1)
, we know that Pr(−2<Z<2)≈0.95
. To be more precise:

Pr(−1.96<Z<1.96)=0.95.
```

Q14

Our researcher wants to be correct about their outcome 95% of the time, or the researcher is willing to be incorrect 5% of the time. Probabilities are stated as decimals, with 1.0 being completely positive (100%) and 0 being completely negative (0%). Thus, the researcher who wants to be 95% sure about the outcome of their study is willing to be wrong 5% of the time about the study result. The alpha is the decimal expression of how much they are willing to be wrong. For the current example, the alpha is 0.05. We now have the level of uncertainty the researcher is willing to accept (alpha or significance level) of 0.05 or 5% chance they are not correct about the outcome of the study. Probabilities are stated as decimals, with 1.0 being completely positive (100%) and 0 being completely negative (0%). Thus, the researcher who wants to be 95% sure about the outcome of their study is willing to be wrong 5% of the time about the study result.

Q15 H0 : σ A 2 = σ B 2 H1 : σ A 2 ≠ σ B 2 (two-tailed) Significance level, α = 0.05 Since σˆ A < σˆ B and the larger value is always placed in the numerator,

ν1 = nB −1 = 14 and ν2 = nA −1 = 9 Using interpolation, the upper 2.5% point for

F( ) 14,9 = F( ) 12,9 − 2 3 F( ) 12,9 − F ( ) ( ) 15,9 = 3.868 − 2 3 ( ) 3.868 − 3.769 = 3.802 Thus critical region is F > 3.802 Test statistic is

F = σˆ B 2 σˆ A 2 = 52 32 = 2.78 This value does not lie in the critical region. Thus there is no evidence, at the 5% level of significance, of a difference in t.

In [ ]:
```
Q16
Two groups of students are given different study materials to prepare for a test. T
30) has a mean score of 80 with a standard deviation of 10, and the second group (n
score of 75 with a standard deviation of 8. Test the hypothesis that the population
groups are equal with a significance level of 0.01.
```

Q17 Is the value t = 1.5 statistically significant at the 5% level At the 1% level? t = 1.82 is significant at α = 0.05 but not at α = 0.01. these two values 1.12 is between 1.059 (p = 0.15) and 1.318 (p = 0.10).

For example if = 0.05 and it is an upper tailed test, the critical value is 1.645. For a lower tailed test it is -1.645. But if it is two tailed test then the critical values are -1.96 and 1.96.