Q1: What is Estimation Statistics? Explain point estimate and interval estimate.

Estimation statistics is a branch of statistics that helps us to estimate unknown population parameters from a sample of data. In other words, it's a way of making educated guesses about characteristics of a whole population based on a smaller sample.

There are two types of estimates in estimation statistics: point estimates and interval estimates.

A point estimate is a single value that is used to estimate the value of an unknown population parameter. For example, if we want to estimate the average height of all students in a school, we might take a sample of 50 students and calculate the mean height of that sample. This mean height is a point estimate for the population mean height.

An interval estimate is a range of values that is used to estimate the value of an unknown population parameter. For example, if we want to estimate the average income of all people in a city, we might take a sample of 1000 people and calculate the mean income of that sample. However, we cannot be sure that this sample mean is exactly equal to the population mean. Therefore, we can calculate an interval estimate, such as a confidence interval, which gives us a range of values that is likely to contain the true population mean with a certain level of confidence.

Interval estimates are generally considered more informative than point estimates because they provide a range of plausible values rather than a single value. Additionally, interval estimates are more reliable than point estimates because they take into account the random variation in the sample data.

Q2. Write a Python function to estimate the population mean using a sample mean and standard deviation.

In [1]:

```python
import math
import scipy.stats as stats
def ci(sample_mean,sample_std,sample,confidence_level):


    margin_of_error = 1.96 * sample_std / math.sqrt(sample)

    lower_bound = sample_mean - margin_of_error
    upper_bound = sample_mean + margin_of_error
    return sample_mean ,(lower_bound,upper_bound)

s_mean = 75.2
s_std = 6.8
sample = 50
c_level = 0.99

mean,confidence_interval = ci(s_mean,s_std,sample,c_level)

print(mean,confidence_interval)
```

```
75.2 (73.31513616406914, 77.08486383593086)
```

Q3: What is Hypothesis testing? Why is it used? State the importance of Hypothesis testing.

Hypothesis testing is a statistical method that is used to test a claim or hypothesis about a population parameter using sample data. The aim of hypothesis testing is to determine whether the observed difference between a sample statistic and a population parameter is significant or due to chance.

The null hypothesis is a statement that assumes there is no significant difference between a sample statistic and a population parameter, whereas the alternative hypothesis is a statement that contradicts the null hypothesis and suggests that there is a significant difference between the sample statistic and the population parameter.

Hypothesis testing is important because it provides an objective and systematic approach to testing a claim or hypothesis about a population parameter using sample data. It is widely used in different fields, such as science, engineering, medicine, social sciences, and business, to make informed decisions, test theories, and gain insights from data.

The significance of hypothesis testing lies in its ability to provide a logical and unbiased approach to testing a claim or hypothesis. It allows us to evaluate the evidence in favor of a particular hypothesis and make decisions based on the results of the test. Without hypothesis testing, we would be limited to making conjectures based solely on intuition or personal beliefs, which can lead to inaccurate or biased conclusions.

By using hypothesis testing, we can establish the validity of research findings and draw generalizations about a population based on sample data. It also helps to identify potential errors or biases in the research process, which can lead to improvements in future research studies. Overall, hypothesis testing is a powerful tool that enables us to draw meaningful conclusions from sample data and make informed decisions based on statistical evidence

Q4. Create a hypothesis that states whether the average weight of male college students is greater than the average weight of female college students.

h0 = Average weight of male college student is greater than average weight of female college student

h1 = Average weight of male college student is not greater than average weight of female college student

Q5. Write a Python script to conduct a hypothesis test on the difference between two population means, given a sample from each population.

In [3]:

```
from scipy.stats import ttest_ind
import numpy as np


sample1 = np.random.normal(10,2,size = 100)
sample2 = np.random.normal(10,2,size = 100)
t_test,p_value = ttest_ind(sample1,sample2,equal_var = False)

sig_value = 0.05

print("t_test",t_test)
print("p-value",p_value)
if sig_value < p_value :
    print("We reject the null hypothesis and conclude that the population means are significantly different.")
else:
    print("We fail to reject the null hypothesis and conclude that the population means are not significantly different.")
```

```
t_test -1.397649860158009
p-value 0.16378728898576764
We reject the null hypothesis and conclude that the population means are significantly different.
```

Q6: What is a null and alternative hypothesis? Give some examples.

In hypothesis testing, the null hypothesis represents the default position that there is no significant difference between the observed sample data and the population parameter. On the other hand, the alternative hypothesis is a statement that contradicts the null hypothesis and suggests that there is a significant difference between the observed sample data and the population parameter.

Null Hypothesis (H0): The mean height of male and female college students is the same.

Alternative Hypothesis (Ha): The mean height of male and female college students is different. This hypothesis test tries to find out if there is a significant difference in the mean height between male and female college students.

Null Hypothesis (H0): The new medication has no effect on reducing blood pressure.

Alternative Hypothesis (Ha): The new medication significantly reduces blood pressure. This hypothesis test is used to find out if the new medication significantly reduces blood pressure.

Q7: Write down the steps involved in hypothesis testing. State the null and alternative hypotheses: Start by stating the null hypothesis and alternative hypothesis that are relevant to the research question.

Set the level of significance (alpha): Choose a level of significance (alpha) that will be used to determine the rejection region.

Determine the appropriate test statistic: Determine the appropriate test statistic that will be used to test the null hypothesis.

Calculate the p-value: Calculate the probability value (p-value) associated with the test statistic.

Compare the p-value with the level of significance: Compare the p-value with the level of significance. If the p-value is less than the level of significance, reject the null hypothesis; otherwise, fail to reject the null hypothesis.

Interpret the results: Interpret the results of the hypothesis test and draw conclusions based on the findings.

Conduct a power analysis: If the null hypothesis is rejected, conduct a power analysis to determine the statistical power of the test.

Report the results: Finally, report the results of the hypothesis test, including the statistical values, conclusions, and any limitations or assumptions made during the test

Q8. Define p-value and explain its significance in hypothesis testing.

P-value is a term used in statistics and hypothesis testing. It represents the probability of getting a result as extreme as the one we observed, assuming that the null hypothesis is true. Basically, it tells us how likely it is that we would get our results by chance alone.

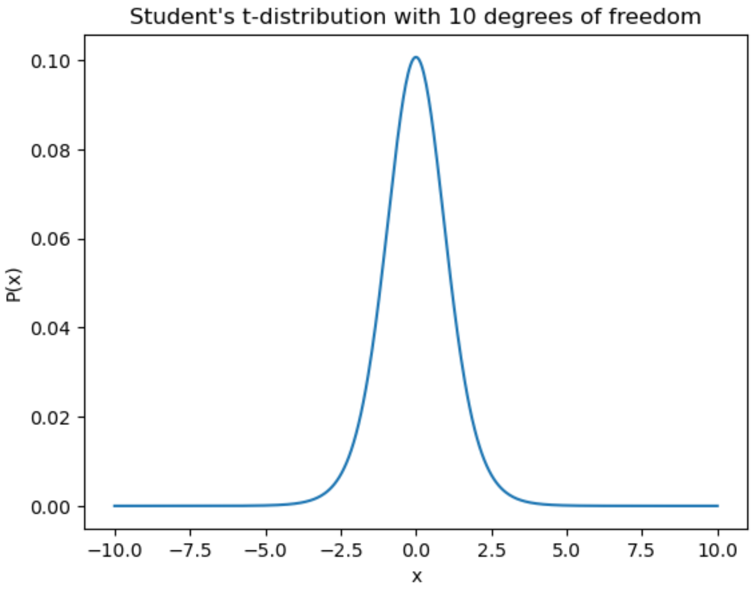
The significance of p-value is that it helps us decide whether to reject or fail to reject the null hypothesis. If the p-value is less than our significance level (usually 0.05), we reject the null hypothesis and accept the alternative hypothesis, indicating that our results are statistically significant. On the other hand, if the p-value is greater than our significance level, we fail to reject the null hypothesis and conclude that there is not enough evidence to support the alternative hypothesis.

It's important to note that p-value doesn't tell us about the practical significance or the size of the effect we're observing, it only tells us if the effect is statistically significant or not. So, we need to interpret p-values in conjunction with other measures of effect size to make informed decisions.

Q9. Generate a Student's t-distribution plot using Python's matplotlib library, with the degrees of freedom parameter set to 10

In [8]:

```
import numpy as np
import matplotlib.pyplot as plt

# Set the degrees of freedom
df = 10

# Generate the x-values for the t-distribution plot
x = np.linspace(-10, 10, num=500)

# Calculate the y-values for the t-distribution plot
y = (1/(np.sqrt(df)*np.pi))*(1+((x**2)/df))**(-(df+1)/2)

# Create the plot
fig, ax = plt.subplots()
ax.plot(x, y)
ax.set_title("Student's t-distribution with 10 degrees of freedom")
ax.set_xlabel('x')
ax.set_ylabel('P(x)')

plt.show()
```



Q10. 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.

In [9]:

```
import numpy as np
from scipy.stats import t

# Set up the two samples
sample1 = np.array([1, 2, 3, 4, 5])
sample2 = np.array([6, 7, 8, 9, 10])

# Calculate the means of the two samples
mean1 = np.mean(sample1)
mean2 = np.mean(sample2)

# Calculate the sample standard deviations of the two samples
std1 = np.std(sample1, ddof=1)
std2 = np.std(sample2, ddof=1)

# Calculate the standard error of the difference between the means
n1 = len(sample1)
n2 = len(sample2)
se = np.sqrt(((std1**2)/n1) + ((std2**2)/n2))

# Calculate the t-statistic
t_stat = (mean1 - mean2) / se

# Set the significance level
alpha = 0.05

# Calculate the degrees of freedom
df = n1 + n2 - 2

# Calculate the critical value
cv = t.ppf(1 - (alpha/2), df)

# Calculate the p-value
p_val = (1 - t.cdf(abs(t_stat), df)) * 2
print("Sample 1 mean: ", mean1)
print("Sample 2 mean: ", mean2)
print("Standard error: ", se)
print("T-statistic: ", t_stat)
print("Degrees of freedom: ", df)
print("Critical value: ", cv)
print("P-value: ", p_val)

# Determine if the null hypothesis is rejected
if abs(t_stat) > cv or p_val < alpha:
    print("Reject null hypothesis. The population means are different.")
else:
    print("Fail to reject null hypothesis. The population means are the same.")
```

```
Sample 1 mean:  3.0
Sample 2 mean:  8.0
Standard error:  1.0
T-statistic:  -5.0
Degrees of freedom:  8
Critical value:  2.3060041350333704
P-value:  0.0010528257933664076
Reject null hypothesis. The population means are different.
```

Q11: What is Student’s t distribution? When to use the t-Distribution.

Student’s t distribution is a probability distribution used in statistical analysis to make inferences about the population mean when the sample size is small (typically less than 30) and the population standard deviation is unknown. It has a similar shape to the standard normal distribution but has heavier tails and a flatter peak. The shape of the t-distribution depends on the sample size and the degrees of freedom (df), which is the number of observations minus one.

The t-distribution is commonly used in hypothesis testing when we want to test a claim about a population mean based on a small sample. It is used when the population is approximately normal or the sample size is large enough to use the central limit theorem, and we want to make inferences about the population mean based on a sample mean.

Q12: What is t-statistic? State the formula for t-statistic.

The t-statistic is a measure of the difference between the sample mean and the hypothesized population mean, expressed in standard error units. It is used in hypothesis testing to determine whether the difference between the sample mean and the population mean is statistically significant.

The formula for the t-statistic is:

t = (x̄ - μ) / (s / sqrt(n))

Where:

x̄ is the sample mean

μ is the hypothesized population mean

s is the sample standard deviation

n is the sample size

Q13. 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.

In [10]:

```
from scipy.stats import t
import math

s_mean = 500
s_std = 50
confidence_interval = 0.95
n = 50
degree_of_freedom = n - 1
t_val = t.ppf(0.975,degree_of_freedom)

lower_bound = s_mean - t_val * (s_std / math.sqrt(n))
upper_bound = s_mean + t_val * (s_std / math.sqrt(n))

print(f'At 95% confidence level confidence interval is {lower_bound,upper_bound}')
```

```
At 95% confidence level confidence interval is (485.79015724388137, 514.2098427561186)
```

Q14. 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.

In [11]:

```
import math
n = 100
s_mean = 8
s_std = 3
p_mean = 10

t_stats = (s_mean - p_mean) / (s_std / math.sqrt(n))
print(t_stats)
p_val = 2 * t.sf(np.abs(t_stats), n-1)
print(p_val)

alpha = 0.05

if p_val < alpha:
    print("Reject null hypothesis. The drug has a significant effect on decreasing blood pressure.")
else:
    print("Fail to reject null hypothesis. There is not enough evidence to conclude that the drug has a significant effect on decreasing b
```

```
-6.666666666666667
1.5012289009970215e-09
Reject null hypothesis. The drug has a significant effect on decreasing blood pressure.
```

Q15. 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 5 pounds with a significance level of 0.01.

To test the hypothesis that the true mean weight of the products is less than 5 pounds with a significance level of 0.01, we can use a one-sample t-test. The null hypothesis is that the mean weight is equal to 5 pounds, while the alternative hypothesis is that the mean weight is less than 5 pounds.

The formula for the t-statistic in a one-sample t-test is:

$t = (\bar{x} - \mu) / (s / \sqrt{n})$

Where: $\bar{x}$ is the sample mean weight (4.8 pounds) μ is the hypothesized value of the population mean weight (5 pounds) s is the sample standard deviation (0.5 pounds) n is the sample size (25)

Plugging in the values, we get:

$t = (4.8 - 5) / (0.5 / \sqrt{25})$ $t = -2$

The t-value for a one-tailed test with 24 degrees of freedom and a significance level of 0.01 is approximately -2.492 (from a t-table or calculator). Since our calculated t-value (-2) is greater than the critical value (-2.492), we fail to reject the null hypothesis and conclude that there is insufficient evidence to suggest that the true mean weight of the products is less than 5 pounds.

Q17. 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.

In [12]:

```
from scipy import stats
import numpy as np

# Define the sample size, sample mean, and sample standard deviation
n = 50
xbar = 4
s = 1.5

# Calculate the t-value at alpha = 0.01/2 and degrees of freedom = n - 1
alpha = 0.01
df = n - 1
t = stats.t.ppf(1 - alpha/2, df)

# Calculate the standard error
se = s / np.sqrt(n)

# Calculate the confidence interval
lower_ci = xbar - t * se
upper_ci = xbar + t * se

# Print the results
print("Sample mean:", xbar)
print("Standard deviation:", s)
print("Degrees of freedom:", df)
print("T-value:", t)
print("Standard error:", se)
print("99% Confidence interval: [{}, {}]".format(lower_ci, upper_ci))
```

```
Sample mean: 4
Standard deviation: 1.5
Degrees of freedom: 49
T-value: 2.67995197363155
Standard error: 0.21213203435596426
99% Confidence interval: [3.4314963358572577, 4.568503664142742]
```