Name: Shreyans Tatiya

Batch: E-2

Roll no: 16010123325

Dept : Computer Engineering

Tutorial 6 - Hypothesis Testing

Q1. A random sample of 200 observations has mean 6.5 cm. Can it be a random sample from a population whose mean is 7 cm and variance is 8.5 cm ?

In [8]:
```
import math
from scipy.stats import norm

pm = 7  # Population mean
sm = 6.5  # Sample mean
variance = 8.5  # Population variance
sd = math.sqrt(variance)  # Population standard deviation
n = 200

# Compute standard error
se = sd / math.sqrt(n)

# Calculate the absolute value of the z-score
zcal = (sm - pm) / se

# Critical value (two-tailed test) at 95% confidence level
z_critical = norm.ppf(0.975)

# Print the results
print("Absolute value of z-calculated is:", zcal)
print("Critical z-value is:", z_critical)

# Hypothesis test conclusion
if abs(zcal) > z_critical:
    print("Reject the null hypothesis: The sample is unlikely from the given population.")
else:
    print("Fail to reject the null hypothesis: The sample could be from the given population.")

print("Name & Roll No, Q.No.: Shreyans Tatiya & 16010123325, Q1")
```

```
Absolute value of z-calculated is: -2.4253562503633295
Critical z-value is: 1.959963984540054
Reject the null hypothesis: The sample is unlikely from the given population.
Name & Roll No, Q.No.: Shreyans Tatiya & 16010123325, Q1
```

Q2. An examination of the weight of 9 apples provided the following data : 150, 152, 149, 151, 148, 152, 150, 151 and 153 grams. Investigate whether the average weight of the apples can be assumed to be 151 grams.

In [9]:
```
import math
from scipy.stats import t

# Sample data
sample_data = [150, 152, 149, 151, 148, 152, 150, 151, 153]
n = len(sample_data)  # Sample size
sample_mean = sum(sample_data) / n  # Sample mean
pop_mean = 151  # Hypothesized population mean

# Compute sample standard deviation
variance = sum((x - sample_mean) ** 2 for x in sample_data) / (n - 1)
sd = math.sqrt(variance)

# Compute t-score
t_score = (sample_mean - pop_mean) / (sd / math.sqrt(n))

# Degrees of freedom
df = n - 1

# Critical t-value for 95% confidence level (two-tailed test)
t_critical = t.ppf(0.975, df)

# Print results
print("Sample Mean:", sample_mean)
print("Sample Standard Deviation:", sd)
print("Computed t-score:", t_score)
print("Critical t-value:", t_critical)

# Hypothesis test conclusion
if abs(t_score) > t_critical:
    print("Reject the null hypothesis: The average weight of apples is significantly different from 151 g.")
else:
    print("Fail to reject the null hypothesis: The average weight of apples can be assumed to be 151 g.")

print("Name & Roll No, Q.No.: Shreyans Tatiya & 16010123325, Q2")
```

```
Sample Mean: 150.66666666666666
Sample Standard Deviation: 1.5811388300841898
Computed t-score: -0.6324555320336938
Critical t-value: 2.3060041350333704
Fail to reject the null hypothesis: The average weight of apples can be assumed to be 151 g.
Name & Roll No, Q.No.: Shreyans Tatiya & 16010123325, Q2
```

Q3. Twenty students participated in a mathematics competition. They were provided with additional tutoring sessions for a month before participating in another similar competition. The scores of each student in both competitions were recorded. Test if the scores provided below indicate that the tutoring sessions had a positive impact on the students' performance. Scores in Competition 1: 85, 78, 72, 90, 93, 65, 79, 81, 70, 75, 87, 69, 82, 74, 86, 88, 91, 73, 77, 84 Scores in Competition 2: 88, 80, 75, 91, 95, 68, 82, 84, 73, 79, 89, 71, 85, 77, 90, 92, 94, 76, 78, 83

In [10]:
```
import math
import numpy as np
from scipy.stats import t

comp1 = np.array([85, 78, 72, 90, 93, 65, 79, 81, 70, 75, 87, 69, 82, 74, 86, 88, 91, 73, 77, 84])
comp2 = np.array([88, 80, 75, 91, 95, 68, 82, 84, 73, 79, 89, 71, 85, 77, 90, 92, 94, 76, 78, 83])

# Compute differences
differences = comp2 - comp1
n = len(differences)

# Compute mean and standard deviation
mean_diff = np.mean(differences)
sample_sd_diff = np.std(differences, ddof=1)  # Sample standard deviation
df = n - 1  # Degrees of freedom

# Compute t-score
t_score = mean_diff / (sample_sd_diff / math.sqrt(n))

# Critical t-value for one-tailed test at 95% confidence level
t_critical = t.ppf(0.95, df)


# Print results
print("Mean difference:", mean_diff)
print("Computed t-score:", t_score)
print("Critical t-value:", t_critical)

if t_score > t_critical:
    print("Reject the null hypothesis: Tutoring sessions had a positive impact on the students' performance.")
else:
    print("Fail to reject the null hypothesis: No significant evidence of improvement.")

print("Name & Roll No, Q.No.: Shreyans Tatiya & 16010123325, Q3")
```

```
Mean difference: 2.55
Computed t-score: 9.575304506946091
Critical t-value: 1.729132811521367
Reject the null hypothesis: Tutoring sessions had a positive impact on the students' performance.
Name & Roll No, Q.No.: Shreyans Tatiya & 16010123325, Q3
```

Q4. Test the significance of the difference between the means of two normal population with the same standard deviation from the following data. Size Mean St. Dev Sample-1 1000 25 5 Sample-2 2000 23 7

In [11]:
```
import math
from scipy.stats import norm

n1 = 1000
mean1 = 25 # Sample mean 1
std1 = 5 # Sd 1

n2 = 2000
mean2 = 23 # Sample mean 2
std2 = 7 # Sd 2

se = math.sqrt((std1**2/n1) + (std2**2/n2))


# Compute z-score for difference of means
zcal = abs((mean1 - mean2) /se)

# Critical value (two-tailed test)
z_critical = norm.ppf(0.975)


print("Absolute value of z-calculated is:", zcal)
print("Critical z-value is:", z_critical)


# Hypothesis test conclusion
if zcal > z_critical:
    print("Reject the null hypothesis: There is a significant difference between the means.")
else:
    print("Fail to reject the null hypothesis: No significant difference between the means.")

print("Name & Roll No, Q.No.: Shreyans Tatiya & 16010123325, Q4")
```

```
Absolute value of z-calculated is: 8.989331499509893
Critical z-value is: 1.959963984540054
Reject the null hypothesis: There is a significant difference between the means.
Name & Roll No, Q.No.: Shreyans Tatiya & 16010123325, Q4
```

Q5. The weights of eight randomly selected athletes are recorded in kilograms: 70, 75, 78, 80, 82,85, 87, 90. The weights of twelve randomly selected basketball players are recorded in kilograms: 72, 74, 76, 78, 79, 80, 82, 83, 84, 85, 87, 88. Can it be concluded that basketball players, on average, weigh more than athletes?

In [12]:
```
import math
from scipy.stats import t

athletes = [70, 75, 78, 80, 82,85, 87, 90]
basketball_players = [72, 74, 76, 78, 79, 80, 82, 83, 84, 85, 87, 88]

# Compute sample statistics
n1 = len(athletes)
n2 = len(basketball_players)
mean1 = sum(athletes) / n1
mean2 = sum(basketball_players) / n2
var1 = sum((x - mean1) ** 2 for x in athletes) / (n1) # Variance of athletes
var2 = sum((x - mean2) ** 2 for x in basketball_players) / (n2) # Variance of basketball players

# Compute pooled standard deviation
sp = math.sqrt(((n1) * var1 + (n2) * var2) / (n1 + n2 - 2))

# Compute t-score
t_score = (mean2 - mean1) / (sp * math.sqrt(1/n1 + 1/n2))

# Degrees of freedom
df = n1 + n2 - 2

# Critical t-value for one-tailed test at 95% confidence level
t_critical = t.ppf(0.95, df)


print("Mean weight of athletes:", mean1)
print("Mean weight of basketball players:", mean2)
print("Computed t-score:", t_score)
print("Critical t-value:", t_critical)

# Hypothesis test conclusion
if t_score > t_critical:
    print("Reject the null hypothesis: Basketball players weigh significantly more on average than athletes.")
else:
    print("Fail to reject the null hypothesis: There is no significant evidence that basketball players weigh more than athletes.")

print("Name & Roll No, Q.No.: Shreyans Tatiya & 16010123325, Q5")
```

```
Mean weight of athletes: 80.875
Mean weight of basketball players: 80.66666666666667
Computed t-score: -0.08014605884111159
Critical t-value: 1.7340636066175354
Fail to reject the null hypothesis: There is no significant evidence that basketball players weigh more than athletes.
Name & Roll No, Q.No.: Shreyans Tatiya & 16010123325, Q5
```

In [ ]: