Tutorial 6 - Hypothesis Testing

Name - Aaryan Sharma

Roll No. - 16010123012

Batch - A1

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 [2]:
```python
import math
from scipy.stats import norm

print("Q.No.: 1")
print("Aaryan Sharma - 16010123012")

# Given data
pm = 7
sm = 6.5
variance = 8.5
sd = math.sqrt(variance)
n = 200

# Calculate the absolute value of the z-score
zcal = abs((sm - pm) / (sd / math.sqrt(n)))

# Critical value for 95% confidence level (two-tailed test)
z_critical = norm.ppf(0.975)  # 1.96 for 95% confidence

# Print the results
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: The sample is unlikely from the given population.")
else:
    print("Fail to reject the null hypothesis: The sample could be from the given population.")
```

```
Q.No.: 1
Aaryan Sharma - 16010123012
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.
```

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 [4]:
```python
import math
from scipy.stats import t

print("Q.No.: 2")
print("Aaryan Sharma - 16010123012")

# Given data
sample_data = [150, 152, 149, 151, 148, 152, 150, 151, 153]
n = len(sample_data)
sample_mean = sum(sample_data) / n
pop_mean = 151

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

# Compute t-score
t_score = (sample_mean - pop_mean) / (sd / math.sqrt(n-1))

# Critical t-value for 95% confidence level (two-tailed test)
df = n - 1
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 is significantly different from 151 grams.")
else:
    print("Fail to reject the null hypothesis: The average weight can be assumed to be 151 grams.")
```

```
Q.No.: 2
Aaryan Sharma - 16010123012
Sample Mean: 150.66666666666666
Sample Standard Deviation: 1.4907119849998598
Computed t-score: -0.632455532033694
Critical t-value: 2.306004135204166
Fail to reject the null hypothesis: The average weight can be assumed to be 151 grams.
```

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 [6]:
```python
import math
from scipy.stats import t

print("Q.No.: 3")
print("Aaryan Sharma - 16010123012")

# Scores before and after tutoring
scores_before = [85, 78, 72, 90, 93, 65, 79, 81, 70, 75, 87, 69, 82, 74, 86, 88, 91, 73, 77, 84]
scores_after = [88, 80, 75, 91, 95, 68, 82, 84, 73, 79, 89, 71, 85, 77, 90, 92, 94, 76, 78, 83]

# Calculate differences
differences = [scores_after[i] - scores_before[i] for i in range(len(scores_before))]
n = len(differences)

# Calculate mean and standard deviation of differences
mean_diff = sum(differences) / n
var_diff = sum((x - mean_diff) ** 2 for x in differences) / (n - 1)
sd_diff = math.sqrt(var_diff)

# Compute t-score for paired data
t_score = mean_diff / (sd_diff / math.sqrt(n))

# Degrees of freedom
df = n - 1

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

# Print results
print("Mean difference:", mean_diff)
print("Sample standard deviation of differences:", sd_diff)
print("Computed t-score:", t_score)
print("Critical t-value (one-tailed, 95%):", t_critical)

# Hypothesis test conclusion
if t_score > t_critical:
    print("Reject the null hypothesis: Tutoring sessions significantly improved students' performance.")
else:
    print("Fail to reject the null hypothesis: No significant evidence that tutoring improved performance.")
```

```
Q.No.: 3
Aaryan Sharma - 16010123012
Mean difference: 2.55
Sample standard deviation of differences: 1.190974832912761
Computed t-score: 9.575304506946091
Critical t-value (one-tailed, 95%): 1.729132811521367
Reject the null hypothesis: Tutoring sessions significantly improved students' performance.
```

Q4. Test the Significance of the Difference Between Means

Test the significance of the difference between the means of two normal populations with the same standard deviation from the following data:

Sample	Size (n)	Mean ($\bar{x}$)	Standard Deviation (σ)
Sample-1	1000	25	5
Sample-2	2000	23	7

In [7]:
```python
import math
from scipy.stats import norm

print("Q.No.: 4")
print("Aaryan Sharma - 16010123012")

# Given data for two samples
n1 = 1000
mean1 = 25
std1 = 5

n2 = 2000
mean2 = 23
std2 = 7

# Since the populations have the same standard deviation, we use S.E.=√((s_1^2)/n_2 +(s_2^2)/n_1 )
se = math.sqrt(( std1**2/n2) + (std2**2/n1))

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

# Critical value for 95% confidence level (two-tailed test)
z_critical = norm.ppf(0.975)  # 1.96 for 95% confidence

# Print the results
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.")
```

```
Q.No.: 4
Aaryan Sharma - 16010123012
Absolute value of z-calculated is: 8.064778385455119
Critical z-value is: 1.959963984540054
Reject the null hypothesis: There is a significant difference between the means.
```

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 [8]:
```python
import math
from scipy.stats import t

print("Q.No.: 5")
print("Aaryan Sharma - 16010123012")

# Weights of athletes and basketball players
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

# Sample variances (corrected by dividing by n-1)
var1 = sum((x - mean1) ** 2 for x in athletes) / (n1 - 1)
var2 = sum((x - mean2) ** 2 for x in basketball_players) / (n2 - 1)

# Pooled standard deviation calculation
sp = math.sqrt(((n1 - 1)*var1 + (n2 - 1)*var2) / (n1 + n2 - 2))

# Compute t-score for difference in means
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 (95% confidence)
t_critical = t.ppf(0.95, df)

# Print results
print("Mean weight of athletes:", mean1)
print("Mean weight of basketball players:", mean2)
print("Computed t-score:", t_score)
print("Critical t-value (one-tailed, 95%):", t_critical)

# Hypothesis test conclusion
if t_score > t_critical:
    print("Reject the null hypothesis: Basketball players weigh significantly more than athletes.")
else:
    print("Fail to reject the null hypothesis: No significant evidence that basketball players weigh more than athlete
s.")
```

```
Q.No.: 5
Aaryan Sharma - 16010123012
Mean weight of athletes: 80.875
Mean weight of basketball players: 80.66666666666667
Computed t-score: -0.0801640588411159
Critical t-value (one-tailed, 95%): 1.7340636066175354
Fail to reject the null hypothesis: No significant evidence that basketball players weigh more than athletes.
```