In [1]:
```
import scipy.stats as stats
import pandas as pd
```

Q1. A company claims that their new battery lasts longer than their previous model, which has a mean lifespan of 100 hours in previous model. A sample of 25 new batteries is tested, yielding a mean lifespan of 105 hours with a standard deviation of 10 hours. At a 5% significance level, can we conclude that the new battery lasts longer? Formulate the null and alternative hypotheses, choose an appropriate test, and conduct the test to reach a conclusion.

Null Hypothesis: New batteries lifespan is not longer than previous model

Alternate Hypothesis: New batteries lifespan is longer than previous models

In [3]:
```
pmean=100
smean=105
std=10
size=25
significane_value=0.05
```

In [4]:
```
t_statistic=(smean-pmean)/(std/(size**0.5))
c_i=stats.t.ppf(1-significane_value,size-1)
p_value=1-stats.t.cdf(t_statistic, size-1)
print(f'T-statistic: {t_statistic}')
print(f'Critical T-value: {c_i}')
print(f'P-value: {p_value}')
if t_statistic>c_i:
    print('Reject the Null Hypothesis, New batteries lifespan is longer than previous models')
else:
    print('Fail to reject the Null Hypothesis, New batteries lifespan is not longer than previous models')
```

```
T-statistic: 2.5
Critical T-value: 1.7108820799094275
P-value: 0.009827087558289316
Reject the Null Hypothesis, New batteries lifespan is longer than previous models
```

Q2. A researcher wants to estimate the average time spent on social media by high school students. A sample of 50 students is selected, and the mean time spent is found to be 3.5 hours with a standard deviation of 0.8 hours. Construct a 95% confidence interval for the true mean time spent on social media by all high school students. Calculate the confidence interval and interpret the results.

In [5]:
```
smean=3.5
sample=50
std=0.8
confidence=0.95
dof=sample-1
alpha=1-confidence
critical_value=stats.t.ppf(1-alpha/2,dof)
marginoferror=critical_value*(std/(sample**0.5))
confidence_interval=(smean-marginoferror,smean+marginoferror)
print(f"Critical Value (t-score): {critical_value:.2f}")
print(f"Margin of Error: {marginoferror:.2f}")
print(f'95% sure that average time spent on social media is between: {confidence_interval}')
```

```
Critical Value (t-score): 2.01
Margin of Error: 0.23
95% sure that average time spent on social media is between: (3.2726425156034162, 3.7273574843965838)
```

Q3. A survey is conducted to examine whether there is an association between gender (male/female) and preference for a new product (like/dislike). The data is summarized in the table below. Conduct a Chi-square test of independence to determine if there is a significant association between gender and product preference. State the null and alternative hypotheses, calculate the Chi-square statistic, and interpret the results. Like Dislike Male 20 30 Female 25 25

Null Hypothesis: There is no association between gender and preference for a new product.

Alternate Hypothesis: There is association between gender and preference for a new product.

In [7]:
```
data=[[20,30],[25,25]]
statistic,p,dof,expected=stats.chi2_contingency(data)
print("statistic:", statistic)
print("P-value:", p)
print("Degrees of freedom:", dof)
print("Expected frequencies:\n", expected)
if p<0.05:
    print("We Reject the Null Hypothesis, there is association between gender and preference for a new product")
else:
    print("We fail to reject the Null Hypothesis, there is no association between gender and preference for a new product")
```

```
statistic: 0.6464646464646464
P-value: 0.4213795037428696
Degrees of freedom: 1
Expected frequencies:
 [[22.5 27.5]
 [22.5 27.5]]
We fail to reject the Null Hypothesis, there is no association between gender and preference for a new product
```

Q4. A group of students takes a math test before and after attending a 4-week tutoring session. The test scores before and after the session are recorded. Conduct a appropriate test to determine if the tutoring session has significantly improved the students' test scores. Formulate the null and alternative hypotheses, choose an appropriate test, and conduct the test to reach a conclusion.

Student Before After

1 60 75

2 72 80

3 63 78

4 80 82

5 69 76

Null Hypothesis: Tutoring session improve the score

Alternate Hypothesis: Tutoring session has not impact on the scores

In [8]:
```
data=pd.DataFrame({'Student':[1,2,3,4,5], 'Before':[60,72,63,80,69],'After':[75,80,78,82,76]})
data
```

Out[8]:

	Student	Before	After
0	1	60	75
1	2	72	80
2	3	63	78
3	4	80	82
4	5	69	76

In [9]:
```
ttest,p_value =stats.ttest_rel(data['Before'],data['After'])
print(f"t-statistic: {ttest:.4f}")
print(f"P-value: {p_value:.4f}")
if p_value<0.05:
    print("We Reject the Null Hypothesis, tutoring session has not impact on the scores")
else:
    print("We fail to reject the Null Hypothesis, tutoring session improve the score")
```

```
t-statistic: -3.7570
P-value: 0.0198
We Reject the Null Hypothesis, tutoring session has not impact on the scores
```

Q5. A botanist is studying the effect of three different fertilizers on the growth of a particular plant species. The plants are divided into three groups, each receiving a different type of fertilizer (Fertilizer A, Fertilizer B, Fertilizer C). After 8 weeks, the height of the plants in each group is measured in centimeters. Perform a ANOVA to determine if there is a significant difference in the average plant height between the three fertilizers. Report the F-statistic, p-value, and your conclusion.

Fertilizer Fertilizer A Fertilizer B Fertilizer C

Plant 1 24 30 22

Plant 2 26 32 20

Plant 3 28 29 23

Plant 4 22 31 21

Plant 5 25 33 24

Null Hypothesis: There is no significant effect of Fertilizers on the growth.

Aleternate Hypothesis: There is significant effect of Fertilizers on the growth.

In [10]:
```
data= pd.DataFrame({'FertilizerA':[24,26,28,22,25], 'FertilizerB':[30,32,29,31,33], 'FertilizerC':[22,20,23,21,24]},index=['Pl
data
```

Out[10]:

	FertilizerA	FertilizerB	FertilizerC
Plant1	24	30	22
Plant2	26	32	20
Plant3	28	29	23
Plant4	22	31	21
Plant5	25	33	24

In [11]:
```
f_oneway=stats.f_oneway
fstatistics,p_value=f_oneway(data['FertilizerA'], data['FertilizerB'], data['FertilizerC'])
print('F_statistics', fstatistics)
print('p_value',p_value)
if p_value<0.05:
    print("Reject the Null Hypothesis, There is significant effect of Fertilizers on the growth")
else:
    print("Fail to rejct the Null Hypothesis, There is no significant effect of Fertilizers on the growth")
```

```
F_statistics 31.5
p_value 1.6777216000000003e-05
Reject the Null Hypothesis, There is significant effect of Fertilizers on the growth
```

Q6. You have a bag containing 4 red balls, 3 blue balls, and 3 green balls. You randomly draw two balls from the bag, without replacement.

Task: a) What is the probability that both balls are red? b) What is the probability that the first ball is red and the second ball is blue? c) If you know the first ball drawn was red, what is the probability that the second ball drawn is also red? Use conditional probability.

In [12]:
```
Total_balls=10
Red_balls=4
Blue_balls=3
Green_balls=3
```

In [13]:
```
#a) Both balls are red
first_ball_is_red= Red_balls/Total_balls
second_ball_is_red=(Red_balls-1)/(Total_balls-1)
Both_balls_is_red=first_ball_is_red*second_ball_is_red
print('Prob of both balls is red is',Both_balls_is_red)
```

```
Prob of both balls is red is 0.13333333333333333
```

In [14]:
```
#b) First ball is red and second ball is blue
first_ball_is_red= Red_balls/Total_balls
second_ball_is_blue=Blue_balls/(Total_balls-1)
first_is_red_second_is_blue=first_ball_is_red*second_ball_is_blue
print('Prob of first is red and second is blue is',first_is_red_second_is_blue)
```

```
Prob of first is red and second is blue is 0.13333333333333333
```

In [15]:
```
#c) If you know the first ball drawn was red, what is the probability that the second ball drawn is also red?
second_ball_is_red=(Red_balls-1)/(Total_balls-1)
print('Porb of second ball is red is',second_ball_is_red)
```

```
Porb of second ball is red is 0.3333333333333333
```

Q7. Explain Bayes theorem in detail with some example.

Bayes' theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event. It's a way to update our beliefs about something given new evidence.

Formula: P(A|B) = [P(B|A) * P(A)] / P(B)

Where:

- P(A|B): Posterior probability – the probability of event A happening, *given* that event B has already occurred. This is what we want to find.
- P(B|A): Likelihood – the probability of event B happening, *given* that event A has already occurred.
- P(A): Prior probability – the initial probability of event A happening *before* considering any new evidence (B).

- P(B): Evidence probability – the probability of event B happening *regardless* of A.

For Example lets take this question.

Suppose a weather forecast predicts a 30% chance of rain tomorrow (P(Rain) = 0.30). You also know that the weather forecast is 80% accurate when it predicts rain (P(Prediction|Rain) = 0.80), and 70% accurate when it predicts no rain (P(Prediction|No Rain) = 0.70). You wake up and see the weather forecast predicting rain. What is the probability that it will actually rain given this prediction?

In [16]:
```
Prob_Rain = 0.30
Prob_No_Rain = 0.70
Prob_Prediction_given_Rain = 0.80
Prob_Prediction_given_No_Rain = 0.30

#total probability of the prediction of rain
Prob_Prediction = (Prob_Prediction_given_Rain * Prob_Rain) + (Prob_Prediction_given_No_Rain * Prob_No_Rain)

#posterior probability of rain given the prediction
Prob_Rain_given_Prediction = (Prob_Prediction_given_Rain * Prob_Rain) / Prob_Prediction

# Output the results
print(f'Total probability of the prediction (P(Pred)): {Prob_Prediction:.4f}')
print(f'Probability that it will rain given the prediction (P(Rain|Pred)): {Prob_Rain_given_Prediction:.4f}')
```

```
Total probability of the prediction (P(Pred)): 0.4500
Probability that it will rain given the prediction (P(Rain|Pred)): 0.5333
```