HYPOTHESIS TESTING

Hypothesis testing is used to make inferences about a population parameter based on sample data, evaluating whether observed results are statistically significant or occurred by chance.

Step-by-Step Guide to Hypothesis Testing

Step 1: Formulate Hypotheses

- **Null Hypothesis (H_0)**: The new drug has no effect on blood pressure.
  - $H_0 : \mu_{\text{after}} - \mu_{\text{before}} = 0$
- **Alternate Hypothesis (H_1)**: The new drug has an effect on blood pressure.
  - $H_1 : \mu_{\text{after}} - \mu_{\text{before}} \neq 0$ (two-tailed test)

The null hypothesis

is a statement of no effect, no difference, or no relationship between variables or conditions. It represents the status quo or the default position that there is no change, no effect, or no association in the population parameter being studied.

Typically, the null hypothesis is represented as:

$$H_0 : \theta = \theta_0$$

where θ is the population parameter of interest (e.g., population mean, population proportion) and θ_0 is a specific hypothesized value (often based on theoretical expectations or previous research).

Alternative Hypothesis (

Definition: The alternative hypothesis is the hypothesis that the researcher wants to test. It typically states that there is an effect, a difference, or a relationship between variables or conditions in the population, which is not due to random chance.

Symbolic Representation: The alternative hypothesis can take several forms depending on the nature of the research question:

- For a two-tailed test:

  $$H_1 : \theta \neq \theta_0$$

  This indicates that the population parameter θ is not equal to the hypothesized value θ_0, suggesting there is some effect, difference, or relationship present.

- For a one-tailed test (either direction):

  $$H_1 : \theta > \theta_0$$

  or

In [1]:
```
import numpy as np
from scipy import stats #lib--module--function

# Data
before_treatment = np.array([120, 122, 118, 130, 125, 128, 115, 121, 123, 119])
after_treatment = np.array([110, 100, 112, 100, 122, 125, 110, 117, 119, 113])

# Step 1: Null and Alternate Hypotheses
# Null Hypothesis: The new drug has no effect on blood pressure.
# Alternate Hypothesis: The new drug has an effect on blood pressure.
alternate_hypothesis = "The new drug has an effect on blood pressure."
```

step 2: test

z test, f test, t test, chi squre............

sample size=10,<30== t test

Use a paired T-test when comparing two related groups (e.g., before and after treatment measurements on the same subjects). Use an independent T-test when comparing two independent groups (e.g., measurements from different subjects).

Use the paired T-test to compare the means of paired samples (before and after treatment).

population===parameter

sample=== statistic

In [2]:
```
# Paired T-test
t_statistic, p_value = stats.ttest_rel(before_treatment,after_treatment)
t_statistic
```

Out[2]:
```
3.1875827806998402
```

In [3]:
```python
#p_value

#  P stands for probability and measures how likely that  any observed difference between groups is due to cha

# A lower p-value suggests stronger evidence against the null hypothesis.
p_value
```

Out[3]:
```
0.01104937716788195
```

step 3: significance level

Significance Level (α)

Typically set at 0.05 (5%), representing the probability of rejecting the null hypothesis when it is actually true.

confidence= 99, 95, 90

significane= 1, 5, 10

In many fields, certain significance levels are commonly used.

✅**For example, 5% (α = 0.05) is widely accepted in many scientific disciplines.**

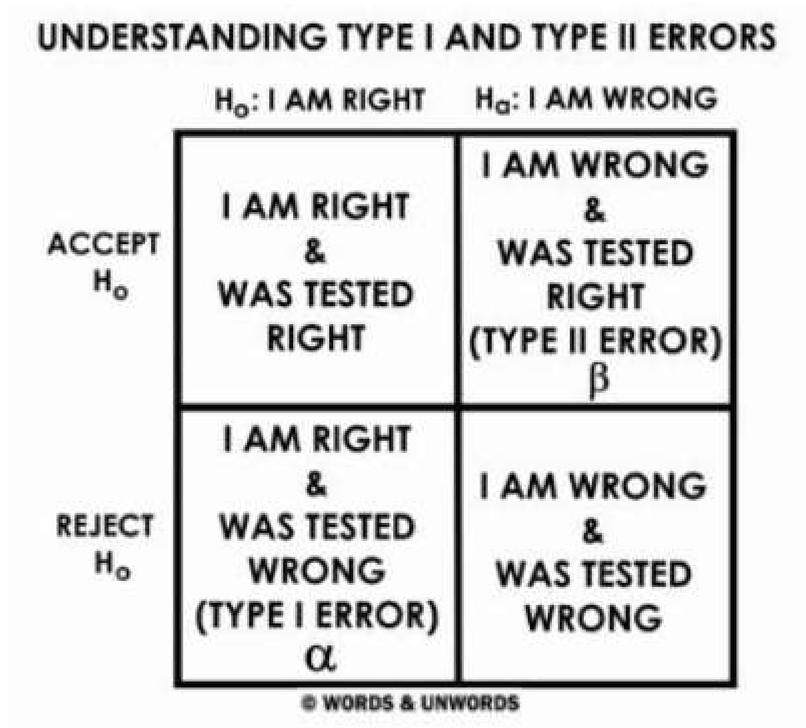
⭐**A lower significance level (e.g., 1% or 0.1%) reduces the likelihood of Type I errors but increases the chance of Type II errors.**

✨ **Conversely, a higher significance level (e.g., 10%) decreases the chance of Type II errors but increases the chance of Type I errors.**

What is a Type I error and Type II error in hypothesis testing?

Answer:

☝Type I error: Rejecting the null hypothesis when it is actually true.



In [4]:
```
0.028935374046502243<.05
```

Out[4]:
```
True
```

In [5]:
```

# Decision
significance = 0.05 #5% signifnace
if p_value < 0.05:
    decision = "Reject null hypo " #  reject no differnce----their is a differnce()
else:
    decision = "accept null "
```

In [6]:
```
# conclusion
5/100
```

Out[6]:
```
0.05
```

In [7]:
```
# Conclusion
if decision == "Reject":
    conclusion = "There is statistically significant evidence that the average blood pressure before and afte
else:
    conclusion = "There is insufficient evidence to claim a significant difference in average blood pressure

# Display results
print("T-statistic (from scipy):", t_statistic)
print("P-value (from scipy):", p_value)
print("Decision:",decision, "the null hypothesis at alpha=",significance)
print("Conclusion:", conclusion)
```

```
T-statistic (from scipy): 3.1875827806998402
P-value (from scipy): 0.01104937716788195
Decision: Reject null hypo  the null hypothesis at alpha= 0.05
Conclusion: There is insufficient evidence to claim a significant difference in average blood pressure befor
e and after treatment with the new drug.
```

SIMPLY

In [8]:

```python
import numpy as np
from scipy import stats

# Data
before_treatment = np.array([120, 122, 118, 130, 125, 128, 115, 121, 123, 119])
after_treatment = np.array([110, 100, 112, 100, 122, 125, 110, 117, 119, 113])

# Step 1: Null and Alternate Hypotheses
# Null Hypothesis: The new drug has no effect on blood pressure.
# Alternate Hypothesis: The new drug has an effect on blood pressure.
null_hypothesis = "The new drug has no effect on blood pressure."
alternate_hypothesis = "The new drug has an effect on blood pressure."

# Step 2: Significance Level
alpha = 0.05
# Step 3: Paired T-test
t_statistic, p_value = stats.ttest_rel(after_treatment, before_treatment)


# Step 5: Decision
if p_value <= alpha:
    decision = "Reject null "
else:
    decision = "accept null"

# Conclusion
if decision == "Reject":
    conclusion = "There is statistically significant evidence that the average blood pressure before and afte
else:
    conclusion = "There is insufficient evidence to claim a significant difference in average blood pressure

# Display results
print("T-statistic (from scipy):", t_statistic)
print("P-value (from scipy):", p_value)
print(f"Decision: {decision} the null hypothesis at alpha={alpha}.")
print("Conclusion:", conclusion)
```

```
T-statistic (from scipy): -3.1875827806998402
P-value (from scipy): 0.01104937716788195
Decision: Reject null  the null hypothesis at alpha=0.05.
Conclusion: There is insufficient evidence to claim a significant difference in average blood pressure befor
e and after treatment with the new drug.
```

Q 2 - LEFT TAILED

Suppose a company claims that the mean weight of their product is at least 500 grams. You suspect that the actual mean weight is less than 500 grams and you want to test this claim using a one-sided t-test. You collect a sample of 10 products with the following weights (in grams):

[495 , 490 , 505 , 500 , 499 , 498 , 497 , 496 , 502 , 494]

You want to test this claim at the 0.05 significance level.

In [9]:
```
import numpy as np
from scipy import stats

# Sample data
data = [495, 490, 505, 500, 499, 498, 497, 496, 502,500]

# Hypothesized population mean
mu_0 = 500

#null= no differnce, equal
null="their is no differnce b/w actal mean=calculated mean"


alternative="actual mean weight is less than 500 grams"  # < ,left
#alternative="actual mean weight is greater than 500 grams" # right
#alternative="actual mean weight is not equal to than 500 grams" # 2 tailed
```

In [10]:
```
t,p_val = stats.ttest_1samp(data,500,alternative="less") #alternative='two-sided',less,greater
t
```

Out[10]:
```
-1.386896841731297
```

In [11]:
```
p_val
```

Out[11]:
```
0.0994287292511126
```

In [12]:
```
if p_val <= .05:
    print( "Reject null")
else:
    print("Fail to reject null") # accpet null

#0.15020834437485206>0.05
```

```
Fail to reject null
```

IN GENERAL

In [13]:

```python
import numpy as np
from scipy import stats

# Sample data
data = [495, 490, 505, 500, 499, 498, 497, 496, 502, 300]

# Hypothesized population mean
mu_0 = 500

# Significance level
alpha = 0.05

# Perform one-sided (left-tailed) t-test
t_statistic, p_value = stats.ttest_1samp(data, mu_0, alternative='less')

# Calculate degrees of freedom
df = len(data) - 1

# Calculate the critical t value for a left-tailed test
t_critical = stats.t.ppf(alpha, df)

# Print results
print(f"t Statistic: {t_statistic}")
print(f"P-Value: {p_value}")
print(f"Critical Value (t_critical): {t_critical}")

# Decision based on t-statistic and critical value
if t_statistic < t_critical:
    print("Reject the null hypothesis: The mean weight is significantly less than the hypothesized mean based
else:
    print("Fail to reject the null hypothesis: There is not enough evidence to suggest the mean weight is les

# Decision based on p-value
if p_value < alpha:
    print("Reject the null hypothesis: The mean weight is significantly less than the hypothesized mean based
else:
    print("Fail to reject the null hypothesis: There is not enough evidence to suggest the mean weight is les
```

```
t Statistic: -1.0987082008162468
P-Value: 0.15020834437485206
Critical Value (t_critical): -1.8331129326536337
Fail to reject the null hypothesis: There is not enough evidence to suggest the mean weight is less than the hypothesized mean based on the t-statistic.
Fail to reject the null hypothesis: There is not enough evidence to suggest the mean weight is less than the hypothesized mean based on the p-value.
```

Chi-Square Test Example

You have a dataset with the observed frequencies of people preferring two different product types (A and B) across genders. You want to test if there's an association between gender and product preference.

In [14]:

```python
import pandas as pd
# Observed frequencies
observed = pd.DataFrame([[20, 15], [30, 35]], columns=["Product A", "Product B"], index=["Male", "Female"])
observed
```

Out[14]:

	Product A	Product B
Male	20	15
Female	30	35

ho; no relationship

h1: a relation ship

In [15]:

```python
# Chi-Square Test of Independence
chi2, p, dof, expected = stats.chi2_contingency(observed)
print(f"Chi-Square Statistic: {chi2}, P-value: {p}")
```

```
Chi-Square Statistic: 0.7032967032967032, P-value: 0.4016781664697727
```

In [16]:
```
p
#0.4016781664697727<.05 # rej null, relation
```

Out[16]:
```
0.4016781664697727
```

In [17]:
```
if p <= .05:
    print( "Reject null")
else:
    print("Fail to reject null") # accpet null
```

```
Fail to reject null
```

Since the p-value (0.402) is higher than a typical significance level (e.g., 0.05), we fail to reject the null hypothesis. This means that there is no statistically significant association between gender and the preference for Product A or Product B in this sample.

In other words, based on this data, we do not have enough evidence to conclude that gender impacts the choice between Product A and Product B.

In [ ]:
```

```