HYPOTHESIS TESTING

Hypothesis Testing :

A statistical hypothesis test is a method of statistical inference used to decide whether the data at hand sufficiently support a particular hypothesis. Hypothesis testing allows us to make probabilistic statements about population parameters.

Example of a hypothesis:

Let's say the average view time of my YouTube videos is 5 minutes before I added thumbnails. After adding thumbnails, I created 5 new videos, which have an average view time of 7 minutes. Can I say that adding thumbnails to videos significantly increased the average view time of my YouTube videos?

Note :

I made 35 videos with thumbnails, which have an average view time greater than videos without thumbnails. However, this doesn’t mean all my videos with thumbnails will have a higher average view time. The 35 videos with thumbnails are a sample, and we use them to make inferences about the population data. This means we can’t conclusively say that videos with thumbnails will always have a greater average view time.

Steps involved in Hypothesis Testing :

Rejection Region Approach

1. Formulate a Null and Alternate hypothesis
2. Select a significance level(This is the probability of rejecting the null hypothesis when it is actually true, usually set at 0.05 or 0.01)
3. Check assumptions (example distribution)
4. Decide which test is appropriate(Z-test, T-test, Chi-square test, ANOVA)
5. State the relevant test statistic
6. Conduct the test
7. Reject or not reject the Null Hypothesis.
8. Interpret the result

HYPOTHESIS TESTING

Performing a Z test Example 1

Suppose a company is evaluating the impact of a new training program on the productivity of its employees. The company has data on the average productivity of its employees before implementing the training program. The average productivity was 50 units per day with a known population standard deviation of 5 units. After implementing the training program, the company measures the productivity of a random sample of 30 employees. The sample has an average productivity of 53 units per day. The company wants to know if the new training program has significantly increased productivity.

Perform Test :

Given Information

- **Population mean (μ_0) = 50 units per day (before training)**
- **Population standard deviation (σ) = 5 units per day**
- **Sample size (n) = 30**
- **Sample mean ($\bar{x}$) = 53 units per day (after training)**
- **Significance level (α) = Typically 0.05, unless otherwise specified**

Step-by-Step Z-Test

1. **Set Up Hypotheses:**
    - **Null Hypothesis (H_0): μ = 50 (The training program has no effect on productivity; the mean productivity remains 50 units per day).**
    - **Alternative Hypothesis (H_1): μ > 50 (The training program increases productivity, so the mean productivity is greater than 50 units per day).**
2. **Calculate the Test Statistic (Z-score): The Z-score formula for a one-sample Z-test is:**

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

where:

- **x¯ is the sample mean,**
- **μ0 is the population mean,**
- **σ is the population standard deviation, and**
- **n is the sample size.**

Plugging in the values:

$$Z = \frac{53 - 50}{\frac{5}{\sqrt{30}}}$$

3. **Compute the Z-score:**

Let's calculate it:

The calculated Z-score is approximately 3.29.

4. **Determine the Critical Value and P-value:**
   - **For a one-tailed test at a 0.05 significance level, the critical Z-value is approximately 1.645.**
   - **Since 3.29 > 1.645, the test statistic falls in the rejection region.**
5. **Conclusion:**
   - **Decision: Since the Z-score is greater than the critical value, we reject the null hypothesis.**
   - **Interpretation: There is statistically significant evidence to suggest that the new training program has increased the average productivity of employees.**

Example 2

Suppose a snack food company claims that their Lays wafer packets contain an average weight of 50 grams per packet. To verify this claim, a consumer watchdog organization decides to test a random sample of Lays wafer packets. The organization wants to determine whether the actual average weight differs significantly from the claimed 50 grams. The organization collects a random sample of 40 Lays wafer packets and measures their weights. They find that the sample has an average weight of 49 grams, with a known population standard deviation of 4 grams.

Perform Test :

HYPOTHESIS TESTING

Given Information

- **Claimed average weight (Population mean, μ_0) = 50 grams**
- **Population standard deviation (σ) = 4 grams**
- **Sample size (n) = 40**
- **Sample mean ($\bar{x}$) = 49 grams**
- **Significance level (α) = Typically 0.05 for a two-tailed test, unless otherwise specified**

Step-by-Step Z-Test

1. **Set Up Hypotheses:**
   - **Null Hypothesis (H_0): $\mu = 50$ grams (The average weight is as claimed, 50 grams per packet).**
   - **Alternative Hypothesis (H_1): $\mu \neq 50$ grams (The average weight differs from 50 grams per packet).**
2. **Calculate the Test Statistic (Z-score): The Z-score formula for a one-sample Z-test is:**

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

where:

- **$\bar{x}$ is the sample mean,**
- **$\mu 0$ is the population mean (claimed weight),**
- **σ is the population standard deviation, and**
- **nis the sample size.**

Plugging in the values:

$$Z = \frac{49 - 50}{\frac{4}{\sqrt{40}}}$$

3. **Compute the Z-score:**

Let's calculate it:

The calculated Z-score is approximately -1.58.

4. **Determine the Critical Value and P-value:**

Yogeshchouhan263@gmail.com

HYPOTHESIS TESTING

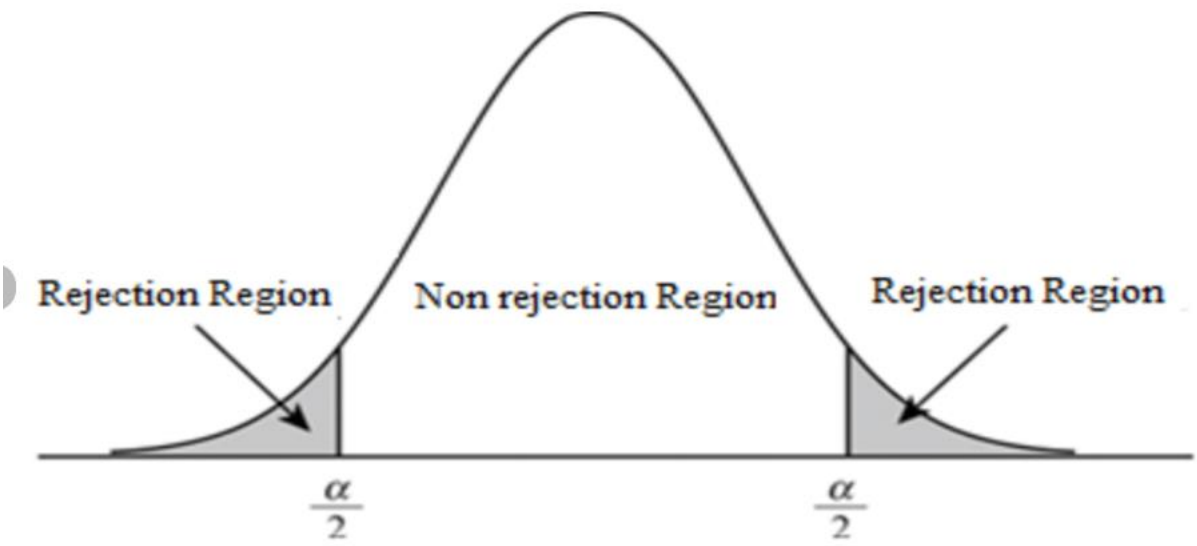
  - For a two-tailed test at a 0.05 significance level, the critical Z-values are approximately ±1.96.
  - Since −1.58-1.58−1.58 lies within the range −1.96-1.96−1.96 to +1.96+1.96+1.96, it does not fall in the rejection region.

5. **Conclusion:**
  - **Decision: We fail to reject the null hypothesis since the Z-score does not exceed the critical values.**
  - **Interpretation: There is no statistically significant evidence to suggest that the actual average weight of the Lays wafer packets differs from the claimed 50 grams.**

Rejection Region :

Significance level - denoted as α (alpha), is a predetermined threshold used in hypothesis testing to determine whether the null hypothesis should be rejected or not. It represents the probability of rejecting the null hypothesis when it is actually true, also known as Type 1 error. The critical region is the region of values that corresponds to the rejection of the null hypothesis at some chosen probability level.

The critical region is the region of values that corresponds to the rejection of the null hypothesis at some chosen probability level.



HYPOTHESIS TESTING

Problem with Rejection Region Approach :

1. All-or-Nothing Decision: It gives a strict "reject" or "do not reject" outcome based on a fixed threshold (like 0.05). This can oversimplify results, as it doesn't show how strongly the data supports rejecting the null hypothesis.
2. Ignores P-value Nuances: It doesn't consider the actual p-value, which tells us the strength of the evidence. Instead, it only checks if the p-value is below a set significance level, missing out on more nuanced information.
3. Less Informative: It doesn't indicate the probability of a Type I error (rejecting a true null) or Type II error (failing to reject a false null), making it less informative for understanding potential errors in decision-making.

Type 1 vs Type 2 Error :

In hypothesis testing, there are two types of errors that can occur when making a decision about the null hypothesis: Type I error and Type II error.

	H0 True	H0 False
Reject H0	TYPE 1 ERROR	CORRECT DICISION
Acept H0	CORRECT DICISION	TYPE 2 ERROR

Type-I (False Positive) : error occurs when the sample results, lead to the rejection of the null hypothesis when it is in fact true. In other words, it's the mistake of finding a significant effect or relationship when there is none. The probability of committing a Type I error is denoted by α (alpha), which is also known as the significance level. By choosing a significance level, researchers can control the risk of making a Type I error.

Type-II (False Negative): error occurs when based on the sample results, the null hypothesis is not rejected when it is in fact false. This means that the researcher fails to detect a significant effect or relationship when one actually exists. The probability of committing a Type II error is denoted by β (beta). Trade-off between Type 1 and Type 2 errors.

Trade-Off:

- If we lower the probability of making a Type I error (by choosing a stricter significance level, like 0.01 instead of 0.05), we make it harder to reject the null hypothesis. This reduces the chances of a false positive but increases the risk of a **Type II error** (missing a real effect).

- Conversely, if we make it easier to detect an effect (increasing the significance level), we reduce the chance of a **Type II error** but increase the risk of a **Type I error**.

In essence, there's a balance: **minimizing one type of error generally increases the other**.

One sided vs two sided test :

One-sided (one-tailed) test: A one-sided test is used when the researcher is interested in testing the effect in a specific direction (either greater than or less than the value specified in the null hypothesis). The alternative hypothesis in a one-sided test contains an inequality (either ">" or "<").

Example: A researcher wants to test whether a new medication increases the average recovery rate compared to the existing medication.

Two-sided (two-tailed) test: A two-sided test is used when the researcher is interested in testing the effect in both directions (i.e., whether the value specified in the null hypothesis is different, either greater or lesser). The alternative hypothesis in a two-sided test contains a "not equal to" sign (≠).

Example: A researcher wants to test whether a new medication has a different average recovery rate compared to the existing medication. The main difference between them lies in the directionality of the alternative hypothesis andhow the significance level is distributed in the critical regions.

Advantages and Disadvantages?

Two-tailed test (two-sided):

Advantages:

Detects effects in both directions: Two-tailed tests can detect effects in both directions, which makes them suitable for situations where the direction of the effect is uncertain or when researchers want to test for any difference between the groups or variables.

1.More conservative: Two-tailed tests are more conservative because the significance level (α) is split between both tails of the distribution. This reduces the risk of Type I errors in cases where the direction of the effect is uncertain.

2.Disadvantages:

Less powerful: Two-tailed tests are generally less powerful than one-tailed tests because the significance level (α) is divided between both tails of the distribution. This means the test requires a larger effect size to reject the null hypothesis, which could lead to a higher risk of Type II errors (failing to reject the null hypothesis when it is false).

1.Not appropriate for directional hypotheses: Two-tailed tests are not ideal for cases where the research question or hypothesis is directional, as they test for differences in both directions, which may not be of interest or relevance.

One-tailed test (one-sided):

Advantages:

More powerful: One-tailed tests are generally more powerful than two-tailed tests, as the entire significance level (α) is allocated to one tail of the distribution. This means that the test is more likely to detect an effect in the specified direction, assuming the effect exists.

1.Directional hypothesis: One-tailed tests are appropriate when there is a strong theoretical or practical reason to test for an effect in a specific direction.

2.Disadvantages:

Missed effects: One-tailed tests can miss effects in the opposite direction of the specified alternative hypothesis. If an effect exists in the opposite direction, the test will not be able to detect it, which could lead to incorrect conclusions.

1.Increased risk of Type I error: One-tailed tests can be more prone to Type I errors if the effect is actually in the opposite direction than the one specified in the alternative hypothesis.

P-value :

P-value is the probability of getting a sample as or more extreme(having more evidence against H0) than our own sample given the Null Hypothesis(H0) is true.

In simple words p-value is a measure of the strength of the evidence against the Null Hypothesis that is provided by our sample data.

Interpreting p-value :

1. Very small p-values (e.g., $p < 0.01$) indicate strong evidence against the null hypothesis,suggesting that the observed effect or difference is unlikely to have occurred by chance alone.

2. Small p-values (e.g., $0.01 \le p < 0.05$) indicate moderate evidence against the null hypothesis, suggesting that the observed effect or difference is less likely to have occurred by chance alone.

3.Large p-values (e.g., $0.05 \le p < 0.1$) indicate weak evidence against the null hypothesis, suggesting that the observed effect or difference might have occurred by chance alone, but there is still some level of uncertainty.

4. Very large p-values (e.g., $p \ge 0.1$) indicate weak or no evidence against the null hypothesis, suggesting that the observed effect or difference is likely to have occurred by chance alone.

P-value in context of Z-test:

Problem. Suppose a company is evaluating the impact of a new training program on the productivity of its employees. The company has data on the average productivity of its employees before implementing the training program. The average productivity was 50 units per day. After implementing the training program, the company measures the productivity of a random sample of 30 employees. The sample has an average productivity of 53 units per day and the pop std is 4. The company wants to know if the new training program has significantly increased productivity.

Given Information

- **Claimed average productivity (Population mean, μ_0) = 50 grams**
- **Population standard deviation (σ) = 4 grams**
- **Sample size (n) = 30**
- **Sample mean ($\bar{x}$) = 53 grams**
- **Significance level (α) = Typically 0.05 for a two-tailed test, unless otherwise specified**

Step-by-Step Z-Test

1. **Set Up Hypotheses:**
    - **Null Hypothesis (H_0): $\mu = 50$ grams (The training program has no effect on productivity).**
    - **Alternative Hypothesis (H_1): $\mu > 50$ grams (The training program has increased productivity).**
2. **Calculate the Test Statistic (Z-score): The Z-score formula for a one-sample Z-test is:**

Yogeshchouhan263@gmail.com

HYPOTHESIS TESTING

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

where:

- **x̄ is the sample mean,**
- **μ_0 is the population mean (claimed weight),**
- **σ is the population standard deviation, and**
- **nis the sample size.**

Plugging in the values:

$$z = \frac{53 - 50}{\frac{4}{\sqrt{30}}} = \frac{3}{\frac{4}{5.477}} = \frac{3}{0.7303} = 4.11$$

3. Find the p-value

Now, we find the p-value corresponding to the z-score of 4.11 using standard normal distribution tables or a calculator. For a z-value of 4.11:

- **The p-value is the area to the right of z = 4.11 on the standard normal curve.**
- **Using a z-table or calculator, we find the p-value ≈ 0.00002.**

4. Compare the p-value with the significance level

Let's assume a significance level α=0.05.

Since the p-value (0.00002) is much smaller than 0.05, we reject the null hypothesis.

5. Conclusion

We have sufficient evidence to conclude that the new training program has significantly increased employee productivity. The p-value is very small, indicating strong evidence against the null hypothesis.

6. Strength of Evidence

The p-value of 0.00002 is extremely small, which means the evidence is very strong. It suggests that the observed increase in productivity (from 50 to 53 units) is highly unlikely to have occurred by chance, and the training program has a significant impact on productivity.

HYPOTHESIS TESTING

T – tests:

In real-world scenarios, we usually don't know the population standard deviation, so we typically use the t-test instead of the z-test .

A t-test is a statistical test used in hypothesis testing to compare the means of two samples or to compare a sample mean to a known population mean. The t-test is based on the t- distribution, which is used when the population standard deviation is unknown and the sample size is small.

There are three main types of t-tests:

One-sample t-test: The one-sample t-test is used to compare the mean of a single sample to a known population mean. The null hypothesis states that there is no significant difference between the sample mean and the population mean, while the alternative hypothesis states that there is a significant difference.

Independent two-sample t-test: The independent two-sample t-test is used to compare the means of two independent samples. The null hypothesis states that there is no significant difference between the means of the two samples, while the alternative hypothesis states that there is a significant difference.

Paired t-test (dependent two-sample t-test): The paired t-test is used to compare the means of two samples that are dependent or paired, such as pre-test and post-test scores for the same group of subjects or measurements taken on the same subjects under two different conditions. The null hypothesis states that there is no significant difference between the means of the paired differences, while the alternative hypothesis states that there is a significant difference.

Single Sample t-test :

A one-sample t-test checks whether a sample mean differs from the population mean.

Assumptions for a single sample t-test

1. Normality - Population from which the sample is drawn is normally distributed

2.Independence - The observations in the sample must be independent, which means that the value of one observation should not influence the value of another observation.

3.Random Sampling - The sample must be a random and representative subset of the population.

4. Unknown population std - The population std is not known.

HYPOTHESIS TESTING

Problem. Suppose a manufacturer claims that the average weight of their new chocolate bars is 50 grams, we highly doubt that and want to check this so we drew out a sample of 25 chocolate bars and measured their weight, the sample mean came out to be 49.7 grams and the sample std deviation was 1.2 grams. Consider the significance level to be 0.05.

Given Information

- **Claimed average weight(Population mean, μ_0) = 50 grams**
- **Sample standard deviation (s) = 1.2 grams**
- **Sample size (n) = 25**
- **Sample mean ($\bar{x}$) = 49.7 grams**
- **Significance level (α) = Typically 0.05 for a two-tailed test, unless otherwise specified**

Step-by-Step T-Test

1. **Set Up Hypotheses:**
   - **Null Hypothesis (H_0): μ = 50 grams (The average weight of the chocolate bars is 50 grams).**
   - **Alternative Hypothesis (H_1): μ != 50 grams (The average weight of the chocolate bars is not 50 grams).**
2. **Check Assumptions for the One-Sample t-test:**

   1.Normality: For this we use Shapiro wilk test (The Shapiro-Wilk test returns two values: the test statistic (W) and the p-value.
   If the p-value is greater than your chosen significance level (α = 0.05), you can assume the data comes from a normally distributed population.
   If the p-value is less than or equal to the significance level, the normality assumption is not met,
   and you should consider applying a data transformation or using a non-parametric test like the Mann-Whitney U test.
).
   Let assume our population data is normally distributed and all the 3 assumptions are true.

3. Calculate the Test Statistic

Given the assumptions are met, we proceed with the test statistic calculation:

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

HYPOTHESIS TESTING

Substitute the values:

$$t = \frac{49.7 - 50}{\frac{1.2}{\sqrt{25}}} = \frac{-0.3}{0.24} = -1.25$$

4. Degrees of Freedom

The degrees of freedom (df) for a one-sample t-test is n−1.

Df = 25−1 = 24

5. Find the Critical t-value and p-value(using T-Table)

At a significance level of α=0.05for a two-tailed test and 24 degrees of freedom, we can look up the critical t-value:

- **The critical t-value for a two-tailed test with α = 0.05 and df = 24 is approximately ±2.064.**

Since |t|=1.25 is less than 2.064, we do not fall into the rejection region.

6. Calculate the p-value

```
from scipy.stats import t
# set the t value and degree of freedom
t_stats = -1.25
df = 24

# calculate the cdf value
p_value = t.cdf(t_stats,df)
p_value

# for Two-Tails test Multiply By 2
print('P_value:',p_value*2)
```

```
P_value: 0.2233514781656204
```

The p-value for t = −1.25 with 24 degrees of freedom is approximately 0.224 for a two-tailed test.

7. Decision

Since the p-value (0.224) is greater than the significance level (0.05), we fail to reject the null hypothesis.

8. Conclusion

HYPOTHESIS TESTING

At the 5% significance level, there is not enough evidence to conclude that the average weight of the chocolate bars is different from the claimed 50 grams.

Independent 2 sample t-test:

An independent two-sample t-test, also known as an unpaired t-test, is a statistical method used to compare the means of two independent groups to determine if there is a significant difference between them.

Assumptions for the test:

1.Independence of observations: The two samples must be independent, meaning there is no relationship between the observations in one group and the observations in the other group. The subjects in the two groups should be selected randomly and independently.

2.Normality: The data in each of the two groups should be approximately normally

distributed. The t-test is considered robust to mild violations of normality, especially when the sample sizes are large (typically n ≥ 30) and the sample sizes of the two groups are similar. If the data is highly skewed or has substantial outliers, consider using a non-parametric test, such as the Mann-Whitney U test.

3.Equal variances (Homoscedasticity): The variances of the two populations should be approximately equal. This assumption can be checked using F-test for equality of variances. If this assumption is not met, you can use Welch's t-test, which does not require equal variances.

4. Random sampling: The data should be collected using a random sampling method from the respective populations. This ensures that the sample is representative of the population and reduces the risk of selection bias.

PROBLEM : Suppose a website owner claims that there is no difference in the average time spent on their website between desktop and mobile users. To test this claim, we collect data from 30 desktop users and 30 mobile users regarding the time spent on the website in minutes. The sample statistics are as follows:

desktop users = [12, 15, 18, 16, 20, 17, 14, 22, 19, 21, 23, 18, 25, 17, 16, 24, 20, 19, 22, 18, 15,14, 23, 16, 12, 21, 19, 17, 20, 14]

mobile_users = [10, 12, 14, 13, 16, 15, 11, 17, 14, 16, 18, 14, 20, 15, 14, 19, 16, 15, 17, 14, 12,11, 18, 15, 10, 16, 15, 13, 16, 11]

HYPOTHESIS TESTING

Given Information

Desktop users :

- Sample size (n1): 30
- Sample mean (mean1): 18.5 minutes
- Sample standard deviation (std_dev1): 3.5 minutes

Mobile users:

- Sample size (n2): 30
- Sample mean (mean2): 14.3 minutes
- Sample standard deviation (std_dev2): 2.7 minutes

Steps for Two-Sample t-Test

1. State the Hypotheses

- **Null hypothesis (H_0):** There is no difference in the average time spent on the website between desktop and mobile users.

  H0: μ desktop = μ mobile

- **Alternative hypothesis (H_1):** There is a difference in the average time spent on the website between desktop and mobile users.

  H1: μ desktop ≠ μ mobile

This is a two-tailed test because we are testing for a difference in either direction.

2. Check Assumptions

The assumptions for a two-sample t-test include:

1. **Independence**: The two groups (desktop and mobile users) are independent of each other.
2. **Normality**: The distribution of time spent for each group should be approximately normal. Since we have 30 samples in each group, the Central Limit Theorem suggests that the t-test is robust to minor deviations from normality. (and let say if we have sample size <30 then we perform Shapiro wilk test)
3. **Equal Variances**: Ideally, the variances of the two groups should be equal. We can test this assumption by using Levene's test or use the Welch's t-test if the variances are not equal.

HYPOTHESIS TESTING

```python
from scipy.stats import shapiro

# Decide significance label
alpha = 0.05

# Input the data as lists
desktop_users = [12, 15, 18, 16, 20, 17, 14, 22, 19, 21, 23, 18, 25, 17, 16, 24, 20, 19, 22, 18, 15, 14, 23, 16,
mobile_users = [10, 12, 14, 13, 16, 15, 11, 17, 14, 16, 18, 14, 20, 15, 14, 19, 16, 15, 17, 14, 12, 11, 18, 15,

# Perform the Shapiro-Wilk test for both desktop and mobile users
shapiro_desktop = shapiro(desktop_users)
shapiro_mobile = shapiro(mobile_users)

print("Shapiro-Wilk test for desktop users:", shapiro_desktop)
print("Shapiro-Wilk test for mobile users:", shapiro_mobile)

# Interpret Results
desktop_pvalue = shapiro_desktop[1]
mobile_pvalue = shapiro_mobile[1]
if desktop_pvalue>alpha:
  print('Desktop Data Normally Distributed')
else:
  print('Deskto Data Not Normally Distributed')
if mobile_pvalue>alpha:
  print('Mobile Data Normally Distributed')
else:
  print('Mobile Data Not Normally Distributed')
```

```
Shapiro-Wilk test for desktop users: ShapiroResult(statistic=0.9783115512411942, pvalue=0.7791003299808725)
Shapiro-Wilk test for mobile users: ShapiroResult(statistic=0.9714355768676655, pvalue=0.5791606602037616)
Desktop Data Normally Distributed
Mobile Data Normally Distributed
```

```python
# If the p-value from Levene's test is greater than your chosen significance level (α = 0.05), you can assume equal variances
#  If the p-value is less than or equal to the significance level, the assumption of equal variances is not met,
# and you should consider using Welch's t-test instead of the regular independent two-sample t-test.

from scipy.stats import levene
# Input the data as lists
desktop_users = [12, 15, 18, 16, 20, 17, 14, 22, 19, 21, 23, 18, 25, 17, 16, 24, 20, 19, 22, 18, 15, 14, 23, 16, 12, 21, 19, 17
mobile_users = [10, 12, 14, 13, 16, 15, 11, 17, 14, 16, 18, 14, 20, 15, 14, 19, 16, 15, 17, 14, 12, 11, 18, 15, 10, 16, 15, 13,

# Perform Levene's test
levene_test = levene(desktop_users, mobile_users)
print(levene_test)

# Decide signigicance level
alpha = 0.05

# Since We P-value is Greater then significance level
# we Can say our equal variance assumption is True
```

```
LeveneResult(statistic=2.94395488191752, pvalue=0.09153720526741761)
```

Yogeshchouhan263@gmail.com

3. Calculate the Test Statistic

```
from scipy.stats import ttest_ind

# Data for desktop and mobile users
desktop_users = [12, 15, 18, 16, 20, 17, 14, 22, 19, 21, 23, 18, 25, 17, 16, 24, 20, 19,
                 14, 23, 16, 12, 21, 19, 17, 20, 14]
mobile_users = [10, 12, 14, 13, 16, 15, 11, 17, 14, 16, 18, 14, 20, 15, 14, 19, 16, 15, 1
                11, 18, 15, 10, 16, 15, 13, 16, 11]

# Perform the two-sample t-test
t_stat, p_value = ttest_ind(desktop_users, mobile_users, equal_var=True)

# Display the results
print("t-statistic:", t_stat)
print("p-value:", p_value)
```

```
t-statistic: 4.625335930681123
p-value: 2.1422811334975257e-05
```

The p-value for t = 4.6253359 is approximately 0.0000214 (approximately)

4. Decision

Since the p-value is much smaller than the common significance level (e.g., 0.05), we reject the null hypothesis.

5.Conclusion

This suggests that there is a statistically significant difference in the average time spent on the website between desktop and mobile users.

2 Sample t test on Titanic Dataset: (In Notebook)

Paired 2 sample t-test:

A paired two-sample t-test, also known as a dependent or paired-samples t-test, is a statistical test used to compare the means of two related or dependent groups.

Common scenarios where a paired two-sample t-test is used include:

1.Before-and-after studies: Comparing the performance of a group before and after an intervention or treatment.

2. Matched or correlated groups: Comparing the performance of two groups that are matched or correlated in some way, such as siblings or pairs of individuals with similar characteristics.

HYPOTHESIS TESTING

Assumptions

1.Paired observations: The two sets of observations must be related or paired in some way, such as before-and-after measurements on the same subjects or observations from matched or correlated groups.

2.Normality: The differences between the paired observations should be approximately normally distributed. This assumption can be checked using graphical methods (e.g., histograms, Q-Q plots) or statistical tests for normality (e.g., Shapiro-Wilk test).

Note that: the t-test is generally robust to moderate violations of this assumption when the sample size is large.

3.Independence of pairs: Each pair of observations should be independent of other pairs. In other words, the outcome of one pair should not affect the outcome of another pair. This assumption is generally satisfied by appropriate study design and random sampling.

Problem. Let's assume that a fitness center is evaluating the effectiveness of a new 8-week weight loss program. They enroll 15 participants in the program and measure their weights before and after the program. The goal is to test whether the new weight loss program leads to a significant reduction in the participants' weight.

Before the program:

[80, 92, 75, 68, 85, 78, 73, 90, 70, 88, 76, 84, 82, 77, 91]

After the program:

[78, 93, 81, 67, 88, 76, 74, 91, 69, 88, 77, 81, 80, 79, 88]

Steps for Paired Two-Sample t-Test

1. State the Hypotheses

- **Null hypothesis (H_0):** There is no significant difference in weights before and after the program.

  H0: μ Before = μ after

- **Alternative hypothesis (H_1):** There is a significant reduction in weights after the program.

  H1: μ Before > μ after

2. Check Assumptions

1. Paired observations: True

HYPOTHESIS TESTING

2.Normality:

```
# Input the data as lists
Before_program = [80, 92, 75, 68, 85, 78, 73, 90, 70, 88, 76, 84, 82, 77, 91]
After_program = [78, 93, 81, 67, 88, 76, 74, 91, 69, 88, 77, 81, 80, 79, 88]

# Perform the Shapiro-Wilk test for both desktop and mobile users
shapiro_before= shapiro(Before_program)
shapiro_after = shapiro(Before_program)

print("Shapiro-Wilk test for shapiro_before :", shapiro_before)
print("Shapiro-Wilk test for shapiro_after:", shapiro_after)

# Since Pvalue Is Greater Then alpha = 0.05 We Can Say Our Normality Assumption is True
```

```
Shapiro-Wilk test for shapiro_before : ShapiroResult(statistic=0.960822970929243, pvalue=0.7066999995867655)
Shapiro-Wilk test for shapiro_after: ShapiroResult(statistic=0.960822970929243, pvalue=0.7066999995867655)
```

3 . independence of Pairs : True

3. Calculate the Test Statistic

Calculate T-stats For Paired Two Sample T Test

```
# Input the data as lists
Before_program = np.array([80, 92, 75, 68, 85, 78, 73, 90, 70, 88, 76, 84, 82, 77, 91])
After_program = np.array([78, 93, 81, 67, 88, 76, 74, 91, 69, 88, 77, 81, 80, 79, 88])

# Formula : t_stats = difference.mean() / (difference.std()/ sqrt(sample size ))

difference = Before_program - After_program

mean_diff = difference.mean()
std_diff = difference.std()
sample_size = len(Before_program)

t_stats = mean_diff / (std_diff / np.sqrt(sample_size))
print('t_statistics : ',t_stats)
```

```
t_statistics :  -0.10850778933039285
```

4. Calculate p_value:

```
# Calculate P_value

from scipy.stats import t
# set the t value and degree of freedom
t_statistics = t_stats
df = sample_size - 1

# calculate the cdf value
p_value = t.cdf(t_statistics,df)
p_value

print('P_value:',p_value)
```

```
P_value: 0.45756637096100794
```

Yogeshchouhan263@gmail.com

5. Result :

Since P_value is Greater Then significance label (alpha = 0.05) .We Can not Reject The Null Hypothesis. That Means There is a significant reduction in weights after the program.

HYPOTHESIS TESTING

Yogeshchouhan263@gmail.com