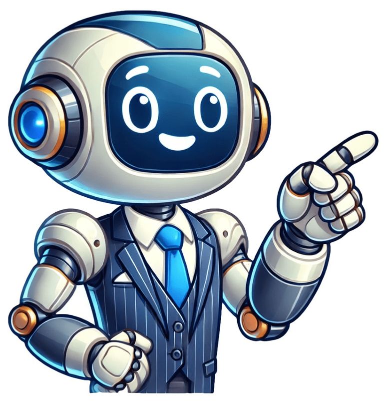

We are living in the most dynamic time in human history. Things around us are changing at a rapid rate, mostly fueled by rapid advancement in the fields of machine learning and artificial intelligence. Businesses have now at their disposal significantly improved decision making tools than a decade or so ago.But with the rise in the availability of means to make a decision, the onus is now on engineers and business to leverage technology to support rapid decision making.Consider a software testing platform provider wants to revamp their UI to know if the changes in UI improve their user experience and lead to more users utilizing their software testing platform. The correct way to assess if this proposed change is really correct is to do an A/B test, i.e. release the new feature to a controlled group of users and draw insights from their experience of the new UI.A/B testing falls under Hypothesis testing, a statistical method to determine if there is enough evidence from a sample data to draw conclusions about a wider population.In this blog we will explore in detail what is Hypothesis testing, what is the advantage of Hypothesis testing and types of hypothesis testing using Python, a popular open source language.Hypothesis testing is defined as a statistical tool that helps us verify or deny if an assumption about a population based on sample data Hypothesis testing is a statistical technique that allows us to draw conclusions about a population based on a sample of data. It is often used in fields like medicine, psychology, and economics to test the effectiveness of new treatments, analyze consumer behavior, or estimate the impact of policy changes.In Python, hypothesis testing is facilitated by modules such as scipy.stats and statsmodels.stats. In this article, well explore three examples of hypothesis testing in Python: the one sample t-test, the two sample t-test, and the paired samples t-test.For each test, well provide a brief explanation of the underlying concepts, an example of a research question that can be answered using the test, and a step-by-step guide to performing the test in Python. Lets get started!One Sample t-testThe one sample t-test is used to compare a sample mean to a known or hypothesized population mean. This allows us to determine whether the sample mean is significantly different from the population mean.The test assumes that the data are normally distributed and that the sample is randomly drawn from the population. Example research question: Is the mean weight of a species of turtle significantly different from a known or hypothesized value?Step-by-step guide:Define the null hypothesis (H0) and alternative hypothesis (Ha).The null hypothesis is typically that the sample mean is equal to the population mean. The alternative hypothesis is that they are not equal.For example:H0: The mean weight of a species of turtle is 100 grams. Ha: The mean weight of a species of turtle is not 100 grams.Collect a random sample of data.This can be done using Pythons random module or by importing data from a file. For example: weight_sample = [95, 105, 110, 98, 102, 116, 101, 99, 104, 108]Calculate the sample mean (x), sample standard deviation (s), and standard error (SE).For example: x = sum(weight_sample)/len(weight_sample) s = np.std(weight_sample) SE = s / (len(weight_sample)**0.5)Calculate the t-value using the formula: t = (x -) / (SE), where is the hypothesized population mean.For example: t = (x - 100) / SECalculate the p-value using a t-distribution table or a Python function like scipy.stats.ttest_1samp().For example: p_value = scipy.stats.ttest_1samp(weight_sample, 100).pvalueCompare the p-value to the level of significance (), typically set to 0.05.If the p-value is less than , reject the null hypothesis and conclude that there is sufficient evidence to support the alternative hypothesis.If the p-value is greater than , fail to reject the null hypothesis and conclude that there is insufficient evidence to support the alternative hypothesis. For example: if p_value < 0.05: print("Reject the null hypothesis.") else: print("Fail to reject the null hypothesis.") Two Sample t-testThe two sample t-test is used to compare the means of two independent samples. This allows us to determine whether the means are significantly different from each other.The test assumes that the data are normally distributed and that the samples are randomly drawn from their respective populations. Example research question: Is the mean weight of two different species of turtles significantly different from each other?Step-by-step guide:Define the null hypothesis (H0) and alternative hypothesis (Ha).The null hypothesis is typically that the sample means are equal. The alternative hypothesis is that they are not equal.For example:H0: The mean weight of species A is equal to the mean weight of species B. Ha: The mean weight of species A is not equal to the mean weight of species B.Collect two random samples of data.This can be done using Python's random module or by importing data from a file. For example: species_a = [95, 105, 110, 98, 102] species_b = [116, 101, 99, 104, 108]Calculate the sample means (x1, x2), sample standard deviations (s1, s2), and pooled standard error (SE).For example: x1 = sum(species_a)/len(species_a) x2 = sum(species_b)/len(species_b) s1 = np.std(species_a) s2 = np.std(species_b) n1 = len(species_a) n2 = len(species_b) SE = (((n1-1)*s1**2 + (n2-1)*s2**2)/(n1+n2-2))**0.5 * (1/n1 + 1/n2)**0.5Calculate the t-value using the formula: t = (x1 - x2) / (SE), where x1 and x2 are the sample means.For example: t = (x1 - x2) / SECalculate the p-value using a t-distribution table or a Python function like scipy.stats.ttest_ind().For example: p_value = scipy.stats.ttest_ind(species_a, species_b).pvalueCompare the p-value to the level of significance (), typically set to 0.05.If the p-value is less than , reject the null hypothesis and conclude that there is sufficient evidence to support the alternative hypothesis.If the p-value is greater than , fail to reject the null hypothesis and conclude that there is insufficient evidence to support the alternative hypothesis. For example: if p_value < 0.05: print("Reject the null hypothesis.") else: print("Fail to reject the null hypothesis.") Paired Samples t-testThe paired samples t-test is used to compare the means of two related samples. This allows us to determine whether the means are significantly different from each other, while accounting for individual differences between the samples.The test assumes that the differences between paired observations are normally distributed. Example research question: Is there a significant difference in the max vertical jump of basketball players before and after a training program?Step-by-step guide:Define the null hypothesis (H0) and alternative hypothesis (Ha).The null hypothesis is typically that the mean difference is equal to zero. The alternative hypothesis is that it is not equal to zero.For example:H0: The mean difference in max vertical jump before and after training is zero. Ha: The mean difference in max vertical jump before and after training is not zero.Collect two related samples of data.This can be done by measuring the same variable in the same subjects before and after a treatment or intervention. For example: before = [72, 69, 77, 71, 76] after = [80, 70, 75, 74, 78]Calculate the differences between the paired observations and the sample mean difference (d), sample standard deviation (s), and standard error (SE).For example: differences = [after[i]-before[i] for i in range(len(before))] d = sum(differences)/len(differences) s = np.std(differences) SE = s / (len(differences)**0.5)Calculate the t-value using the formula: t = (d -) / (SE), where is the hypothesized population mean difference (usually zero).For example: t = (d - 0) / SECalculate the p-value using a t-distribution table or a Python function like scipy.stats.ttest_rel().For example: p_value = scipy.stats.ttest_rel(after, before).pvalueCompare the p-value to the level of significance (), typically set to 0.05.If the p-value is less than , reject the null hypothesis and conclude that there is sufficient evidence to support the alternative hypothesis.If the p-value is greater than , fail to reject the null hypothesis and conclude that there is insufficient evidence to support the alternative hypothesis. For example: if p_value < 0.05: print("Reject the null hypothesis.") else: print("Fail to reject the null hypothesis.") Two Sample t-test in PythonThe two sample t-test is used to compare two independent samples and determine if there is a significant difference between the means of the two populations. In this test, the null hypothesis is that the means of the two samples are equal, while the alternative hypothesis is that they are not equal.Example research question: Is the mean weight of two different species of turtles significantly different from each other? Step-by-step guide:Define the null hypothesis (H0) and alternative hypothesis (Ha). The null hypothesis is that the mean weight of the two turtle species is the same.The alternative hypothesis is that they are not equal. For example:H0: The mean weight of species A is equal to the mean weight of species B.Ha: The mean weight of species A is not equal to the mean weight of species B. 2.Collect a random sample of data for each species. For example: species_a = [4.3, 3.9, 5.1, 4.6, 4.2, 4.8] species_b = [4.9, 5.2, 5.5, 5.3, 5.0, 4.7]Calculate the sample mean (x1, x2), sample standard deviation (s1, s2), and pooled standard error (SE). For example: import numpy as np from scipy.stats import ttest_ind x1 = np.mean(species_a) x2 = np.mean(species_b) s1 = np.std(species_a) s2 = np.std(species_b) n1 = len(species_a) n2 = len(species_b) SE = np.sqrt(s1**2/n1 + s2**2/n2)Calculate the t-value using the formula: t = (x1 - x2) / (SE), where x1 and x2 are the sample means.For example: t = (x1 - x2) / SECalculate the p-value using a t-distribution table or a Python function like ttest_ind().For example: p_value = ttest_ind(species_a, species_b).pvalueCompare the p-value to the level of significance (), typically set to 0.05.If the p-value is less than , reject the null hypothesis and conclude that there is sufficient evidence to support the alternative hypothesis. If the p-value is greater than , fail to reject the null hypothesis and conclude that there is insufficient evidence to support the alternative hypothesis.For example: alpha = 0.05 if p_value < alpha: print("Reject the null hypothesis.") else: print("Fail to reject the null hypothesis.")In this example, if the p-value is less than 0.05, we would reject the null hypothesis and conclude that there is a significant difference between the mean weight of the two turtle species.Paired Samples t-test in PythonThe paired samples t-test is used to compare the means of two related samples. In this test, the null hypothesis is that the difference between the two means is equal to zero, while the alternative hypothesis is that they are not equal.Example research question: Is there a significant difference in the max vertical jump of basketball players before and after a training program? Step-by-step guide:Define the null hypothesis (H0) and alternative hypothesis (Ha). The null hypothesis is that the mean difference in max vertical jump before and after the training program is zero.The alternative hypothesis is that it is not zero. For example:H0: The mean difference in max vertical jump before and after the training program is zero.Ha: The mean difference in max vertical jump before and after the training program is not zero. 2.Collect two related samples of data, such as the max vertical jump of basketball players before and after a training program. For example: before_training = [58, 64, 62, 70, 68] after_training = [62, 66, 64, 74, 70]Calculate the differences between the paired observations and the sample mean difference (d), sample standard deviation (s), and standard error (SE). For example: differences = [after_training[i]-before_training[i] for i in range(len(before_training))] d = np.mean(differences) s = np.std(differences) n = len(differences) SE = s / np.sqrt(n)Calculate the t-value using the formula: t = (d -) / (SE), where is the hypothesized population mean difference (usually zero).For example: t = d / SECalculate the p-value using a t-distribution table or a Python function like ttest_rel().For example: p_value = ttest_rel(after_training, before_training).pvalueCompare the p-value to the level of significance (), typically set to 0.05.If the p-value is less than , reject the null hypothesis and conclude that there is sufficient evidence to support the alternative hypothesis.If the p-value is greater than , fail to reject the null hypothesis and conclude that there is insufficient evidence to support the alternative hypothesis. For example: if p_value < alpha: print("Reject the null hypothesis.") else: print("Fail to reject the null hypothesis.")In this example, if the p-value is less than 0.05, we would reject the null hypothesis and conclude that there is a significant difference in the max vertical jump of basketball players before and after the training program.ConclusionHypothesis testing is an essential tool in statistical analysis, which gives us insights into populations based on limited data. The two sample t-test and paired samples t-test are two popular statistical methods that enable researchers to compare means of samples and determine whether they are significantly different.With the help of Python, hypothesis testing in practice is made more accessible and convenient than ever before. In this article, we have provided a step-by-step guide to performing these tests in Python, enabling researchers to perform rigorous analyses that generate meaningful and accurate results.In conclusion, hypothesis testing in Python is a crucial step in making conclusions about populations based on data samples. The three common hypothesis tests in Python; one-sample t-test, two-sample t-test, and paired samples t-test can be effectively applied to explore various research questions.By setting null and alternative hypotheses, collecting data, calculating mean and standard deviation values, computing t-value, and comparing it with the set significance level of , we can determine if there's enough evidence to reject the null hypothesis. With the use of such powerful methods, scientists can give more accurate and informed conclusions to real-world problems and take critical decisions when needed.Continual learning and expertise with hypothesis testing in Python tools can enable researchers to leverage this powerful statistical tool for better outcomes. Hypothesis Testing is a statistical method used to make inferences or decisions about a population based on sample data. It starts with a null hypothesis (H0), which represents a default stance or no effect, and an alternative hypothesis (H1 or Ha), which represents what we aim to prove or expect to find. The process involves using sample data to determine whether to reject the null hypothesis in favor of the alternative hypothesis, based on the likelihood of observing the sample data under the null hypothesis. So, if you want to learn how to perform Hypothesis Testing, this article is for you. In this article, Ill take you through the task of Hypothesis Testing using Python. So, Hypothesis Testing is a fundamental process in data science for making data-driven decisions and inferences about populations based on sample data. Below is the process we can follow for the task of Hypothesis Testing: Gather the necessary data required for the hypothesis test. Define Null (H0) and Alternative Hypothesis (H1 or Ha). Choose the Significance Level (), which is the probability of rejecting the null hypothesis when it is true. Select the appropriate statistical tests. Examples include t-tests for comparing means, chi-square tests for categorical data, and ANOVA for comparing means across more than two groups. Perform the chosen statistical test on your data. Determine the p-value and interpret the results of your statistical tests. To get started with Hypothesis Testing, we need appropriate data. I found an ideal dataset for this task. You can download the dataset from here. Now, lets get started with the task of Hypothesis Testing by importing the necessary Python libraries and the dataset: import pandas as pdfrom scipy.stats import ttest_inddf = pd.read_csv("website_ab_test.csv")print(df.head()) Theme Click Through Rate Conversion Rate Bounce Rate \0 Light Theme 0.054920 0.282367 0.405085 1 Light Theme 0.113932 0.032973 0.732759 2 Dark Theme 0.323352 0.178763 0.296543 3 Light Theme 0.485836 0.325225 0.245001 4 Light Theme 0.034783 0.196766 0.765100 Scroll_Depth Age Location Session_Duration Purchases Added_to_Cart 0 72.489458 25 Chennai 1535 No Yes 1 61.858568 19 Pune 303 No Yes 2 45.737376 47 Chennai 563 Yes Yes 3 76.305298 58 Pune 385 Yes No 4 48.927407 25 New Delhi 1437 No No So, the dataset is based on the performance of two themes on a website. Our task is to find which theme performs better using Hypothesis Testing. Lets go through the summary of the dataset, including the number of records, the presence of missing values, and basic statistics for the numerical columns: # dataset summarysummary = { 'Number of Records': df.shape[0], 'Number of Columns': df.shape[1], 'Missing Values': df.isnull().sum(), 'Numerical Columns Summary': df.describe()}summary {'Number of Records': 1000, 'Number of Columns': 10, 'Missing Values': Theme 0 Click Through Rate 0 Conversion Rate 0 Bounce Rate 0 Scroll_Depth 0 Age 0 Location 0 Session_Duration 0 Purchases 0 Added_to_Cart 0 dtype: int64, 'Numerical Columns Summary': Click Through Rate Conversion Rate Bounce Rate Scroll_Depth \ count 1000.000000 1000.000000 1000.000000 1000.000000 mean 0.256048 0.253312 0.505758 50.319494 std 0.139265 0.139092 0.172195 16.895269 min 0.010767 0.010881 0.200720 20.011738 25% 0.140794 0.131564 0.353609 35.655167 50% 0.253715 0.252823 0.514049 51.130712 75% 0.370674 0.373040 0.648557 64.666258 max 0.499989 0.498916 0.799658 79.997108 Age Session_Duration count 1000.000000 1000.000000 mean 41.528000 924.999000 std 14.114334 508.231723 min 18.000000 38.000000 25% 29.000000 466.500000 50% 42.000000 931.000000 75% 54.000000 1375.250000 max 65.000000 1797.000000 } The dataset contains 1,000 records across 10 columns, with no missing values. Heres a quick summary of the numerical columns: Click Through Rate: Ranges from about 0.01 to 0.50 with a mean of approximately 0.26. Conversion Rate: Also ranges from about 0.01 to 0.50 with a mean close to the Click Through Rate, approximately 0.25. Bounce Rate: Varies between 0.20 and 0.80, with a mean around 0.51. Scroll Depth: Shows a spread from 20.01 to nearly 80, with a mean of 50.32. Age: The age of users ranges from 18 to 65 years, with a mean age of about 41.5 years. Session Duration: This varies widely from 38 seconds to nearly 1800 seconds (30 minutes), with a mean session duration of approximately 925 seconds (about 15 minutes). Now, lets move on to comparing the performance of both themes based on the provided metrics. Well look into the average Click Through Rate, Conversion Rate, Bounce Rate, and other relevant metrics for each theme. Afterwards, we can perform hypothesis testing to identify if theres a statistically significant difference between the themes: # grouping data by theme and calculating mean values for the metricstheme_performance = df.groupby('Theme').mean()# sorting the data by conversion rate for a better comparisontheme_performance_sorted = theme_performance.sort_values(by='Conversion Rate', ascending=False)print(theme_performance_sorted) Click Through Rate Conversion Rate Bounce Rate Scroll_Depth \Theme Light Theme 0.247109 0.255459 0.499035 50.735232 Dark Theme 0.264501 0.251282 0.512115 49.926404 Age Session_Duration Theme Light Theme 41.734568 930.833333 Dark Theme 41.332685 919.482490 The comparison between the Light Theme and Dark Theme on average performance metrics reveals the following insights: Click Through Rate (CTR): The Dark Theme has a slightly higher average CTR (0.2645) compared to the Light Theme (0.2471). Conversion Rate: The Light Theme leads with a marginally higher average Conversion Rate (0.2555) compared to the Dark Theme (0.2513). Bounce Rate: The Bounce Rate is slightly higher for the Dark Theme (0.5121) than for the Light Theme (0.4990). Scroll Depth: Users on the Light Theme scroll slightly further on average (50.74%) compared to those on the Dark Theme (49.93%). Age: The average age of users is similar across themes, with the Light Theme at approximately 41.73 years and the Dark Theme at 41.33 years. Session Duration: The average session duration is slightly longer for users on the Light Theme (930.83 seconds) than for those on the Dark Theme (919.48 seconds). From these insights, it appears that the Light Theme slightly outperforms the Dark Theme in terms of Conversion Rate, Bounce Rate, Scroll Depth, and Session Duration, while the Dark Theme leads in Click Through Rate. However, the differences are relatively minor across all metrics. Well use a significance level (alpha) of 0.05 for our hypothesis testing. It means well consider a result statistically significant if the p-value from our test is less than 0.05. Lets start with hypothesis testing based on the Conversion Rate between the Light Theme and Dark Theme. Our hypotheses are as follows: Null Hypothesis (H0):There is no difference in Conversion Rates between the Light Theme and Dark Theme. Alternative Hypothesis (Ha):There is a difference in Conversion Rates between the Light Theme and Dark Theme. Well use a two-sample t-test to compare the means of the two independent samples. Lets proceed with the test: # extracting conversion rates for both themesconversion_rates_light = df[df['Theme'] == 'Light Theme']['Conversion Rate']conversion_rates_dark = df[df['Theme'] == 'Dark Theme']['Conversion Rate']# performing a two-sample t-testt_stat, p_value = ttest_ind(conversion_rates_light, conversion_rates_dark, equal_var=False)t_stat, p_value (0.4748494462782632, 0.6349982678451778) The result of the two-sample t-test gives a p-value of approximately 0.635. Since this p-value is much greater than our significance level of 0.05, we do not have enough evidence to reject the null hypothesis. Therefore, we conclude that there is no statistically significant difference in Conversion Rates between the Light Theme and Dark Theme based on the data provided. Now, lets conduct hypothesis testing based on the Click Through Rate (CTR) to see if theres a statistically significant difference between the Light Theme and Dark Theme regarding how often users click through. Our hypotheses remain structured similarly: Null Hypothesis (H0):There is no difference in Click Through Rates between the Light Theme and Dark Theme. Alternative Hypothesis (Ha):There is a difference in Click Rates between the Light Theme and Dark Theme. Well perform a two-sample t-test on the CTR for both themes. Lets proceed with the calculation: # extracting click through rates for both themesctr_light = df[df['Theme'] == 'Light Theme']['Click Through Rate']ctr_dark = df[df['Theme'] == 'Dark Theme']['Click Through Rate']# performing a two-sample t-testt_stat_ctr, p_value_ctr = ttest_ind(ctr_light, ctr_dark, equal_var=False)t_stat_ctr, p_value_ctr (-1.9781708664172253, 0.04818435371010704) The two-sample t-test for the Click Through Rate (CTR) between the Light Theme and Dark Theme yields a p-value of approximately 0.048. This p-value is slightly below our significance level of 0.05, indicating that there is a statistically significant difference in Click Through Rates between the Light Theme and Dark Theme, with the Dark Theme likely having a higher CTR given the direction of the test statistic. Now, lets perform Hypothesis Testing based on two other metrics: bounce rate and scroll depth, which are important metrics for analyzing the performance of a theme or a design on a website. Ill first perform these statistical tests and then create a table to show the report of all the tests we have done: # extracting bounce rates for both themesbounce_rates_light = df[df['Theme'] == 'Light Theme']['Bounce Rate']bounce_rates_dark = df[df['Theme'] == 'Dark Theme']['Bounce Rate']# performing a two-sample t-test for bounce ratet_stat_bounce, p_value_bounce = ttest_ind(bounce_rates_light, bounce_rates_dark, equal_var=False)# extracting scroll depths for both themesscroll_depth_light = df[df['Theme'] == 'Light Theme']['Scroll_Depth']scroll_depth_dark = df[df['Theme'] == 'Dark Theme']['Scroll_Depth']# performing a two-sample t-test for scroll deptht_stat_scroll, p_value_scroll = ttest_ind(scroll_depth_light, scroll_depth_dark, equal_var=False)# creating a table for comparisoncomparison_table = pd.DataFrame({ 'Metric': ['Click Through Rate', 'Conversion Rate', 'Bounce Rate', 'Scroll Depth'], 'T-Statistic': [t_stat_ctr, t_stat, t_stat_bounce, t_stat_scroll], 'P-Value': [p_value_ctr, p_value, p_value_bounce, p_value_scroll]})comparison_table So, heres a table comparing the performance of the Light Theme and Dark Theme across various metrics based on hypothesis testing: Click Through Rate: The test reveals a statistically significant difference, with the Dark Theme likely performing better (P-Value = 0.048). Conversion Rate: No statistically significant difference was found (P-Value = 0.635). Bounce Rate: Theres no statistically significant difference in Bounce Rates between the themes (P-Value = 0.230). Scroll Depth: Similarly, no statistically significant difference is observed in Scroll Depths (P-Value = 0.450). In summary, while the two themes perform similarly across most metrics, the Dark Theme has a slight edge in terms of engaging users to click through. For other key performance indicators like Conversion Rate, Bounce Rate, and Scroll Depth, the choice between a Light Theme and a Dark Theme does not significantly affect user behaviour according to the data provided. So, Hypothesis Testing is a statistical method used to make inferences or decisions about a population based on sample data. It starts with a null hypothesis (H0), which represents a default stance or no effect, and an alternative hypothesis (H1 or Ha), which represents what we aim to prove or expect to find. The process involves using sample data to determine whether to reject the null hypothesis in favor of the alternative hypothesis, based on the likelihood of observing the sample data under the null hypothesis. I hope you liked this article on Hypothesis Testing using Python. Feel free to ask valuable questions in the comments section below. You can follow me onInstagramfor many more resources. Hypotheses are claims, and we can use statistics to prove or disprove them. At this point, hypothesis testing structures the problems so that we can use statistical evidence to test these claims. So we can check whether or not the claim is valid.In this article, I want to show hypothesis testing with Python on several questions step-by-step. But before, let me explain the hypothesis testing process briefly. If you wish, you can move to the questions directly. First of all, we should understand which scientific question we are looking for an answer to, and it should be formulated in the form of the Null Hypothesis (H) and the Alternative Hypothesis (H or H). Please remember that H and H must be mutually exclusive, and H shouldnt contain equality:H: =x, H: xH: x, H: >xH: x, H: H: The data is normally distributed.H: The data is not normally distributed.Assume that =0.05. If the p-value is >0.05, it can be said that data is normally distributed.For checking normality, I used Shapiro-Wilks W test which is generally preferred for smaller samples however there are other options like Kolmogorov-Smirnov and DAgostino and Pearsons test. Please visit for more information. p value:0.6556Fail to reject null hypothesis >> The data is normally distributedp value:0.0803Fail to reject null hypothesis >> The data is normally distributedH: The variances of the samples are the same.H: The variances of the samples are different. It tests the null hypothesis that the population variances are equal (called homogeneity of variance or homoscedasticity). Suppose the resulting p-value of Levenes test is less than the significance level (typically 0.05). In that case, the obtained differences in sample variances are unlikely to have occurred based on random sampling from a population with equal variances.For checking variance homogeneity, I preferred Levenes test but you can also check Bartletts test from here: p value:0.8149Fail to reject null hypothesis >> The variances of the samples the are same. Since assumptions are satisfied, we can perform the parametric version of the test for 2 groups and unpaired data. p value:0.00753598since the hypothesis is one sided >> use p_value/2 >> p_value_one_sided:0.0038Reject null hypothesis At this significance level, there is enough evidence to conclude that the average grade of the students who follow the course synchronously is higher than the students who follow the course asynchronously. Photo by Christian Bowen on UnsplashA pediatrician wants to see the effect of formula consumption on the average monthly weight gain (in gr) of babies. For this reason, she collected data from three different groups. The first group is exclusively breastfed children (receives only breast milk), the second group is children who are fed with only formula and the last group is both formula and breastfed children. These data are as below.only_breast=[794.1, 716.9, 993. , 724.7, 760.9, 908.2, 659.3 , 690.8, 768.7, 717.3 , 630.7, 729.5, 714.1, 810.3, 583.5, 679.9, 865.1]only_formula=[898.8, 881.2, 940.2, 966.2, 957.5, 1061.7, 1046.2, 980.4, 895.6, 919.7, 1074.1, 952.5, 796.3, 859.6, 871.1 , 1047.5, 919.1 , 1160.5, 996.9]both=[976.4, 656.4, 861.2, 706.8, 718.5, 717.1, 759.8, 894.6, 867.6, 805.6, 765.4, 800.3, 789.9, 875.3, 740. , 799.4, 790.3, 795.2 , 823.6, 818.7, 926.8, 791.7, 948.3]According to this information, conduct the hypothesis testing to check whether there is a difference between the average monthly gain of these three groups by using a 0.05 significance level. If there is a significant difference, perform further analysis to find what caused the difference. Before doing hypothesis testing, check the related assumptions. H: == or The mean of the samples is the same.H: At least one of them is different. H: The data is normally distributed.H: The data is not normally distributed.H: The variances of the samples are the same.H: The variances of the samples are different. p value:0.4694Fail to reject null hypothesis >> The data is normally distributedp value:0.8879Fail to reject null hypothesis >> The data is normally distributedp value:0.7973Fail to reject null hypothesis >> The data is normally distributedp value:0.7673Fail to reject null hypothesis >> The variances of the samples are same. Since assumptions are satisfied, we can perform the parametric version of the test for more than 2 groups and unpaired data. p value:0.000000Reject null hypothesis At this significance level, it can be concluded that at least one of the groups has a different average monthly weight gain. To find which group or groups cause the difference, we need to perform a posthoc test/pairwise comparison as below.Note: To avoid family-wise p-value inflation, I used Bonferroni adjustment. You can see your other alternative from here: At this significance level, it can be concluded that:"only breast" is different than "only formula""only formula" is different than both "only breast" and "both""both" is different than "only formula" Photo by Alex Kotliarskyi on UnsplashA human resource specialist working in a technology company is interested in the overwork time of different teams. To investigate whether there is a difference between overtime of the software development team and the test team, she selected 17 employees randomly in each of the two teams and recorded their weekly average overwork time in terms of an hour. The data is below.test_team=[6.2, 7.1, 1.5, 2,3 , 2, 1.5, 6.1, 2.4, 2.3, 12.4, 1.8, 5.3, 3.1, 9.4, 2.3, 4.1]developer_team=[2.3, 2.1, 1.4, 2.0, 8.7, 2.2, 3.1, 4.2, 3.6, 2.5, 3.1, 6.2, 12.1, 3.9, 2.2, 1.2 ,3.4]According to this information, conduct the hypothesis testing to check whether there is a difference between the overwork time of two teams by using a 0.05 significance level. Before doing hypothesis testing, check the related assumptions. H: H: > H: The data is normally distributed.H: The data is not normally distributed.H: The variances of the samples are same.H: The variances of the samples are different. p value:0.0046Reject null hypothesis >> The data is not normally distributedp value:0.0005Reject null hypothesis >> The data is not normally distributedp value:0.5410Fail to reject null hypothesis >> The variances of the samples are same. There are two groups, and data is collected from different individuals, so it is not paired. However, the normality assumption is not satisfied; therefore, we need to use the nonparametric version of 2 group comparison for unpaired data: the Mann-Whitney U Test. p-value:0.8226Fail to recejt null hypothesisAt this significance level, it can be said that there is no statistically significant difference between the average overwork time of the two teams. Photo by Aman Pal on UnsplashAn e-commerce company regularly advertises on YouTube, Instagram, and Facebook for its campaigns. However, the new manager was curious about if there was any difference between the number of customers attracted by these platforms. Therefore, she started to use Adjust, an application that allows you to find out where your users come from. The daily numbers reported from Adjust for each platform are as below.Youtube=[1913, 1879, 1939, 2146, 2040, 2127, 2122, 2156, 2036, 1974, 1956, 2146, 2151, 1943, 2125]Instagram = [2305., 2355., 2203., 2231., 2185., 2420., 2386., 2410., 2340., 2349., 2241., 2396., 2244., 2267., 2281.]Facebook = [2133., 2522., 2124., 2551., 2293., 2367., 2460., 2311., 2178., 2113., 2048., 2443., 2265., 2095., 2528.]According to this information, conduct the hypothesis testing to check whether there is a difference between the average customer acquisition of these three platforms using a 0.05 significance level. If there is a significant difference, perform further analysis to find that caused the difference. Before doing hypothesis testing, check the related assumptions. H: == or The mean of the samples is the same.H: At least one of them is different. H: The data is normally distributed.H: The data is not normally distributed.H: The variances of the samples are the same.H: The variances of the samples are different. p value:0.0285Reject null hypothesis >> The data is not normally distributedp value:0.4156Fail to reject null hypothesis >> The data is normally distributedp value:0.1716Fail to reject null hypothesis >> The data is normally distributedp value:0.0012Reject null hypothesis >> The variances of the samples are different. The normality and variance homogeneity assumptions are not satisfied, therefore we need to use the nonparametric version of ANOVA for unpaired data (the data is collected from different sources). p value:0.000015Reject null hypothesisAt this significance level, at least one of the average customer acquisition number is different.Note: Since the data is not normal, the nonparametric version of posthoc test is used. The average number of customers coming from YouTube is different than the other (actually smaller than the others). Photo by Brooke Lark on UnsplashThe University Health Center diagnosed eighteen students with high cholesterol in the previous semester. Healthcare personnel told these patients about the dangers of high cholesterol and prescribed a diet program. One month later, the patients came for control, and their cholesterol level was reexamined. Test whether there is a difference in the cholesterol levels of the patients.According to this information, conduct the hypothesis testing to check whether there is a decrease in the cholesterol levels of the patients after the diet by using a 0.05 significance level. Before doing hypothesis testing, check the related assumptions. Comment on the resultstest_results_before_diet=[224, 235, 223, 253, 253, 224, 244, 225, 259, 220, 242, 240, 239, 229, 276, 254, 237, 227]test_results_after_diet=[198, 195, 213, 190, 246, 206, 225, 199, 214, 210, 188, 205, 200, 220, 190, 199, 191, 218] H: d>=0 or The true mean difference is equal to or bigger than zero.H: d> The data is normally distributedp value:0.1003Fail to reject null hypothesis >> The data is normally distributed The data is paired since data is collected from the same individuals and assumptions are satisfied, then we can use the dependent t-test. p value:0.000008 one tailed p value:0.000004Reject null hypothesis At this significance level, there is enough evidence to conclude mean cholesterol level of patients has decreased after the diet. A venture capitalist wanted to invest in a startup that provides data compression without any loss in quality, but there are two competitors: PiedPiper and EndFrame. Initially, she believed the performance of the EndFrame could be better but still wanted to test it before the investment. Then, she gave the same files to each company to compress and recorded their performance scores. The data is below.piedpiper=[4.57, 4.55, 5.47, 4.67, 5.41, 5.55, 5.53, 5.63, 3.86, 3.97, 5.44, 3.93, 5.31, 5.17, 4.39, 4.28, 5.25]endframe = [4.27, 3.93, 4.01, 4.07, 3.87, 4. , 4. , 3.72, 4.16, 4.1 , 3.9 , 3.97, 4.08, 3.96, 3.96, 3.77, 4.09]According to this information, conduct the related hypothesis testing by using a 0.05 significance level. Before doing hypothesis testing, check the related assumptions. Comment on the results. Since the performance scores are obtained from the same files, the data is paired.H: d>=0 or The true mean difference is equal to or bigger than zero.H: d> The data is not normally distributedp value:0.9587Fail to reject null hypothesis >> The data is normally distributed The normality assumption is not satisfied; therefore, we need to use the nonparametric version of the paired test, namely the Wilcoxon Signed Rank test. p-value:0.000214 >> one_tailed_pval:0.000107one sided pvalue:0.000107Reject null hypothesisAt this significance level, there is enough evidence to conclude that the performance of the PiedPaper is better than the EndFrame. Photo by Kelly Sikkema on UnsplashA researcher was curious about whether there is a difference between the methodology she developed, C, and baseline methods A and B in terms of performance. Therefore, she decided to design different experiments and recorded the achieved accuracy by each method. The below table shows the achieved accuracy on test sets by each method. Please note that the same train and test sets were used for each method.According to this information, conduct the hypothesis testing to check whether there is a difference between the performance of the methods by using a 0.05 significance level. If there is a significant difference, perform further analysis to find which one caused the difference. Before doing hypothesis testing, check the related assumptions. Comment on the results. H: == or The mean of the samples is the same.H: At least one of them is different. H: The data is normally distributed.H: The data is not normally distributed.H: The variances of the samples are the same.H: The variances of the samples are different. p value:0.3076Fail to reject null hypothesis >> The data is normally distributedp value:0.0515Fail to reject null hypothesis >> The data is normally distributedp value:0.0016Reject null hypothesis >> The data is not normally distributedp value:0.1953Fail to reject null hypothesis >> The variances of the samples are same. There are three groups, but the normality assumption is violated. So, we need to use the nonparametric version of ANOVA for paired data since the accuracy scores are obtained from the same test sets. p value:0.0015Reject null hypothesis89.35 89.49 90.49At this significance level, at least one of the methods has a different performance.Note: Since the data is not normal, the nonparametric version of the posthoc test is used. Method C outperformed others and achieved better accuracy scores than the others. Photo by janilson furtado on UnsplashAn analyst of a financial investment company is curious about the relationship between gender and risk appetite. A random sample was taken of 660 customers from the database. The customers in the sample were classified according to their gender and their risk appetite. The result is given in the following table.Test the hypothesis that the risk appetite of the customers in this company is independent of their gender. Use = 0.01. H: Gender and risk appetite are independent.H: Gender and risk appetite are dependent. chi2 test should be used for this question. This test is known as the goodness-of-fit test. It implies that if the observed data are very close to the expected data. The assumption of this test every Ei 5 (in at least 80% of the cells) is satisfied. expected frequencies: [[43.21 24.74 28.23 32.41 101.41] [80.79 46.26 52.77 60.59 189.59]]degrees of freedom: 4test stat :7.0942p value:0.1310 critical stat:13.2767Since the p-value is larger than =0.01 (or calculated statistic=7.14 is smaller than the critical statistic=13.28) Fail to Reject H. At this significance level, it can be concluded that gender and risk appetite are independent. You can visit to see the full implementation. Statistical hypothesis testing is a fundamental aspect of data analysis in many fields, especially in data science and analytics. It allows us to make inferences or draw conclusions about a population based on sample data. In this post, we will delve into the concepts of statistical hypothesis testing using Python, understand its principles, and apply them practically through a case study. A hypothesis is a statement or assumption about a population parameter. In statistical terms, we often differentiate between two types of hypotheses: Null Hypothesis (H0): This hypothesis states that there is no effect or difference, and it is the hypothesis that we seek to test. For example, a null hypothesis could state that the mean height of a population is equal to a certain value. Alternative Hypothesis (Ha): This hypothesis asserts that there is an effect or a difference. It is what we hope to support with evidence from our data. For instance, the alternative hypothesis might state that the mean height is greater than a certain value. The process of hypothesis testing generally involves the following steps: Formulate Hypotheses: Define the null and alternative hypotheses. Select Significance Level (): Choose the significance level, typically set at 0.05 or 0.01. Collect Data: Gather data relevant to the hypotheses. Conduct Statistical Test: Perform the appropriate statistical test to analyze the data. Make a Decision: Based on the test results, determine whether to reject or not reject the null hypothesis. Interpret the Results: Draw conclusions from the test result, including the p-value. The p-value is a critical concept in hypothesis testing, indicating the probability of observing the test results under the null hypothesis. A smaller p-value suggests that the observed data is inconsistent with the null hypothesis. Generally, if the p-value is less than the significance level (), we reject the null hypothesis. In this section, we will perform a hypothesis test using Python. We will use the SciPy library to conduct a t-test, which is useful for determining if there is a statistically significant difference between the means of two groups. Before we start coding, lets ensure that we have the necessary library installed. We will need the SciPy and NumPy libraries: pip install scipy numpy Lets consider a case study where we want to evaluate if a new teaching method improves student performance in mathematics. We will compare the exam scores from two groups of students: one that underwent the traditional teaching method (Group A) and another that was taught using the new method (Group B). In this case, our hypotheses would be:Null Hypothesis (H0): There is no difference in average exam scores between the two groups (Group A and Group B).Alternative Hypothesis (Ha): There is a significant difference in average exam scores between the two groups. We will set our significance level () at 0.05, which is a common threshold used in hypothesis testing. We will assume the following exam scores for both groups:Group A (Traditional): [78, 85, 88, 90, 76, 84, 92, 81, 79, 77]Group B (New Method): [82, 89, 91, 95, 86, 90, 93, 88, 87, 92] Now we will conduct a t-test using Python to compare the means of the two groups. Lets write the code: import numpy as npfrom scipy import stats# Data for both groupsgroup_a = np.array([78, 85, 88, 90, 76, 84, 92, 81, 79, 77])group_b = np.array([82, 89, 91, 95, 86, 90, 93, 88, 87, 92])# Perform t-testt_statistic, p_value = stats.ttest_ind(group_a, group_b)# Output resultsprint('T-statistic:', t_statistic)print('P-value:', p_value) Upon executing the code, we will receive the t-statistic and the p-value. Based on the p-value, we will determine whether to reject the null hypothesis: Assuming the output is:T-statistic: -1.258P-value: 0.239 Since the p-value (0.239) is greater than our significance level (0.05), we do not reject the null hypothesis. This suggests that there is not enough evidence to support the claim that the new teaching method significantly improved student performance. From our hypothesis testing process, we found that there is no statistically significant difference in exam scores between the two groups. This serves as an important insight for educators and policymakers as they decide about implementing new teaching methods. However, it is essential to consider other factors, such as sample size and context, before final conclusions. Statistical hypothesis testing is a crucial tool in data analysis that provides a structured approach to making informed decisions based on data. Through our case study, we demonstrated how to perform hypothesis testing using Python effectively. Remember, the choice of tests and interpretation of results should be done considering the context of the data and the specific questions at hand. By familiarizing yourself with various hypothesis tests and their applications, you can enhance your analytical skills and contribute to meaningful insights in your field. A hypothesis test is a formal statistical test we use to reject or fail to reject some statistical hypothesis.This tutorial explains how to perform the following hypothesis tests in Python:One sample t-testTwo sample t-testPaired samples t-testLets jump in!Example 1: One Sample t-test in PythonA one sample t-test is used to test whether or not the mean of a population is equal to some value.For example, suppose we want to know whether or not the mean weight of a certain species of some turtle is equal to 310 pounds.To test this, we go out and collect a simple random sample of turtles with the following weights:Weights: 300, 315, 320, 311, 314, 309, 300, 308, 305, 303, 305, 301, 303The following code shows how to use the ttest_1samp() function from the scipy.stats library to perform a one sample t-test:import scipy.stats as stats #define data data = [300, 315, 320, 311, 314, 309, 300, 308, 305, 303, 305, 301, 303] #perform one sample t-test stats.ttest_1samp(a=data, popmean=310) Ttest_1sampResult(statistic=-1.5848116313861254, pvalue=0.1389944275158753) The t test statistic is-1.5848 and the corresponding two-sided p-value is0.1389.The two hypotheses for this particular one sample t-test are as follows:H0: = 310 (the mean weight for this species of turtle is 310 pounds)HA: 310 (the mean weight is not310 pounds)Because the p-value of our test (0.1389) is greater than alpha = 0.05, we fail to reject the null hypothesis of the test.We do not have sufficient evidence to say that the mean weight for this particular species of turtle is different from 310 pounds.Example 2: Two Sample t-test in PythonA two sample t-test is used to test whether or not the means of two populations are equal.For example, suppose we want to know whether or not the mean weight between two different species of turtles is equal.To test this, we collect a simple random sample of turtles from each species with the following weights:Sample 1: 300, 315, 320, 311, 314, 309, 300, 308, 305, 303, 305, 301, 303Sample 2: 335, 329, 322, 321, 324, 319, 304, 308, 305, 311, 307, 300, 305The following code shows how to use the ttest_ind() function from the scipy.stats library to perform this two sample t-test:import scipy.stats as stats #define array of turtle weights for each sample sample1 = [300, 315, 320, 311, 314, 309, 300, 308, 305, 303, 305, 301, 303] sample2 = [335, 329, 322, 321, 324, 319, 304, 308, 305, 311, 307, 300, 305] #perform two sample t-test stats.ttest_ind(a=sample1, b=sample2) Ttest_indResult(statistic=-2.1009029257555696, pvalue=0.04633501389516516) The t test statistic is 2.1009 and the corresponding two-sided p-value is 0.0463.The two hypotheses for this particular two sample t-test are as follows:H0:1 = 2 (the mean weight between the two species is equal)HA:1 2 (the mean weight between the two species is not equal)Since the p-value of the test (0.0463) is less than .05, we reject the null hypothesis.This means we have sufficient evidence to say that the mean weight between the two species is not equal.Example 3: Paired Samples t-test in PythonA paired samples t-test is used to compare the means of two samples when each observation in one sample can be paired with an observation in the other sample.For example, suppose we want to know whether or not a certain training program is able to increase the max vertical jump (in inches) of basketball players.To test this, we may recruit a simple random sample of 12 college basketball players and measure each of their max vertical jumps. Then, we may have each player use the training program for one month and then measure their max vertical jump again at the end of the month.The following data shows the max jump height (in inches) before and after using the training program for each player:Before: 22, 24, 20, 19, 19, 20, 22, 25, 24, 23, 22, 21After: 23, 25, 20, 24, 18, 22, 23, 28, 24, 25, 24, 20The following code shows how to use the ttest_rel() function from the scipy.stats library to perform this paired samples t-test:import scipy.stats as stats #define before and after max jump heights before = [22, 24, 20, 19, 19, 20, 22, 25, 24, 23, 22, 21] after = [23, 25, 20, 24, 18, 22, 23, 28, 24, 25, 24, 20] #perform paired samples t-test stats.ttest_rel(a=before, b=after) Ttest_relResult(statistic=-2.5289026942943655, pvalue=0.02802807458682508) The t test statistic is 2.5289and the corresponding two-sided p-value is 0.0280.The two hypotheses for this particular paired samples t-test are as follows:H0:1 = 2 (the mean jump height before and after using the program is equal)HA:1 2 (the mean jump height before and after using the program is not equal)Since the p-value of the test (0.0280) is less than .05, we reject the null hypothesis.This means we have sufficient evidence to say that the mean jump height before and after using the training program is not equal.Additional ResourcesYou can use the following online calculators to automatically perform various t-tests:One Sample t-test CalculatorTwo Sample t-test CalculatorPaired Samples t-test Calculator Hypothesis testing is a statistical method used to evaluate whether a hypothesis about a population parameter is true or false based on a sample. It is an essential tool in statistics and data analysis. In this article, we will discuss the concept of hypothesis testing, its importance, and how to perform it using Python. Hypothesis testing involves making an assumption, called the null hypothesis (H0), about a population parameter, and then using data from a sample to either reject or fail to reject the null hypothesis. The alternative hypothesis (H1) is the opposite of the null hypothesis. The goal is to determine which hypothesis is more likely to be true based on the evidence from the sample. The hypothesis testing process can be mathematically represented as: Given a sample X1, X2, , Xn from the population, we can calculate the test statistic, which is a measure of how far the sample is from the hypothesized population parameter. We compare the test statistic to the critical value(s) from the distribution under the null hypothesis. If the test statistic falls in the critical region, we reject the null hypothesis, otherwise, we fail to reject it. Python provides several libraries to perform hypothesis testing, including SciPy, Statsmodels, and ScipyStats. In this example, we will use ScipyStats. Suppose we have a sample of 10 measurements for a population mean, and we want to test whether the population mean is greater than 50. We will use a one-tailed test. import numpy as npfrom scipy.stats import t# Sample datasample = np.array([52.1, 53.2, 55.3, 56.4, 57.5, 58.6, 59.7, 60.1, 61.2, 62.3])# Population mean hypothesismu0 = 50# Degrees of freedomdf = len(sample) - 1# Calculate test statistict_stat = np.mean(sample) - mu0t_value = t.ppf(1 - 0.05, df)# Reject null hypothesis if test statistic > critical valueif abs(t_stat) > t_value: print("Reject null hypothesis. The population mean is likely greater than 50.")else: print("Fail to reject null hypothesis. The population mean may or may not be greater than 50.") In some cases, we may want to test whether a population parameter is significantly different from a hypothesized value in either direction. This is called a two-tailed test. import numpy as npfrom scipy.stats import t# Sample datasample = np.array([52.1, 53.2, 55.3, 56.4, 57.5, 58.6, 59.7, 60.1, 61.2, 62.3])# Population mean hypothesismu0 = 50# Degrees of freedomdf = len(sample) - 1# Calculate test statistic and critical valuest_stat = np.abs(np.mean(sample) - mu0)t_critical = t.ppf(0.025, df)# Reject null hypothesis if test statistic > critical valueif t_stat > t_critical: print("Reject null hypothesis. The population mean is significantly different from 50.")else: print("Fail to reject null hypothesis. The population mean may or may not be significantly different from 50.") For the one-tailed test example, the output would look like: Reject null hypothesis. The population mean is likely greater than 50. For the two-tailed test example, the output would look like: Reject null hypothesis. The population mean is significantly different from 50.Hypothesis Testing with PythonData is the basis for all decision-makingindatascience, and with hypothesis testing, you can optimize these decisions.Hypothesis testing helps data scientists to confirm if their findings, theories, and assumptions occurred by chance or sound statistical analysis.Many industries and domains, such as pharmaceuticals, manufacturing, insurance, and research institutions, use hypothesis testing to make better predictions and draw more accurate conclusions from datasets.So, as a data scientist, its vital to familiarize yourself with this powerful statistical tool.In this article, youll learn the following:What is a hypothesis?Parts of a hypothesisWhat is hypothesis testing?Hypothesis testing outcomes: Type I and Type II errorsT-Test: Types and how to perform it with PythonZ-Test: Types and how to perform it with PythonSimilarities and differences between T-Test and Z-TestPrerequisite: This article assumes you are familiar with statistical terms. If not, kindly read this article before youproceed.What is a Hypothesis?Photo by AuthorBased on what you know, what do you think will happen? Thats the question a hypothesis answers.A hypothesis is a prediction or an assumption made based on existing knowledge. Its not a random guess or a question, but an inferential statement that introduces a question and proposes an answer.For example, the statement men suffer from diabetics more than women is a hypothesis. A good hypothesis is specific, testable, and amendable for further research that will be performed on the collected data.A hypothesis has two (2) variables: the independent variable (cause) and the dependent variable (effect). The independent variable affects what happens to the dependent variable.To know if our example hypothesis (men suffer from diabetics more than women) is true, it has to be tested. This can be done by running a survey on people who have suffered from diabetes. The collected data is then analyzed with the help of hypothesis testing.What are the Parts of a Hypothesis?There are two parts of a hypothesis:Null Hypothesis or H: A null hypothesis assumes that there are no effects or differences between the variables of two or more samples. For example, age has no impact on health status is a null hypothesis. This means that, whether you are aged 17 or 89, your health status is not affected by the number of years you have lived. H is always the hypothesis that is tested and then accepted or rejected.Alternative Hypothesis or H: An alternative hypothesis assumes that theres indeed an effect and difference between the variables of two or more samples. For example, age impacts health status is an alternative hypothesis.Some Important Terms Associated with Hypothesis TestingBelow are some terms that often mentioned alongside hypothesis testing:Level of Significance: Also known as the significance level or threshold level, it defines the distance the sample statistic parameter must be from the null hypothesis (H) to be considered statistically significant. In essence, its the probability of making the wrong decision when H is true. The level of significance is also the threshold value that we measure the p-value against. Its denoted by and is usually set at 5% (0.05). Meanwhile, its always determined before testing the hypothesis.P-value: This is the probability of the observed result obtained from testing the hypothesis. It gauges the power of evidence against the null hypothesis(H). If the p-value is less than the level of significance, then H is true, so we reject H. If the p-value is greater than the level of significance, then H is true, so we accept H.Confidence Level: Confidence level is the percentage of probability that the confidence interval will contain the true population parameter when you draw a random sample from the population many times. For example, if we have a population of sufferers of diabetics, the confidence level is the probability that no matter how many times we draw random samples from the population, there will always be a higher percentage of men suffering from diabetes. The formula for confidence level is 1- , where is the level of significance set at 5%. Hence, the confidence level is 95%.Confidence Interval: When we draw random samples from the population of sufferers of diabetes, we will get different values, depending on the variation between the variables in the population. For example, we could get 49%, 46%, 52%, and 44% from four different groups because there is always sampling error. So, it is always good practice to express the population estimate as a range of numbers. This range is the confidence interval. The confidence interval is a set of numbers defined by the confidence level that shows the probability that a particular population parameter will fall between a set of values. From the exampleabove, the confidence interval is 44%52%.Confidence level can be zero, but level of significance cant be zero.What is Hypothesis Testing?Hypothesis testing is performed to approve or disapprove a statement made about a sample drawn from a population. A hypothesis test verifies the credibility of the hypothesis and the likelihood that a finding observed from the sample data occurred by chance. We can use Python programming language to test hypotheses.To perform a hypothesis testing, you have to:Specify the null and alternative hypotheses.Set the significance level or which is usually 5% (0.05).Collect and analyze data.Calculate the test statistic and corresponding p-value.Draw conclusions.Hypothesis Testing Outcomes: Type I and Type II errorsType I error or -error occurs when the null hypothesis (H) is true but it is rejected.Type II error or -error occurs when the null hypothesis (H) is false but it is accepted.Both errors can not simultaneously be zero. If one of them is zero, then the other automatically becomes one. However, this doesnt mean that + = 1. In most cases, is set at 0.05, which is also the level of significance.What is T-Test in Hypothesis Testing?T-test is used to compare the mean of two different samples or groups when the sample size is 30 and the data follows a normal or Gaussian distribution, that is, the data is symmetrically distributed.There are three different types of T-tests. You decide which T-Test to use based on the nature of the data or design of the experiment.One Sample T-test: This test is used to compare the mean value of a sample with a known mean or average. The null and alternative hypotheses of this T-test are as follows:i. Null Hypothesis: = Mii. Alternative Hypothesis: MWhere is the sample mean and M is the known mean.2. Independent Sample T-test: This test is used to compare the mean values of two different samples from two different populations.i. Null Hypothesis: = ii. Alternative Hypothesis: Where and are the sample means for the two independent samples.3. Dependent or Paired sample T-test: This test is used to compare the means of two dependent samples taken from the same population at different points in time.i. Null Hypothesis: = 0ii. Alternative Hypothesis: 0Where and are the sample means for the two dependent samples.How to Perform T-test with Pythonimport numpy as npimport randomimport scipy.stats as stats# We will use random data for the samples and label them as 'a' and 'b' respectivelya = [random.gauss(50, 20) for x in range(30)]b = [random.gauss(55, 15) for x in range(30)]# One Sample T-test for sample 'a'# Null Hypothesis: The sample mean = population mean(45)t_stat, p_value = stats.ttest_1samp(a, 45, axis=0)print("T-statistic:", t_stat)print("P-value:", p_value)print(np.mean(a))# Interpretation of the results: The p-value is less than 0.05(signifance level) and the sample mean(appro. 53) is also population mean(45).# We reject the null hypothesisOne Sample T-test2. Independent Sample T-testimport numpy as npimport randomimport scipy.stats as statsa = [random.gauss(50,20) for x in range (30)]b = [random.gauss(55,15) for x in range (30)]#Independent sample test# Null hypothesis: mean of a = mean of bt_stat, p_value = stats.ttest_ind(a,b,equal_var = False)print("T-statistic:", t_stat)print("P-value:", p_value)print(np.mean(a), np.mean(b))# Interpretation of the results: The p-value is greater than 0.05(signifance level) and the mean of a = mean of b# We accept the null hypothesisIndependent Sample T-test3. Dependent or Paired Sample T-testimport numpy as npimport randomimport scipy.stats as statsa = [random.gauss(50,20) for x in range (30)]b = [random.gauss(55,15) for x in range (30)]#Paired sample t-test# Null hypothesis: mean of b - mean of a = 0a = [random.gauss(50,20) for x in range(30)]b = [random.gauss(60,25) for x in range(30)]t_stat,p_value = stats.ttest_rel(a,b)print("T-statistic:", t_stat)print("P-value:", p_value)print(np.mean(a) - np.mean(b))# Interpretation of the results: The p-value is less than 0.05(signifance level) and the mean of b - mean of a 0# We reject the null hypothesisPaired Sample T-testWhat is Z-Test in Hypothesis Testing?Z-test is used to compare the difference between a sample and a known population mean or between the mean of two independent samples. We use the z-test when the sample size is 30 and the population standard deviation (SD) is known.There are two different types of Z-test:One-sample Z-test: This test is used to compare one sample mean with a known population mean.i. Null Hypothesis: = ii. Alternative Hypothesis: Where is the population mean and is a specified value you compare with the population mean.2. Two-sample Z-test: This test is used to compare the mean values of two independent samples.i. Null Hypothesis: = ii. Alternative Hypothesis: Where and are the population means of the two independent samples you are comparing.How to Perform Z-test with PythonTo perform a Z-test inPython, you need to import the statsmodels library.import numpy as npimport randomfrom statsmodels.stats.weightstats import ztest as ztest# to reproduce the same random values random.seed(20)# We will use random data for the samplea = [random.gauss(100, 15) for x in range(40)] # where 100 is the mean(specified value) and 15 is the standard deviation# One Sample Z-test for sample 'a'# Null Hypothesis: Mean of a = 100 z_stat, p_value = ztest(a, value=100)print("Z-statistic:", z_stat)print("P-value:", p_value)print(np.mean(a))# Interpretation of the results: The p-value is greater than 0.05, and the mean of a is close to 100# We accept the null hypothesisOne Sample Z-Test2. Two Sample Z-Testimport numpy as npimport randomfrom statsmodels.stats.weightstats import ztest as ztest# to reproduce the same random values random.seed(20)# We will use random data as the sample# Null Hypothesis: Mean of a = mean of ba = [random.gauss(100, 15) for x in range(40)]b = [random.gauss(120, 15) for x in range(40)]# Two Sample Z-test for samples 'a' and 'b'# Null Hypothesis: Mean of a = Mean of 'b'z_stat, p_value = ztest(a, b, value=0)print("Z-statistic:", z_stat)print("P-value:", p_value)print(np.mean(a), np.mean(b))# Interpretation of the results: the p-value is less than 0.05 and mean of a mean of b# We reject the null hypothesisTwo Sample Z-TestSimilarities and Differences between T-Test and Z-TestBoth tests are used to compare the means of one or two samples. Also, the results obtained from both tests are usually very close when compared.DifferencesFor Z-test, the population mean and variance or standard deviation must be known. However, T-test doesnt need the population mean or standard deviation.Another difference between t-test and z-test is in their sample size. T-test doesnt require a large sample size, while Z-test works with a large sample size.Ultimately, both tests are vital in hypothesis testing.ConclusionHypothesis testing will help you to draw conclusions about a population from data samples, identify meaningful distinctions between groups or samples, and add credibility to your results. This guide should help you perform hypothesis testing with Python. Goodluck!If you found this article helpful, please clap, drop a comment, share with others, and follow me on LinkedInReferencesPhoto by Mikael BlomkvistStep into the intriguing world of hypothesis testing, where your natural curiosity meets the power of data to reveal truths!This article is your key to unlocking how those everyday huncheslike guessing a groups average income or figuring out who owns their homecan be thoroughly checked and proven with data.I am going to take you by the hand and show you, in simple steps, how to use Python to explore a hypothesis about the average yearly income.By the time were done, youll not only get the hang of creating and testing hypotheses but also how to use statistical tests on actual data.Perfect for up-and-coming data scientists, anyone with a knack for analysis, or just if youre keen on data, get ready to gain the skills to make informed decisions and turn insights into real-world actions.Join me as we dive deep into the data, one hypothesis at a time!Before we get started, elevate your data skills with my expert eBooksthe culmination of my experiences and insights.Support my work and enhance your journey. Check them out:My Gumroad Shop Checkout for more such resources: hypothesis is like a guess or prediction about something specific, such as the average income or the percentage of homeowners in a group of people.Its based on theories, past observations, or questions that spark our curiosity.For instance, you might predict that the average yearly income of potential customers is over $50,000 or that 60% of them own their homes.To see if your guess is right, you gather data from a smaller group within the larger population and check if the numbers (like the average income, percentage of homeowners, etc.) from this smaller group match your initial prediction.You also set a rule for how sure you need to be to trust your findings, often using a 5% chance of error as a standard measure. This means youre 95% confident in your results. Level of Significance (0.05)There are two main types of hypotheses: the null hypothesis, which is your baseline saying theres no change or difference, and the alternative hypothesis, which suggests there is a change or difference.For example,If you start with the idea that the average yearly income of potential customers is $50,000,The alternative could be that its not $50,000it could be less or more, depending on what youre trying to find out.To test your hypothesis, you calculate a test statistica number that shows how much your sample data deviates from what you predicted.How you calculate this depends on what youre studying and the kind of data you have. For example, to check an average, you might use a formula that considers your samples average, the predicted average, the variation in your sample data, and how big your sample is.This test statistic follows a known distribution (like the t-distribution or z-distribution), which helps you figure out the p-value.The p-value tells you the odds of seeing a test statistic as extreme as yours if your initial guess was correct.A small p-value means your data strongly disagrees with your initial guess.Finally, you decide on your hypothesis by comparing the p-value to your error threshold.If the p-value is smaller or equal, you reject the null hypothesis, meaning your data shows a significant difference thats unlikely due to chance.If the p-value is larger, you stick with the null hypothesis, suggesting your data doesnt show a meaningful difference and any change might just be by chance.Well go through an example that tests if the average annual income of prospective customers exceeds $50,000.This process involves stating hypotheses, specifying a significance level, collecting and analyzing data, and drawing conclusions based on statistical tests.Null Hypothesis (H0): The average annual income of prospective customers is $50,000.Alternative Hypothesis (H1): The average annual income of prospective customers is more than $50,000.Significance Level: 0.05, meaning were 95% confident in our findings and allow a 5% chance of error.Well use theProspectiveBuyer table, assuming it's a random sample from the population.This table has 2,059 entries, representing prospective customers' annual incomes.In Python, we can use libraries like Pandas and Numpy to calculate the sample mean and standard deviation.import pandas as pdimport numpy as npdf = pd.read_csv('ProspectiveBuyer.csv')sample_mean = df['YearlyIncome'].mean()sample_sd = df['YearlyIncome'].std()sample_size = len(df)print(f"Sample Mean: {sample_mean}")print(f"Sample Standard Deviation: {sample_sd}")print(f"Sample Size: {sample_size}")SampleMean: 56,992.43SampleSD: 32,079.16SampleSize: 2,059We use the t-test formula to calculate how significantly our sample mean deviates from the hypothesized mean.Pythons Scipy library can handle this calculation:from scipy import stats# Hypothesized meanmu = 50000t_statistic, p_value = stats.ttest_1samp(df['YearlyIncome'], mu)print(f"T-Statistic: {t_statistic}")Result:T-Statistic: 4.62The p-value is already calculated in the previous step using Scipy's ttest_1samp function, which returns both the test statistic and the p-value.print(f"P-Value: {p_value/2}") #specific to one-tailed testsResult:P-Value = 0.0000021We compare the p-value with our significance level to decide on our hypothesis:Since the p-value is less than 0.05, we reject the null hypothesis in favor of the alternative.Conclusion:Theres strong evidence to suggest that the average annual income of prospective customers is indeed more than $50,000.This example illustrates how Python can be a powerful tool for hypothesis testing, enabling us to derive insights from data through statistical analysis.Choosing the right test statistic is crucial and depends on what youre trying to find out, the kind of data you have, and how that data is spread out.Here are some common types of test statistics and when to use them:This ones great for checking out the average of a group when your data follows a normal distribution or when youre comparing the averages of two such groups.The t-test follows a special curve called the t-distribution. This curve looks a lot like the normal bell curve but with thicker ends, which means more chances for extreme values.The t-distributions shape changes based on something called degrees of freedom, which is a fancy way of talking about your sample size and how many groups youre comparing.Use this when youre looking at the average of a normally distributed group or the difference between two group averages, and you already know the standard deviation for all in the population.The z-test follows the standard normal distribution, which is your classic bell curve centered at zero and spreading out evenly on both sides.This is your go-to for checking if theres a difference in variability within a normally distributed group or if two categories are related.The chi-square statistic follows its own distribution, which leans to the right and gets its shape from the degrees of freedombasically, how many categories or groups youre comparing.This one helps you compare the variability between two groups or see if the averages of more than two groups are all the same, assuming all groups are normally distributed.The F-test follows the F-distribution, which is also right-skewed and has two types of degrees of freedom that depend on how many groups you have and the size of each group.In simple terms, the test you pick hinges on what youre curious about, whether your data fits the normal curve, and if you know certain specifics, like the populations standard deviation.Each test has its own special curve and rules based on your samples details and what youre comparing.Join my community of learners! Subscribe to my newsletter for more tips, tricks, and exclusive content on mastering Data Science & AI. Visit My Gumroad Shop: br>

Python hypothesis testing. Python hypothesis. Hypothesis test case python.