|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Methods** | | **Purpose** | | **Assumption** | **Hypothesis** | ​**SPSS Procedure** |
| **Parametric Date (assuming data from a normal distribution)** | | One sample t-test | | Whether the mean of a variable is less than, greater than, or equal to a specific value. Usually, the known value is a population mean. | The dependent variable is normally distributed | Null: There is no significant difference between the sample mean and the population mean.  Alternate: There is a significant difference between the sample mean and the population mean. | Go to **Analyze** menu, choose **Compare Means**, then **One-Sample T Test**. Move the dependent variable into the "**Test Variables**" box. Type in the value you wish to compare your sample to in the box called "**Test Value**." |
| Paired sample t-test | | Determine whether two population means are equal. It tests whether the difference between the two variables is significantly different from zero or not. | Both variables should be normally distributed. | Null: There is no significant difference between the means of the two variables.  Alternate: There is a significant difference between the means of the two variables. | Go to the **Analyze** menu, choose **Compare Means**, then choose **Paired Samples T Test**. Click on both variables you wish to compare, then put them into the **Paired Variables box**. |
| Independent Samples t-test | | Compares the mean difference between two independent groups on a given variable | 1.The dependent variable is normally distributed.   2. The two groups have approximately equal variance.   3. The two groups are independent of one another. | Null: The means of the two groups are not significantly different.  Alternate: The means of the two groups are significantly different. | Go to the **Analyze** menu, choose **Compare Means**, the **Independent Samples T Test**. Move your dependent variable into the box marked "**Test Variable**." Move your independent variable into the box marked "**Grouping Variable**." Click on the box marked "**Define Groups**" and specify the value labels of the two groups you wish to compare. |
| One-Way ANOVA | | Compares the mean of k groups based on one independent variable. | 1. The dependent variable is normally distributed.   2. The two groups have approximately equal variance on the dependent variable. | Null: There are no significant differences between the groups' mean scores.  Alternate: There is a significant difference between the groups' mean scores. | Go to the **Analyze** menu, choose **Compare Means**, then choose **One-Way ANOVA**. Move all dependent variables into the box labeled "**Dependent List**," and move the independent variable into the box labeled "**Factor**." Click on the button labeled "**Options**," and check off the boxes for **Descriptives and Homogeneity of Variance**. Click on the box marked "**Post Hoc**" and choose the appropriate post hoc comparison. |
| Two-way ANOVA | | Determines how a response is affected by two factors. | 1. The standard deviations (SD) of the populations for all groups are equal - this is sometimes referred to as an assumption of the homogeneity of variance.  2. The samples are randomly selected from the population | The null hypothesis is that there is no interaction between columns (data sets) and rows. More precisely, the null hypothesis states that any systematic differences between columns are the same for each row and that any systematic differences between rows are the same for each column. | Select **Analyze**, **General Linear Model**, **Univariate**; enter the **dependent variable** and the **independents** (factors); if you want to test interactions, click **Model** and select Custom, Model (Interaction) and enter interaction terms (ex. gender\*race); click **Plots** to set plot options; click **Options** to set what predicted group and interaction means are desired. |
| Pearson Regression | | Test magnitude and direction of the linear association between two variables that are on an interval or ratio scale. | Both variables are normally distributed. | Null: There is no association between the two variables.  Alternate: There is an association between the two variables. | Under the **Analyze** menu, choose **Correlations-Bivariate**. Move the variables you wish to correlate into the "**Variables**" box. Under the "**Correlation Coefficients**," be sure that the "**Pearson**" box is checked off. |
| Partial Regression | | Describe the linear relationship between two variables while controlling for the effects of one or more additional variables. | The Partial Correlations procedure assumes that each pair of variables is bivariate normal. | Null: There is no association between the two variables.  Alternate: There is an association between the two variables. | Under the **Analyze** menu, choose **Correlations-Partial**. Select two or more numeric variables for which partial correlations are to be computed into the "**Variables**" box. Select one or more numeric control variables into the “**controlling for**” box. |
| Simple Linear Regression | | Amount of variance accounted for by one variable in predicting another variable | 1. The data are linear   2. The dependent variable is normally distributed. | Null: The slope equals zero (there is no slope)  Alternate: The slope is not equal to zero | Under the **Analyze** menu, choose **Regression**, then choose **Linear**. Enter the dependent and independent variables in the appropriate places, then click OK. |
| **Nonparametric test (assuming data is not normally distributed)** | | Chi Square Goodness of Fit | | Determines if the observed frequencies are different from what we would expect to find. | 1.None of the expected values may be less than 1  2.No more than 20% of the expected values may be less than 5 | Null: There are approximately equal numbers of cases in each group  Alternate: There are not equal numbers of cases in each group | Go to **Analyze** menu, choose **Nonparametric Tests**, then **Chi Square**. Move the variable you wish to look at into the "**Test Variables**" box, then click OK. |
| Chi Square Test of Independence | | Determine the association between 2 categorical variables. | 1. None of the expected values may be less than 1  2. No more than 20% of the expected values may be less than 5 | Null: There is no association between the two variables.  Alternate: There is an association between the two variables. | Go to the **Analyze** menu, choose **Descriptive Statistics**, then choose **Crosstabs**. Move one variable into the box marked "**rows**" and the other into the box marked "**columns**." Under the "**Statistics**" button, be sure to check off **Chi Square**. Under the "**Cells**" button, be sure to check off. |
| Two independent- samples test | | Compares difference of two independent groups of cases on one variable. | 1. Random samples from populations  2. Independence within samples and mutual independence between samples  3. Measurement scale is at least ordinal | Null: The shapes of the two groups are not significantly different.  Alternate: The shapes of the two groups are significantly different. | Go to **Analyze** menu, choose **Nonparametric Tests**, then **2 independent samples**. Move the variable you wish to look at into the "**Test Variables**" box, and move the variable that divide the value into groups into “**Grouping Variable**” box, check a method in the “**Test type**” then click OK. |
| Two dependent- samples test | | Compares the distributions of two dependent variables. | The population distribution of the paired differences is assumed to be symmetric. | Null: There is no significant difference between the shapes of the two variables.  Alternate: There is a significant difference between the shapes of the two variables. | Go to **Analyze** menu, choose **Nonparametric Tests**, then **2 related samples**. Move the variable you wish to look at into the "**Test pairs (list)**" box, check a method in the “**Test type**” then click OK. |
| Kendall's Rank Correlation | | Test association for ordinal or ranked variables that take ties into account. The sign of the coefficient indicates the direction of the relationship, and its absolute value indicates the strength, with larger absolute values indicating stronger relationships. | Both variables are NOT normally distributed. | Null: There is no association between the two variables.  Alternate: There is an association between the two variables. | Under the **Analyze** menu, choose **Correlations-Bivariate**. Move the variables you wish to correlate into the "**Variables**" box. Under the "**Correlation Coefficients**," be sure that the "**Kendall’s tau-b**" box is checked off. |
| Spearman Rho correlation | | Test magnitude and direction of the association between two variables that are on an interval or ratio scale | Both variables are NOT normally distributed. | Null: There is no association between the two variables.  Alternate: There is an association between the two variables. | Under the **Analyze** menu, choose **Correlations-Bivariate**. Move the variables you wish to correlate into the "**Variables**" box. Under the "**Correlation Coefficients**," be sure that the "**Spearman**" box is checked off. |
| **T-TEST**  ## Import the packages | |
|  | |

|  |
| --- |
| import numpy as np |
|  |

|  |
| --- |
| from scipy import stats |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| ## Define 2 random distributions |
|  |

|  |
| --- |
| #Sample Size |
|  |

|  |
| --- |
| N = 10 |
|  |

|  |
| --- |
| #Gaussian distributed data with mean = 2 and var = 1 |
|  |

|  |
| --- |
| a = np.random.randn(N) + 2 |
|  |

|  |
| --- |
| #Gaussian distributed data with with mean = 0 and var = 1 |
|  |

|  |
| --- |
| b = np.random.randn(N) |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| ## Calculate the Standard Deviation |
|  |

|  |
| --- |
| #Calculate the variance to get the standard deviation |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| #For unbiased max likelihood estimate we have to divide the var by N-1, and therefore the parameter ddof = 1 |
|  |

|  |
| --- |
| var\_a = a.var(ddof=1) |
|  |

|  |
| --- |
| var\_b = b.var(ddof=1) |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| #std deviation |
|  |

|  |
| --- |
| s = np.sqrt((var\_a + var\_b)/2) |
|  |

|  |
| --- |
| s |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| ## Calculate the t-statistics |
|  |

|  |
| --- |
| t = (a.mean() - b.mean())/(s\*np.sqrt(2/N)) |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| ## Compare with the critical t-value |
|  |

|  |
| --- |
| #Degrees of freedom |
|  |

|  |
| --- |
| df = 2\*N - 2 |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| #p-value after comparison with the t |
|  |

|  |
| --- |
| p = 1 - stats.t.cdf(t,df=df) |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| print("t = " + str(t)) |
|  |

|  |
| --- |
| print("p = " + str(2\*p)) |
|  |

|  |
| --- |
| #Note that we multiply the p value by 2 because its a two tail t-test |
|  |

|  |
| --- |
| ### You can see that after comparing the t statistic with the critical t value (computed internally) we get a good p value of 0.0005 and thus we reject the null hypothesis and thus it proves that the mean of the two distributions are different and statistically significant. |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| ## Cross Checking with the internal scipy function |
|  |

|  |
| --- |
| t2, p2 = stats.ttest\_ind(a,b) |
|  |

|  |
| --- |
| print("t = " + str(t2)) |
|  |

print("p = " + str(2\*p2))