Types of Hypothesis Tests

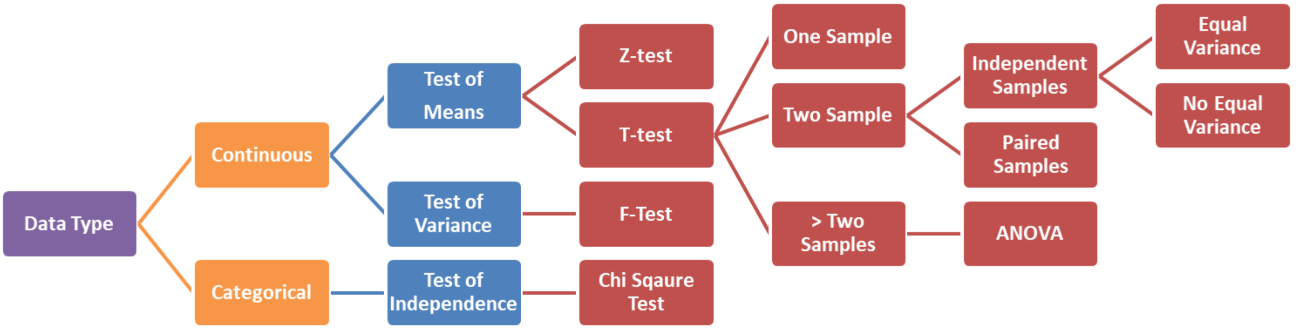
A Quick reference guide

BIREN GANDHI

YOUTUBE.COM/@BIRENGANDHI87

@BIRENGANDHI87

Different Hypothesis Tests



Z - Test

SAMPLE	1 Sample, the Sample size is large
TEST STATISTIC	$$Z = \frac{x - \mu}{\sigma}$$ Z = standard score x = observed value μ = mean of the sample σ = standard deviation of the sample
OUTPUT OF TEST	**Z - Statistic & p-Value** • Compare z stat with z critical to conclude if Ho has to be rejected or not. • At a 95% confidence interval compare the p-value with 0.05 to conclude if Ho has to be rejected or not.

EXCEL FORMULAS	**Left Tailed Test** ◦ p-value: **NORM.S.DIST** (INPUT: z-stat, cumulative = 1) ◦ z statistic: **NORM.S.INV** (INPUT: p-value) **Right Tailed Test** ◦ p-value: **1 - NORM.S.DIST** (INPUT: z-stat, cumulative = 1) ◦ z statistic: **NORM.S.INV** (INPUT: 1 - p-value) **Two-Tailed Test** ◦ p-value: **2*(1 - NORM.S.DIST)** (INPUT: |z-stat|, cumulative = 1) ◦ z statistic: **+/-(NORM.S.INV)** (INPUT: p-value/2)
PYTHON CODE	**import scipy.stats as stats** **Left Tailed Test** • Z-critical: **stats.norm.ppf** for the required level of significance, mean as '0' and std dev as '1' • p-value: **stats.norm.cdf** by passing z statistics. **Right Tailed Test** • Z-critical: **stats.norm.ppf** for the required (1- level of significance), mean as '0' and std dev as '1' • p-value: **1 - stats.norm.cdf** by passing z statistics. **Two-Tailed Test** • Z-critical (+/-): **stats.norm.ppf** for the required (level of significance/2), mean as '0' and std dev as '1' • p-value: **2*(1 - stats.norm.cdf)** by passing |z statistics|.
ASSUMPTIONS	• The samples from the population must be independent of one another • The populations from which the samples are taken must be normally distributed • The population standard deviations is known • The sample size is large (>30)
EXAMPLE	• An investor wants to check if the average daily returns of a stock are greater than 2% . • The standard deviation is assumed as 3.5%. A random sample of 100 returns is picked up to check the assumption.

T - Test : One Sample

SAMPLE	1 Sample, the Sample size is small
TEST STATISTIC	$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ μ: Population mean $\bar{x}$: Sample mean s: Sample standard deviation n: Sample size
OUTPUT OF TEST	**t - Statistic & p-Value** • Compare t - stat with t critical to conclude if Ho has to be rejected or not. • At a 95% confidence interval compare the p-value with 0.05 to conclude if Ho has to be rejected or not.
EXCEL FORMULAS	**Left Tailed Test** ○ p-value: **T.DIST** (INPUT: t-stat, DOF, cumulative = 1) ○ t statistic: **T.INV** (INPUT: p-value, DOF) **Right Tailed Test** ○ p-value: **1 - T.DIST** (INPUT: t-stat, DOF, cumulative = 1) ○ t statistic: **T.INV** (INPUT: 1 - p-value, DOF) **Two-Tailed Test** ○ p-value: **T.DIST.2T** (INPUT: t-stat, DOF) ○ t statistic: **T.INV.2T** (INPUT: p-value, DOF)
PYTHON CODE	**from scipy.stats import ttest_1samp** • **stats import ttest_1samp :** pass the data and population mean • p-value returned is for a two-tailed test • Divide the p-value by 2 if performing the one-tailed test
ASSUMPTIONS	• The samples from the population must be independent • The populations from which the samples are taken must be normally distributed

EXAMPLE	• A tire manufacturing company claims that the diameter of the standard tire product is 20 inches. • No of the samples available is 25.

T - Test : Two Sample Independent

SAMPLE	2 independent samples
TEST STATISTIC	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ **X̄1** is mean of first sample **X̄2** is mean of second sample **μ1** is the mean of first population **μ2** is the mean of second population **s1** is the standard deviation of first sample **s2** is the standard deviation of second sample **n1** is the size of the first sample **n2** is the size of the second sample
OUTPUT OF TEST	**t - Statistic & p-Value** • Compare t - stat with t critical to conclude if Ho has to be rejected or not. • At a 95% confidence interval compare the p-value with 0.05 to conclude if Ho has to be rejected or not.
EXCEL FORMULAS	1. **Go to Data Menu > Data Analysis > select** one of the following based on your case. a. t-Test: Two-Sample for Equal Variances b. t-Test: Two-Sample for Unequal Variance 2. Under Input, select the ranges for both sample 1 and sample 2. 3. In Hypothesized Mean Difference, you'll typically enter zero. This value is the null hypothesis value, which represents no effect. In this case, a mean difference of zero represents no difference between the two methods, which is no effect. 4. Check the Labels checkbox if you have sample labels in row 1. 5. Excel uses a default Alpha value of 0.05, which is usually a good value.

	6. Click OK and interpret the results. Output includes p-value for both one-tailed and two-tailed test
PYTHON CODE	**from scipy.stats import ttest_ind** • **stats import ttest_ind:** pass the 2 samples and mention if the variance is equal or not equal • p-value returned is for a two-tailed test. • Divide the p-value by 2 if performing a one-tailed test to interpret the outcome of the test
ASSUMPTIONS	• The samples from the population must be independent of one another • The populations from which the samples are taken must be normally distributed • Data from each group is obtained by random sampling from the population
EXAMPLE	• There is a complaint from a customer about a private retail brand selling breakfast cereals. • The packs sold by Store 1 are not equal in weight to the same packs sold by Store 2. • Random samples of the packs can be taken from the 2 stores, weights can be measured for the samples, and a two-tailed hypothesis t-test is performed to check if the weights are indeed different.

T - Test: Two Sample Paired

SAMPLE	2 Samples, 1st sample corresponds to initial data and 2nd sample corresponds to revised data after some treatment or process implementation.
TEST STATISTIC	$t_{calc} = \frac{\bar{d}}{s_d / \sqrt{n}}$ • d is the difference between each pair of data • is the average of d • n is the sample size of either population of interest

OUTPUT OF TEST	**t - Statistic & p-Value** • Compare t - stat with t critical to conclude if Ho has to be rejected or not. • At a 95% confidence interval compare the p-value with 0.05 to conclude if Ho has to be rejected or not.
EXCEL FORMULAS	1. **Go to Data Menu > Data Analysis > select t-Test: Paired Two Sample for Means** 2. Under Input, select the ranges for both sample 1 and sample 2. 3. In Hypothesized Mean Difference, you'll typically enter zero. This value is the null hypothesis value, which represents no effect. In this case, a mean difference of zero represents no difference between the two methods, which is no effect. 4. Check the Labels checkbox if you have sample labels in row 1. 5. Excel uses a default Alpha value of 0.05, which is usually a good value. 6. Click OK and interpret the results. Output includes p-value for both one-tailed and two-tailed test
PYTHON CODE	**import scipy.stats as stats** • **stats.ttest_rel:** pass the 2 related samples • p-value returned is for a two-tailed test. • Divide the p-value by 2 if performing a one-tailed test to interpret the outcome of the test
ASSUMPTIONS	• The samples from the population must be independent of one another • The populations from which the samples are taken must be normally distributed • Data from each group is obtained by random sampling from the population
EXAMPLE	• The performance score of 30 employees is recorded • Next, they are provided with high-cost rigorous training for 1 month. • The performance score of the 30 employees is measured again after 3 months of the training to check if there is a significant difference between their scores.

F - Test

SAMPLE	2 Samples
TEST STATISTIC	$\text{F Value} = \frac{\text{Larger Sample Variance}}{\text{Smaller Sample Variance}} = \frac{\sigma_1^2}{\sigma_2^2}$
OUTPUT OF TEST	**f - Statistic & p-Value** • Compare f - stat with f critical to conclude if Ho has to be rejected or not. • At a 95% confidence interval compare the p-value with 0.05 to conclude if Ho has to be rejected or not.
EXCEL FORMULAS	1. **F.TEST** 2. Pass the 2 arrays of data samples that need to be checked for variances. 3. Refer https://support.microsoft.com/en-us/office/f-test-function-100a59e7-4108-46f8-8443-78ffacb6c0a7?ns=excel&version=16&syslcid=1033&uilcid=1033&appver=zxl160&helpid=xlmain11.chm60539&ui=en-us&rs=en-us&ad=us
PYTHON CODE	**import scipy.stats as stats** • Use the formula in python to calculate the F statistic • Use '**stats.f.cdf**' for the F stat along with the degree of freedom of sample 1 and sample 2 to get the p-value and interpret the test results
ASSUMPTIONS	• The samples from the population must be independent of one another • The populations from which the samples are taken must be normally distributed
EXAMPLE	• F-test is used when comparing statistical models that have been fitted to a dataset, in order to identify the best-fit model.

CHI SQUARE Test

SAMPLE	Tests to see whether distributions of categorical variables differ from each another
TEST STATISTIC	$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$ χ^2 = chi squared O_i = observed value E_i = expected value
OUTPUT OF TEST	**Chi-Square Statistic & p-Value** • Compare **Chi-Square Statistic** with **Chi-Square** critical to conclude if Ho has to be rejected or not. • At a 95% confidence interval compare the p-value with 0.05 to conclude if Ho has to be rejected or not.
EXCEL FORMULAS	1. **CHISQ.TEST** 2. Pass the actual and expected range a. Actual Range: The range of data that contains observations to test against expected values. b. Expected Range: The range of data that contains the ratio of the product of row totals and column totals to the grand total. 3. Refer https://support.microsoft.com/en-us/office/chisq-test-function-2e8a7861-b14a-4985-aa93-fb88de3f260f?ns=excel&version=16&syslcid=1033&uilcid=1033&appver=zxl160&helpid=xlmain11.chm60538&ui=en-us&rs=en-us&ad=us
PYTHON CODE	**from scipy.stats import chi2_contingency** • Use '**chi2_contingency**' and pass the contingency table to get the p-value and interpret the test results
	• The samples from the population must be independent of one another • The populations from which the samples are taken must be normally distributed

ASSUMPTIONS	• The data in the contingency table should be frequencies or counts of cases • The levels (or categories) of the variables are mutually exclusive • The study groups must be independent • There are 2 variables, and both are measured as categories, usually at the nominal or ordinal level • The expected value of the number of sample observations in each level of the variable is at least 5
EXAMPLE	• If the Finance Ministry wants to determine if there is a relationship between voter's opinion concerning a new tax reform bill and their level of income, the chi-square test can be used by collecting data for the same.

THANKS FOR
WATCHING

DON'T FORGET TO LIKE,
COMMENT AND SHARE

BIREN GANDHI

SUBSCRIBE

YOUTUBE.COM/@BIRENGANDHI87

@BIRENGANDHI87