Scipy.stats

```
from scipy import stats
```

1. **One-sample T-test:**

scipy.stats.ttest_1samp(a, popmean, axis=0, nan_policy='propagate', alternative='two-sided')

- **a:** array_like - Sample data.
- **popmean:** float or array_like - Expected value in null hypothesis.
- **axis:** int or None, optional - Axis along which to compute test; default is 0.
- **nan_policy:** {'propagate', 'raise', 'omit'}, optional - Defines how to handle when input contains NaN; default is 'propagate'.
- **alternative:** {'two-sided', 'less', 'greater'}, optional - Defines the alternative hypothesis; default is 'two-sided'.

Ex: **t_test, p_value = stats.ttest_1samp(sample, population.mean())**

2. **Independent T-test (two-sample T-test):**

scipy.stats.ttest_ind(a, b, axis=0, equal_var=True, nan_policy='propagate', alternative='two-sided')

- **a, b:** array_like - The arrays must have the same shape, except in the dimension corresponding to axis (the first, by default).
- **axis:** int or None, optional - Axis along which to compute test; default is 0.
- **equal_var:** bool, optional - If True (default), perform a standard independent 2 sample test that assumes equal population variances. If False, perform Welch's t-test, which does not assume equal population variance.
- **nan_policy:** {'propagate', 'raise', 'omit'}, optional - Defines how to handle when input contains NaN; default is 'propagate'.
- **alternative:** {'two-sided', 'less', 'greater'}, optional - Defines the alternative hypothesis; default is 'two-sided'.

Ex: **t_test, p_value = stats.ttest_ind(sample1, sample2)**

3. **Chi-square Test of Independence:**

scipy.stats.chi2_contingency(observed, correction=True, lambda_=None)

- **observed:** array_like - The contingency table. The table contains the observed frequencies.
- **correction:** bool, optional - If True (default), calculate the chi-square statistic with Yates' correction for continuity. The effect of the correction is to adjust each observed value by 0.5 towards the corresponding expected value.
- **lambda_:** float or str, optional - By default, the statistic computed in this test is Pearson's chi-squared statistic. lambda_ allows a statistic from the Cressie-Read power divergence family to be used instead.

Ex: **stats.chi2_contingency(observed_values)**

4. **Percentile Point Function (Inverse of Cumulative Distribution Function) for Chi-square Distribution:**

scipy.stats.chi2.ppf(q, df)

- **q:** array_like - Lower tail probability.

- **df:** array_like - Degrees of freedom.

Ex: **stats.chi2.ppf(q=0.95, df=(4-1)*(6-1))**

5. **Cumulative Distribution Function for Chi-square Distribution:**

scipy.stats.chi2.cdf(x, df)

- **x:** array_like - Percentile values.
- **df:** array_like - Degrees of freedom.

Ex: **p_value = 1 - stats.chi2.cdf(x=chi_sqr, df=(4-1)*(6-1))**

6. **One-way ANOVA test:**

scipy.stats.f_oneway(*args)

- ***args:** array_like - The sample measurements for each group. Accepts two or more array-like objects.

Ex: **f_value, p_value = stats.f_oneway(excel['setosa'], excel['versicolor'], excel['virginica'])**

7. Z-test

scipy.stats.ztest(a, x2=None, value=0, alternative='two-sided', usevar='pooled', ddof=1.0)

- a: This is the array or sequence of sample observations.
- x2 (optional): If provided, this represents the second sample or value to compare with the first sample.
- value (optional): This is the value of the null hypothesis (μ0) you're testing against.
- alternative (optional): This specifies the alternative hypothesis. It can take one of three values: 'two-sided', 'larger', or 'smaller'. The default is 'two-sided', indicating that you're interested in determining if the means are not equal. 'larger' indicates a one-tailed test for greater-than, and 'smaller' indicates a one-tailed test for less-than.
- usevar (optional): This specifies whether to use the pooled sample variance ('pooled') or the unbiased estimate of variance ('unequal'). The default is 'pooled'.
- ddof (optional): This is the degrees of freedom adjustment for the variance. Default is 1.0.

This function returns two values:

- The Z-statistic, which measures how many standard deviations an element is from the mean.
- The p-value for the hypothesis test.