**Cheat Sheet:**

**Mean:**

**Population Variance**

**Sample Variance**

**Population Standard Deviation:**

**Population Variance**:

**Three Equivalent ways to Extract Subsets of Data**

'target\_col\_name’ = Column name of the outcome variable

Example: ‘Blood\_Pressure’

'categorical\_col\_name' = Column name of the categorical variable

Example: ‘Gender’

'category' = value within the category

Example: ‘Male’

df\_house['target\_col\_name'][df\_house['categorical\_col\_name']=='category']

df\_house.loc[df\_house['categorical\_col\_name']=='category']['target\_col\_name']

df\_house.loc[df\_house['categorical\_col\_name']=='category','target\_col\_name']

t-test using stats package:

import scipy.stats as stats

import pandas as pd

stats.ttest\_ind(a, b)

Where ‘a’ and ‘b’ are lists of outcomes for groups ‘a’ and ‘b’.

**Null Hypothesis:**

Low p-value is evidence against the Null. We either ‘Reject’ or ‘Fail to Reject’ the Null.

**Levels:** 1 percent, 5 percent, 10 percent and 15 percent.