**Statistics Formula Reference**

| **Notes** |
| --- |

* Before using the formula on Excel, use the = sign. The variable ‘range’ denotes the datasets.
* To use the formulas in Python, you can import Pandas or the NumPy libraries. NumPy has built-in functions for covariance and correlation as well.
* Save this handout for future sessions.

| **Measure** | **Excel Formula** | **Python (Pandas/NumPy)** | **Power BI (DAX)** | **Regular Formula** |
| --- | --- | --- | --- | --- |
| **Mean** | AVERAGE(range) | df['column'].mean()  or  np.mean(array) | AVERAGE(column) |  |
| **Median** | MEDIAN(range) | df['column'].median()  or  np.median(array) | MEDIAN(column) | Middle value of sorted data |
| **Mode** | MODE.SNGL(range) | df['column'].mode() | MODE(column) | Most frequent value |
| **Range** | MAX(range) - MIN(range) | df['column'].max() - df['column'].min() | MAX(column) - MIN(column) | Range=Max−Min |
| **Variance (Population)** | VARP(range) | np.var(array) | VAR.P(column) |  |
| **Variance (Sample)** | VAR.S(range) | np.var(array, ddof=1) | VAR.S(column) | ​ |
| **Standard Deviation (Population)** | STDEV.P(range) | np.std(array) | STDEV.P(column) |  |
| **Standard Deviation (Sample)** | STDEV.S(range) | np.std(array, ddof=1) | STDEV.S(column) |  |
| **Covariance** | COVARIANCE.P(range1, range2) | np.cov(array1, array2)[0, 1] | COVARIANCE.P(column1, column2) |  |
| **Correlation** | CORREL(range1, range2) | np.corrcoef(array1, array2)[0, 1] | CORREL(column1, column2) |  |