**Multivariate Normal Distribution Analysis**

**Note :**  In Result table for Multi -dimensional data having 10000 rows and 10 features (columns) has been considered.

MVN : Multi Variate Normality

H0 : Data follows Multivariate Normal Distribution

H1 : Data follows Multivariate Normal Distribution

**1) Mardia’s Skewness and Kurtosis Test**

Implementation : Available in R

Result :

1. Non normal data

|  |
| --- |
| Test Statistic p value Result |
| 1 Mardia Skewness 20226.4111606402 0 NO |
| 2 Mardia Kurtosis 48.089367223179 0 NO |
| 3 MVN <NA> <NA> NO |

Time Taken : 6.372337 secs

1. Normal data

|  |
| --- |
| Test Statistic p value Result |
| 1 Mardia Skewness 211.311888427177 0.650946148505642 YES |
| 2 Mardia Kurtosis -0.138113002620644 0.890151109655654 YES |
| 3 MVN <NA> <NA> YES |

Time Taken : 6.384627 secs

Characteristics : Works well for both normal and non-normal data having different shape and size

Limitation : It takes too much time for computation when the size of data is too large ( i.e. 40000-50000 rows and 2-3 features)

Research Paper : PDF Attached

**2) Henze-Zirkler Test**

Implementation : Available in R

Result:

1. Non-Normal data

|  |
| --- |
| 1 Test HZ p value MVN |
| 2 Henze-Zirkler 1.849112 0 NO |

Time Taken : 9.487739 secs

1. Normal data

|  |
| --- |
| Test HZ p value MVN |
| 1 Henze-Zirkler 0.9952236 0.723792 YES |

Time Taken : 9.48199 secs

Limitation : Can be applied to data having any shape and size but gives incorrect outcomes when the data is of large size and more dimension

Research Paper : <https://www.tandfonline.com/doi/abs/10.1080/03610929008830400>

**3) Royston’s Test**

Implementation : Available in R

Result: # Result not given due to major limitation

Limitation : This test can be applied to data having rows <= 5000

Research Paper : <https://academic.oup.com/jrsssc/article-abstract/32/2/121/6985143>

**4) Doornik Hansen Test**

Implementation : Available in R

Result:

1. Non-Normal data

|  |
| --- |
| Test E df p value MVN |
| 1 Doornik-Hansen 33946.97 20 0 NO |

Time Taken : 0.3924642 secs

1. Normal data

|  |
| --- |
| Test E df p value MVN |
| 1 Doornik-Hansen 9.850942 20 0.9707949 YES |

Time Taken : 0.3848441 secs

Limitation : No, works well for all kind of data and any shape and size

Research Paper : PDF Attached

**5) Energy Test**

Implementation : Available in R

Result:

Not given as

Limitation : Works well but takes too much time even more than Mardia’s skewness and kurtosis test , for small size data also it takes too much time

Research Paper :

**6) Shapiro Wilk test for Multivariate Normality**

Implementation : Available in R and Python both

Result (Obtained in Python) :

1. Normal data

ShapiroResult(statistic=0.9999731183052063, pvalue=0.7102762460708618)

p-value > 0.05 Hence Can not reject Null Hypothesis which

Time Taken: 0.0229 sec

1. Normal data

ShapiroResult(statistic=0.4656309485435486, pvalue=0.0

p-value < 0.05 Hence Null Hypothesis is rejected

Time Taken : 0.019

Limitation : In R programming it cannot be applied to data having more than 5000 rows, but works well in Python for any kind data and varying shape - size

Research Paper : PDF Attached

**7) Jarque Bera Test**

Implementation : Available in R and Python both

Result (Python ):

1. Non-Normal data

SignificanceResult(statistic=695163.6873627615, pvalue=0.0)

Time Taken: 0.0069

1. Normal data

SignificanceResult(statistic=4.288522169116266, pvalue=0.11715457239166226)

Time Taken: 0.017

Result (R) :

Non-Normal Data

Normal Data

# Does not work on multidimensional data can be applied on single dimensional data but in Python it works for both univariate and multivariate data

Limitation :

Research Paper : PDF Attached

**8) Lilliefors Test**

Implementation : Available in R

Result (R) :

1. Non-Normal data

data: non\_normal\_data

D = 0.20922, p-value < 2.2e-16

P\_value < 0.05 Hence reject H0

Time Taken : 0.05805802 secs

1. Normal data

data: normal\_data

D = 0.0014596, p-value = 0.8724

P\_value > 0.05 Hence Can’t reject H0

Time Taken : 0.04819107 secs

Result (Python) :

# Does not work on Multidimensional data only works for Single dimension data

Limitation :

Research Paper : PDF Attached