Statistical Research

This document is to provide backing and reasoning for the use of certain statistical testing algorithms.

Anderson-Darling

Data will be sampled in various amounts and tested for normality using the Anderson-Darling test. 1000 of these tests will occur per various sample size.

Blue: 30 Sampled data points

Orange: 50 Sampled data points

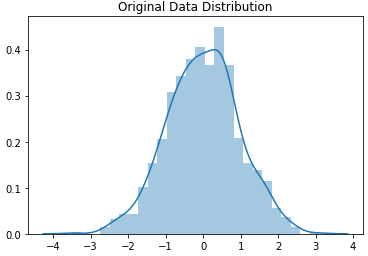
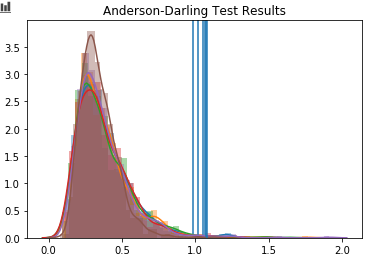
Green: 100 Sampled data points

Red: 150 Sampled data points

Purple: 200 Sampled data points

Baseline:

Normal distribution created with np.random.normal(mu, sigma, n) where mu ,sigma ,n = 0,1,1000.



Baseline results:

Sample size 30: 1.001001001001001% of data misclassified non-normal

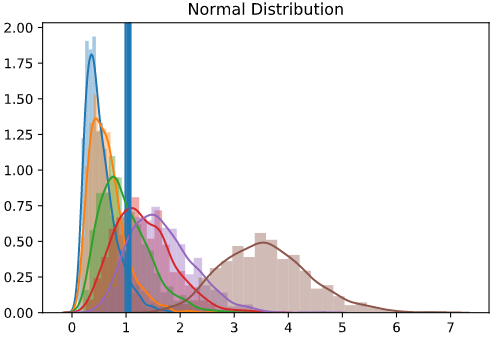
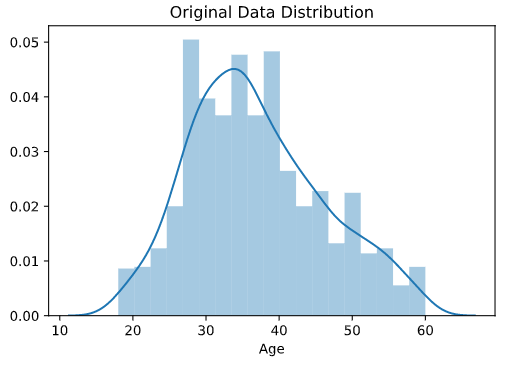
Sample size 50: 1.1011011011011012% of data misclassified non-normal

Sample size 100: 0.9009009009009009% of data misclassified non-normal

Sample size 150: 0.7007007007007007% of data misclassified non-normal

Sample size 200: 0.9009009009009009% of data misclassified non-normal

Sample size 500: 0.10010010010010009% of data misclassified non-normal



Anderson-Darling Test Results

Group of critical values for

0.01 significance level

Normal Distribution:

Sample size 30: 6.706706706706707% of data misclassified non-normal

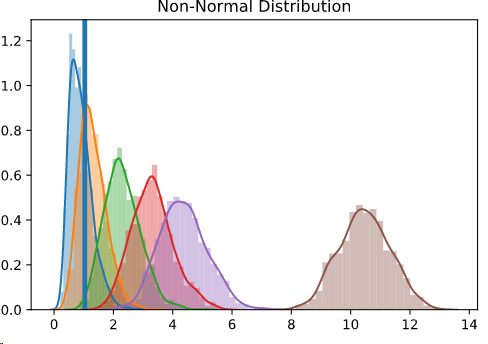
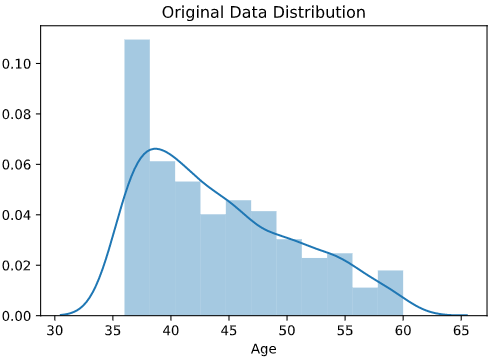
Sample size 50: 14.314314314314313% of data misclassified non-normal

Sample size 100: 40.44044044044044% of data misclassified non-normal

Sample size 150: 64.86486486486487% of data misclassified non-normal

Sample size 200: 83.58358358358359% of data misclassified non-normal

Sample size 500: 100.0% of data misclassified non-normal

**Gumbel**

Anderson-Darling Test Results

Group of critical values for

0.01 significance level

Non-Normal Right Skewed Distribution:

Sample size 30: 66.16616616616616% of data misclassified normal

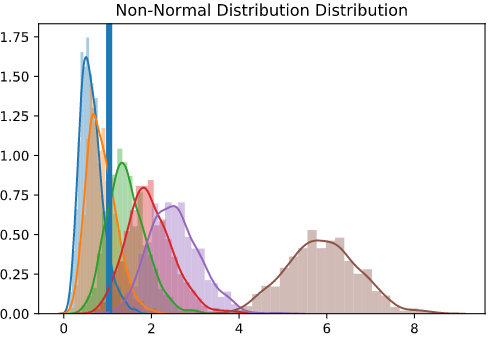
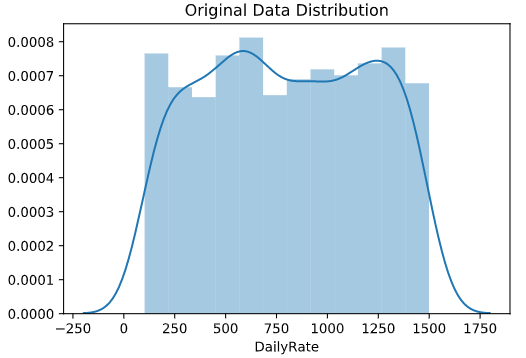
Sample size 50: 31.33133133133133% of data misclassified normal

Sample size 100: 0.7007007007007007% of data misclassified normal

Sample size 150: 0.0% of data misclassified normal

Sample size 200: 0.0% of data misclassified normal

Sample size 500: 0.0% of data misclassified normal



Anderson-Darling Test Results

Group of critical values for

0.01 significance level

Sample size 30: 90.49049049049049% of data misclassified normal Sample size 50: 70.57057057057057% of data misclassified normal Sample size 100: 19.21921921921922% of data misclassified normal Sample size 150: 2.6026026026026026% of data misclassified normal Sample size 200: 0.0% of data misclassified normal Sample size 500: 0.0% of data misclassified normal

Uniform Distribution

Sample size 30: 10.01001001001001% of data misclassified normal

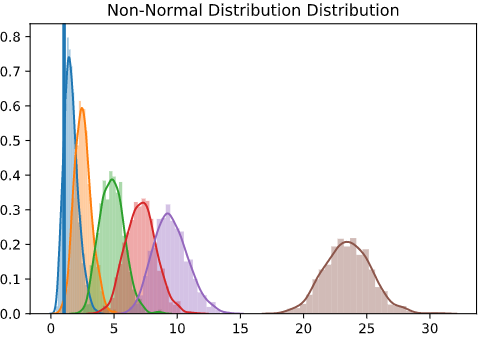
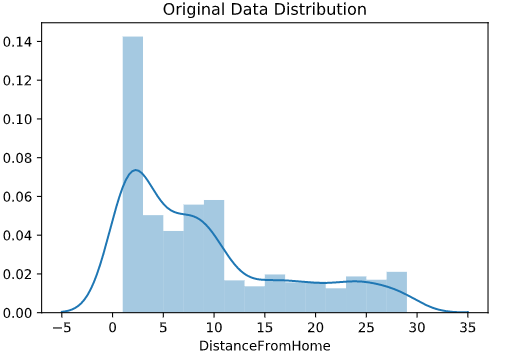
Sample size 50: 28.02802802802803% of data misclassified normal

Sample size 100: 76.67667667667668% of data misclassified normal

Sample size 150: 97.5975975975976% of data misclassified normal

Sample size 200: 99.7997997997998% of data misclassified normal

Sample size 500: 100.0% of data misclassified normal



Anderson-Darling Test Results

Group of critical values for

0.01 significance level

Non-Normal Right Skewed Distribution

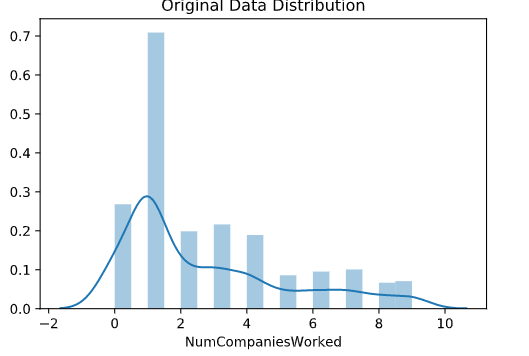
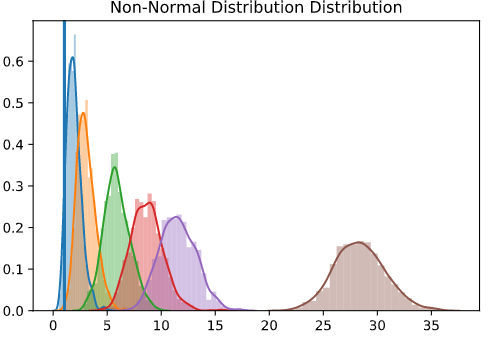
Sample size 30: 11.21121121121121% of data misclassified normal

Sample size 50: 0.7007007007007007% of data misclassified normal

Sample size 100: 0.0% of data misclassified normal

Sample size 150: 0.0% of data misclassified normal

Sample size 200: 0.0% of data misclassified normal

Sample size 500: 0.0% of data misclassified normal

Anderson-Darling Test Results

Group of critical values for

0.01 significance level

Non-Normal Right Skewed Distribution:

Sample size 30: 5.105105105105105% of data misclassified normal

Sample size 50: 0.10010010010010009% of data misclassified normal

Sample size 100: 0.0% of data misclassified normal

Sample size 150: 0.0% of data misclassified normal

Sample size 200: 0.0% of data misclassified normal

Sample size 500: 0.0% of data misclassified normal