Journal of Manufacturing Engineering, June 2024, Vol. 19, Issue. 2, pp 040-056
DOI: https://doi.org/10.37255/jme.v19i2pp040-056

POWER COMPARISON OF TESTS FOR NORMALITY AND RECOMMENDED TEST ORDERS FOR MANUFACTURING ORGANIZATIONS

Lucius LaFromboise

Lucius LaFromboise Enterprises, LLC

Abstract

Manufacturing organizations are increasingly concerned with statistical analysis as a tool for understanding and improving processes. With the release of a new statistical package, QESuite, designed for manufacturers, it is important to directly compare the included normality tests to guide the standards of manufacturing organizations. A common metric of the usefulness of a normality test is its power, the test's ability to identify data that do not follow the normal distribution correctly. Through Monte Carlo simulation, the power of 6 normality tests: Anderson-Darling, Jarque-Bera, Kolmogorov-Smirnov, Lilliefors (corrected KS), Shapiro-Wilk (Royston), and Ryan-Joiner; was evaluated at multiple sample sizes with different original distributions. The sample size of the data being tested greatly impacted the power of all the tests studied. As the sample size increased, the power of almost all the tests studied approached 1.0 (100%). The underlying distribution also showed an effect, with the power being higher for all tests when evaluating asymmetrical distributions than symmetric distributions. When the power was averaged across all distributions and the average ranks of each test across sample sizes were calculated, the following general order of highest power to lowest is recommended: Shapiro-Wilk (Royston), Lilliefors, Ryan-Joiner, Anderson-Darling, Jarque-Bera, Kolmogorov-Smirnov.

Keywords: *quality engineering, normality testing, statistical process control, power comparison, type II error.*

1. Introduction

Today, there are several statistical packages that each offer a variety of methods for evaluating the normality of a data set. With many options for collecting and analyzing data, manufacturing organizations are increasingly adopting statistical tools to increase their knowledge and understanding of manufacturing processes. [1, 11]. Organizations need to use an understanding of the differences between these methods to set standards for analyzing normality that is based on evidence and that upholds the integrity of the conclusions [1]. A standard metric of the usefulness of a test is the test's power. The power of a statistical test represents the ability of the test to correctly reject the null hypothesis when it is false [7]. This is a valuable metric for the selectiveness of many tests for normality because the null hypothesis of these tests is that the data follow a normal distribution [3]; thus, the power for many tests for normality represents the probability that the test will correctly identify data that do not follow the normal distribution.

The power of a normality test to correctly identify data that do not follow the normal distribution is often as important or even more important than identifying data that do follow the normal distribution correctly; this is especially true in manufacturing. The power of a hypothesis can also be expressed as 1-β, where β represents the probability of committing a Type II error [10]. Type II errors are often considered the customer's risk, whereas the producer's risk is the probability of committing a Type I error (α). The probability of committing a Type II error (β) is the consumer's risk because it represents the risk that a failing condition is incorrectly accepted as passing. Because of the complementary relationship between β and power, tests with a higher power present a lower customer risk. In manufacturing, it is often advantageous to minimize the risk to the customer and to be able to provide evidence affirming this to the customer.

Due to the complexity of calculating the power of Type II error [10], it is often more practical to estimate the power of a test through Monte Carlo simulation using data generated from alternate distributions [11]. Many comparisons have been conducted to determine the

**Corresponding Author - E- mail: lucius.lafromboise@engsuite.org*

DOI: https://doi.org/10.37255/jme.v19i2pp040-056

relationship between normality tests and power rankings [6, 9, 11, 12]. The ranking of normality tests by power is especially important in manufacturing, an environment where customer demands drive a desire to minimize Type II errors, increase manufacturing integrity and quality standards, and a desire to minimize non-conformance and customer rejections [2]. As statistical packages become more common and affordable for manufacturers, there has been an increase in the desire for manufacturers to have robust analysis procedures based on scientific and technical research.

Currently, no technical sources perform direct power comparisons of the six normality tests included in the statistical package QESuite. QESuite was specifically designed for use by manufacturing organizations, and thus, it is important to provide a direct comparison of the power rankings to help guide procedures and testing sequences. By leveraging the direct comparison results, organizations can develop standardized testing sequences that can assess normality by beginning with the most powerful test and continuing through less powerful tests if necessary. By designing standards that assess normality in this descending power order, the organization can always ensure that they provide the least risk to the customer while still upholding the normality assumption when appropriate. A deeper understanding of the power of different statistical tests can also allow manufacturers to set limits to the minimum acceptable power, allowing for a higher risk of Type II error on less critical dimensions and posing stricter limits on critical dimensions. This deeper understanding can also allow organizations to provide different testing sequences based on the symmetry of data and sample size, allowing for flexible yet strict standards that leverage statistical knowledge to provide the highest confidence to customers.

2. Materials and Methods

The methodology of this test is based on the methodology used by Razali et al. [9]. It is intended to be a replicate study with the addition of the Ryan-Joiner and Jarque-Bera tests.

2.1. Simulation Procedure

This study used data generated by Monte Carlo simulation of 12 different non-normal populations with 4,000 data points. The alternative distributions studied were five symmetric distributions: Beta(2,2), t(300), t(10), t(7) & t(5) and seven asymmetric distributions: Beta(6,2), Beta(2,1), Beta(3,2), $\chi^2(20)$, Gamma(4,5), $\chi^2(4)$ and Gamma(1,5).

The data was generated in Microsoft Excel, using the RAND() and corresponding inverse distribution functions for each selected distribution.

After generating the data, it was analyzed using the QESuite.js NPM package in a node environment. 10,000 trials of randomly selected samples were generated for each sample size (N) studied. The 10,000 trials were generated using the JavaScript code Math. floor(Math. random()*4000). This script generates a random number between 0 (inclusive) and 1 (exclusive), then multiplies it by 4000 to allow the function to select any value in the population. The random value, now between 0 (inclusive) and 4000 (exclusive), is rounded down to the nearest integer using Math. floor(); this value is either pushed to the array of trial indices if it is unique or a new random value is generated if that index is already present. Using indices allows each sample in the population to be randomly selected only once per trial, as is expected behaviour when working with actual samples in a manufacturing environment. The indices represent the sample that corresponds to the zero-based index in the population; for example, index(random) = 0 is the first sample in the population.

This study was performed at 15 different levels of sample size, where N = 10, 15, 20, 25, 30, 40, 50, 100, 200, 300, 400, 500, 1000, 1500, 2000 samples. These levels were chosen to remain consistent with the comparison study by Razali et al. [9].

After all randomized trials were generated at each sample size level, the JavaScript program iterated through each sample size level and trial. The array of indices for each trial was used to construct arrays from each alternative distribution using the same indices for each trial. The new arrays of randomized samples were then assessed using the QESuite functions for each normality test at a significance level of $\alpha = 0.05$. The raw value of each p-value was then saved in a raw data array for each distribution. The result of the hypothesis was added to the running total of correctly rejected null hypotheses for each distribution at the sample size level if the result was determined significant and the null hypothesis rejected. This total was saved and used to calculate the power of the test for each distribution at each sample size level by dividing it by the number of trials (10,000). This value, being between 0 and 1, represents each test's power.

After performing all normality tests at each sample size, each distribution, and for all trials, the average at each sample size level for all distributions was calculated to determine a general power level at each sample size for all distributions for each test individually. This procedure was also used to calculate the average power of each test (and sample size) for the grouped symmetric tests and asymmetric to determine the effect of symmetric vs asymmetric data.

Scatterplots were generated in Excel to visualize the effect of sample size on the power of each test. A scatterplot was generated for each individual distribution, the symmetric and asymmetric groups, and the overall powers.

3. Results

All normality tests used in this study were performed using the QESuite statistical software, and significance was evaluated at $\alpha = 0.05$. The values in the power comparison tables below are decimals representing the number of times the test correctly identified non-normal data ($P \leq 0.05$) out of the 10,000 trials at each sample size.

3.1. Power Comparison for Symmetric Non-normal Distributions

Each table represents an individual alternative distribution to allow comparison of the power of each normality test based on the effect of sample size for each distribution. Figure 1 is a collection of the charts generated using the data found in the subsequent tables (Table 1 - 5).

3.2. Power Comparison for General Symmetric Non-normal Distributions

Table 6 represents the average power for each normality test at each sample size averaged across all symmetric alternative distributions studied. Table 7 represents the rank of each test at each sample size, with the overall average across sample sizes at the bottom. Figure 2 is the chart generated to visualize the effect of sample size on the average power for all symmetric distributions studied.

3.3. Power Comparison for Asymmetric Non-normal Distributions

Each table represents an individual alternative distribution to allow comparison of the power of each normality test based on the effect of sample size for each distribution. Figure 3 is a collection of the charts generated using the data found in the subsequent tables.

Table 1 Power Comparison for Beta(2,2) Distribution

Beta(2,2)						
N	**Anderson-Darling**	**Jarque-Bera**	**Kolmogorov-Smirnov**	**Lilliefors**	**Ryan-Joiner**	**Shapiro-Wilk (Royston Exp.)**
10	0.045	0.0146	0	0.0432	0.028	0.0388
15	0.0471	0.0045	0	0.0494	0.0207	0.0423
20	0.0577	0.0037	0	0.0498	0.0212	0.0529
25	0.0713	0.0008	0	0.0582	0.0239	0.0677
30	0.0807	0.0012	0	0.0653	0.025	0.0791
40	0.1073	0.0004	0	0.0727	0.0351	0.1123
50	0.1357	0.0001	0	0.0862	0.0472	0.1549
100	0.3262	0.0128	0	1	0.2323	0.4685
200	0.7172	0.61	0	1	0.89	0.9368
300	0.9355	0.9698	0	1	0.9998	0.9985
400	0.991	0.9997	0	1	1	1
500	0.9996	1	0	1	1	1
1000	1	1	0	1	1	1
1500	1	1	0	1	1	1
2000	1	1	0	1	1	1

Table 2 Power Comparison for t(300) Distribution

	t(300)					
N	**Anderson-Darling**	**Jarque-Bera**	**Kolmogorov-Smirnov**	**Lilliefors**	**Ryan-Joiner**	**Shapiro-Wilk (Royston Exp.)**
10	0.0485	0.0429	0.0547	0.9031	0.0449	0.9608
15	0.0497	0.045	0.0252	0.9821	0.0485	0.9957
20	0.0484	0.0419	0.0103	0.9973	0.045	0.9998
25	0.0504	0.0417	0.0047	0.9996	0.0455	1
30	0.0452	0.0408	0.0019	0.9999	0.0414	1
40	0.0488	0.039	0.0004	1	0.0395	1
50	0.0479	0.0353	0	1	0.036	1
100	0.0438	0.0245	0	1	0.0333	1
200	0.0395	0.0202	0	1	0.053	1
300	0.0391	0.0168	0	1	0.1247	1
400	0.0365	0.018	0	1	0.3892	1
500	0.0298	0.0149	0	1	0.9186	1
1000	0.0183	0.0126	0	1	1	1
1500	0.0095	0.0118	0	1	1	1
2000	0.0037	0.0056	0	1	1	1

Table 3 Power Comparison for t(10) Distribution

	t(10)					
N	**Anderson-Darling**	**Jarque-Bera**	**Kolmogorov-Smirnov**	**Lilliefors**	**Ryan-Joiner**	**Shapiro-Wilk (Royston Exp.)**
10	0.0651	0.0771	0.0583	0.9066	0.0757	0.9627
15	0.0794	0.1024	0.0286	0.9833	0.0971	0.996
20	0.0847	0.1154	0.0121	0.9973	0.1062	0.9998
25	0.0877	0.1353	0.0053	0.9997	0.119	1
30	0.0897	0.1441	0.002	0.9999	0.1268	1
40	0.0981	0.1661	0.0005	1	0.1457	1
50	0.1061	0.1947	0	1	0.1661	1
100	0.1439	0.2741	0	1	0.2465	1
200	0.2153	0.4158	0	1	0.4666	1
300	0.2897	0.5251	0	1	0.723	1
400	0.3857	0.6418	0	1	0.9304	1
500	0.4584	0.7201	0	1	0.9986	1
1000	0.798	0.95	0	1	1	1
1500	0.9548	0.9956	0	1	1	1
2000	0.9957	0.9999	0	1	1	1

Table 4 Power Comparison for t(7) Distribution

			t(7)			
N	**Anderson-Darling**	**Jarque-Bera**	**Kolmogorov-Smirnov**	**Lilliefors**	**Ryan-Joiner**	**Shapiro-Wilk (Royston Exp.)**
10	0.0804	0.0959	0.06	0.9077	0.0931	0.963
15	0.1044	0.1402	0.03	0.9834	0.1284	0.9963
20	0.1145	0.1653	0.0127	0.9974	0.151	0.9998
25	0.1237	0.1951	0.0054	0.9998	0.1746	1
30	0.1319	0.2181	0.002	1	0.1922	1
40	0.1528	0.2635	0.0006	1	0.2312	1
50	0.17	0.3118	0	1	0.2721	1
100	0.2639	0.4664	0	1	0.4274	1
200	0.4353	0.6921	0	1	0.7268	1
300	0.5986	0.8247	0	1	0.9153	1
400	0.7332	0.9116	0	1	0.9919	1
500	0.827	0.9512	0	1	1	1
1000	0.9912	0.9995	0	1	1	1
1500	1	1	0	1	1	1
2000	1	1	0	1	1	1

Table 5 Power Comparison for t(5) Distribution

			t(5)			
N	**Anderson-Darling**	**Jarque-Bera**	**Kolmogorov-Smirnov**	**Lilliefors**	**Ryan-Joiner**	**Shapiro-Wilk (Royston Exp.)**
10	0.1085	0.1294	0.0633	0.9128	0.1236	0.964
15	0.1462	0.1934	0.0324	0.984	0.1834	0.9967
20	0.1683	0.2384	0.0133	0.9975	0.221	0.9998
25	0.1983	0.2877	0.0056	0.9998	0.2641	1
30	0.2158	0.3281	0.0022	1	0.2936	1
40	0.2653	0.4042	0.0007	1	0.3667	1
50	0.3063	0.4688	0	1	0.4283	1
100	0.4849	0.6843	0	1	0.6557	1
200	0.7636	0.9045	0	1	0.9205	1
300	0.8989	0.9694	0	1	0.988	1
400	0.9674	0.9936	0	1	0.9997	1
500	0.9889	0.9989	0	1	1	1
1000	1	1	0	1	1	1
1500	1	1	0	1	1	1
2000	1	1	0	1	1	1

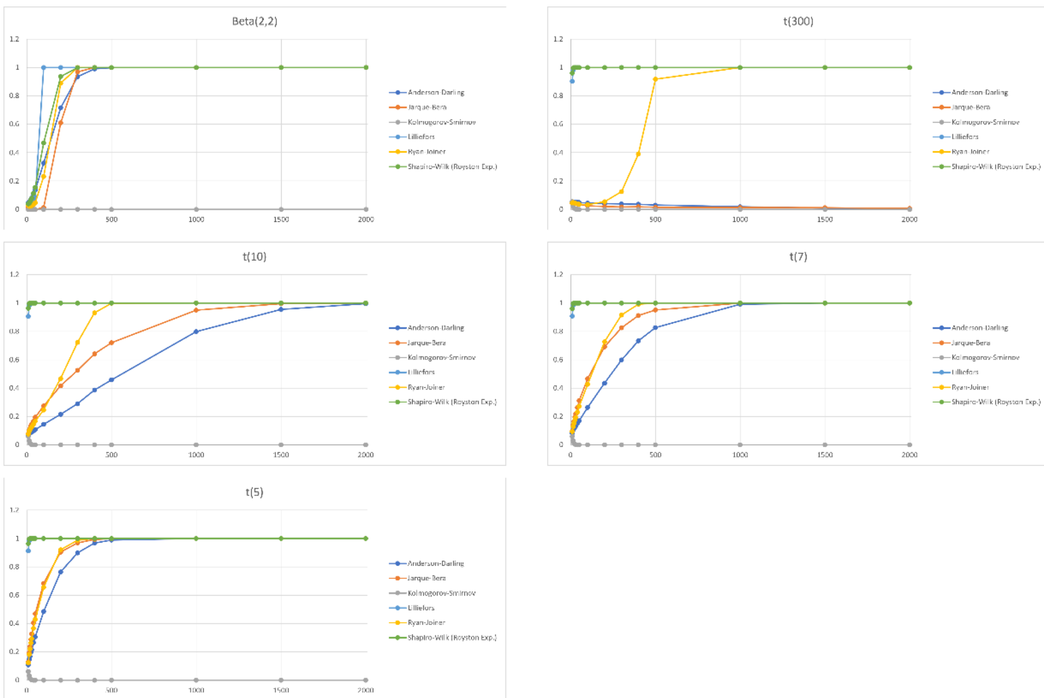

Fig. 1 Collection of Power Comparison Charts for Symmetric Non-normal Distributions

Table 6 Average Power Across All Symmetric Distributions

Average Power Across All Symmetric Distributions						
N	**Anderson-Darling**	**Jarque-Bera**	**Kolmogorov-Smirnov**	**Lilliefors**	**Ryan-Joiner**	**Shapiro-Wilk (Royston Exp.)**
10	0.0695	0.07198	0.04726	0.73468	0.07306	0.77786
15	0.08536	0.0971	0.02324	0.79644	0.09562	0.8054
20	0.09472	0.11294	0.00968	0.80786	0.10888	0.81042
25	0.10628	0.13212	0.0042	0.81142	0.12542	0.81354
30	0.11266	0.14646	0.00162	0.81302	0.1358	0.81582
40	0.13446	0.17464	0.00044	0.81454	0.16364	0.82246
50	0.1532	0.20214	0	0.81724	0.18994	0.83098
100	0.25254	0.29242	0	1	0.31904	0.8937
200	0.43418	0.52852	0	1	0.61138	0.98736
300	0.55236	0.66116	0	1	0.75016	0.9997
400	0.62276	0.71294	0	1	0.86224	1
500	0.66074	0.73702	0	1	0.98344	1
1000	0.7615	0.79242	0	1	1	1
1500	0.79286	0.80148	0	1	1	1
2000	0.79988	0.8011	0	1	1	1

Journal of Manufacturing Engineering, June 2024, Vol. 19, Issue. 2, pp 040-056
DOI: https://doi.org/10.37255/jme.v19i2pp040-056

Table 7 Rank and Average Rank Across All Symmetric Distributions

Rank and Average Rank Across All Symmetric Distributions						
N	Anderson-Darling	Jarque-Bera	Kolmogorov-Smirnov	Lilliefors	Ryan-Joiner	Shapiro-Wilk (Royston Exp.)
10	5	4	6	2	3	1
15	5	3	6	2	4	1
20	5	3	6	2	4	1
25	5	3	6	2	4	1
30	5	3	6	2	4	1
40	5	3	6	2	4	1
50	5	3	6	2	4	1
100	5	4	6	1	3	2
200	5	4	6	1	3	2
300	5	4	6	1	3	2
400	5	4	6	1	3	1
500	5	4	6	1	3	1
1000	5	4	6	1	1	1
1500	5	4	6	1	1	1
2000	5	4	6	1	1	1
Avg	5	3.6	6	1.466666667	3	1.2
Avg Rank	5	4	6	2	3	1

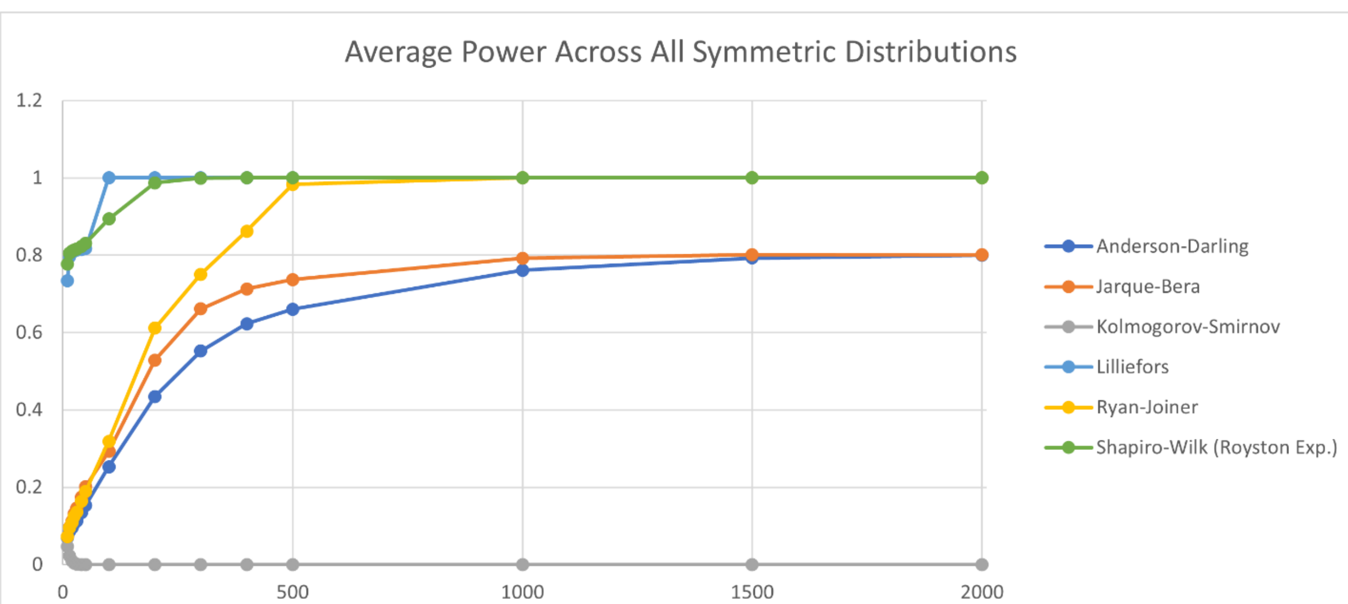

Fig. 2 Average Power Across All Symmetric Distributions

Journal of Manufacturing Engineering, June 2024, Vol. 19, Issue. 2, pp 040-056
DOI: https://doi.org/10.37255/jme.v19i2pp040-056

Table 8 Power Comparison for Beta(6,2) Distribution

			Beta(6,2)			
N	Anderson-Darling	Jarque-Bera	Kolmogorov-Smirnov	Lilliefors	Ryan-Joiner	Shapiro-Wilk (Royston Exp.)
10	0.0951	0.0695	0	0.0866	0.0943	0.094
15	0.1351	0.0907	0	0.1076	0.1297	0.1494
20	0.1836	0.1155	0	0.137	0.1764	0.2116
25	0.2271	0.1273	0	0.1677	0.2142	0.2702
30	0.2666	0.1488	0	0.1944	0.2603	0.3276
40	0.3672	0.2017	0	0.2535	0.3683	0.4637
50	0.4635	0.2611	0	0.3093	0.4793	0.5921
100	0.8374	0.6575	0	1	0.9039	0.9458
200	0.9965	0.9941	0	1	1	1
300	0.9999	1	0	1	1	1
400	1	1	0	1	1	1
500	1	1	0	1	1	1
1000	1	1	0	1	1	1
1500	1	1	0	1	1	1
2000	1	1	0	1	1	1

Table 9 Power Comparison for Beta(2,1) Distribution

			Beta(2,1)			
N	Anderson-Darling	Jarque-Bera	Kolmogorov-Smirnov	Lilliefors	Ryan-Joiner	Shapiro-Wilk (Royston Exp.)
10	0.1227	0.0506	0	0.0987	0.1039	0.1229
15	0.1846	0.0524	0	0.1319	0.1422	0.2014
20	0.2677	0.0508	0	0.1774	0.2088	0.3094
25	0.3506	0.0526	0	0.2306	0.2727	0.4118
30	0.4327	0.057	0	0.2683	0.3497	0.511
40	0.585	0.0776	0	0.3688	0.5166	0.6983
50	0.7352	0.1155	0	0.4635	0.6963	0.8446
100	0.9842	0.7348	0	1	0.9933	0.9989
200	1	1	0	1	1	1
300	1	1	0	1	1	1
400	1	1	0	1	1	1
500	1	1	0	1	1	1
1000	1	1	0	1	1	1
1500	1	1	0	1	1	1
2000	1	1	0	1	1	1

Table 10 Power Comparison for Beta (3,2) Distribution

			Beta (3,2)			
N	**Anderson-Darling**	**Jarque-Bera**	**Kolmogorov-Smirnov**	**Lilliefors**	**Ryan-Joiner**	**Shapiro-Wilk (Royston Exp.)**
10	0.0523	0.0249	0	0.0491	0.0388	0.0471
15	0.0563	0.0189	0	0.0549	0.0368	0.0548
20	0.0727	0.0147	0	0.063	0.0376	0.0713
25	0.0886	0.0109	0	0.0735	0.0441	0.0874
30	0.099	0.0112	0	0.0822	0.0478	0.103
40	0.1363	0.0081	0	0.103	0.0687	0.1432
50	0.1719	0.0111	0	0.1179	0.0914	0.195
100	0.3819	0.0399	0	1	0.3203	0.5154
200	0.7863	0.616	0	1	0.9231	0.9446
300	0.9624	0.9686	0	1	1	0.9989
400	0.995	0.9994	0	1	1	1
500	0.9991	0.9999	0	1	1	1
1000	1	1	0	1	1	1
1500	1	1	0	1	1	1
2000	1	1	0	1	1	1

Table 11 Power Comparison for χ^2 (20) Distribution

			$two\chi^2(20)$			
N	**Anderson-Darling**	**Jarque-Bera**	**Kolmogorov-Smirnov**	**Lilliefors**	**Ryan-Joiner**	**Shapiro-Wilk (Royston Exp.)**
10	0.0819	0.0788	0.3011	0.3623	0.0888	0.3729
15	0.109	0.117	0.4058	0.4711	0.1248	0.5022
20	0.1308	0.1339	0.5093	0.571	0.1443	0.6124
25	0.1604	0.1723	0.5891	0.6548	0.1894	0.7086
30	0.1863	0.2012	0.6587	0.7192	0.2183	0.777
40	0.2387	0.2583	0.7591	0.8164	0.2852	0.8756
50	0.2792	0.3115	0.8347	0.8805	0.3463	0.9285
100	0.526	0.5825	0.9736	1	0.6528	0.9982
200	0.8521	0.9067	0.9993	1	0.9603	1
300	0.9651	0.9876	1	1	0.9998	1
400	0.9947	0.999	1	1	1	1
500	0.9996	1	1	1	1	1
1000	1	1	1	1	1	1
1500	1	1	1	1	1	1
2000	1	1	1	1	1	1

Journal of Manufacturing Engineering, June 2024, Vol. 19, Issue. 2, pp 040-056
DOI: https://doi.org/10.37255/jme.v19i2pp040-056

Table 12 Power Comparison for Gamma(4,2) Distribution

Gamma(4,2)						
N	**Anderson-Darling**	**Jarque-Bera**	**Kolmogorov-Smirnov**	**Lilliefors**	**Ryan-Joiner**	**Shapiro-Wilk (Royston Exp.)**
10	0.1368	0.1237	0.7823	0.828	0.1426	0.839
15	0.2044	0.1919	0.8947	0.9298	0.2233	0.9422
20	0.2653	0.2381	0.9517	0.9728	0.2931	0.9803
25	0.3401	0.3102	0.9752	0.9875	0.3762	0.9922
30	0.398	0.3647	0.9871	0.9947	0.4491	0.9969
40	0.5109	0.4772	0.9962	0.9989	0.5754	0.9997
50	0.6188	0.5822	0.9996	1	0.6869	1
100	0.9219	0.9103	1	1	0.9649	1
200	0.9987	0.9991	1	1	1	1
300	1	1	1	1	1	1
400	1	1	1	1	1	1
500	1	1	1	1	1	1
1000	1	1	1	1	1	1
1500	1	1	1	1	1	1
2000	1	1	1	1	1	1

Table 13 Power Comparison for χ2 (4) Distribution

$\chi^2(4)$						
N	**Anderson-Darling**	**Jarque-Bera**	**Kolmogorov-Smirnov**	**Lilliefors**	**Ryan-Joiner**	**Shapiro-Wilk (Royston Exp.)**
10	0.2335	0.1899	0.222	0.4208	0.2465	0.4362
15	0.368	0.2948	0.2855	0.5559	0.388	0.5874
20	0.4894	0.3896	0.3053	0.661	0.5178	0.706
25	0.6011	0.4823	0.3006	0.7441	0.6354	0.7993
30	0.6891	0.5709	0.2995	0.8058	0.7265	0.861
40	0.8258	0.705	0.3302	0.891	0.8595	0.942
50	0.9028	0.8184	0.3034	0.9347	0.9351	0.9769
100	0.9989	0.9958	0.2351	1	0.9998	1
200	1	1	0.1216	1	1	1
300	1	1	0.0573	1	1	1
400	1	1	0.0301	1	1	1
500	1	1	0.0182	1	1	1
1000	1	1	0.0001	1	1	1
1500	1	1	0	1	1	1
2000	1	1	0	1	1	1

Journal of Manufacturing Engineering, June 2024, Vol. 19, Issue. 2, pp 040-056
DOI: https://doi.org/10.37255/jme.v19i2pp040-056

Table 14 Power Comparison for Gamma(1,5) Distribution

			Gamma(1,5)			
N	**Anderson-Darling**	**Jarque-Bera**	**Kolmogorov-Smirnov**	**Lilliefors**	**Ryan-Joiner**	**Shapiro-Wilk (Royston Exp.)**
10	0.4287	0.2951	0.2085	0.8084	0.4359	0.8214
15	0.6417	0.4558	0.1472	0.9139	0.6495	0.9276
20	0.7923	0.5916	0.0937	0.9635	0.8082	0.9734
25	0.8941	0.7053	0.0537	0.9851	0.9074	0.9926
30	0.9434	0.7928	0.0347	0.9938	0.9566	0.997
40	0.9878	0.9115	0.0161	0.9989	0.9925	0.9997
50	0.9979	0.9692	0.0061	0.9995	0.999	1
100	1	1	0.0001	1	1	1
200	1	1	0	1	1	1
300	1	1	0	1	1	1
400	1	1	0	1	1	1
500	1	1	0	1	1	1
1000	1	1	0	1	1	1
1500	1	1	0	1	1	1
2000	1	1	0	1	1	1

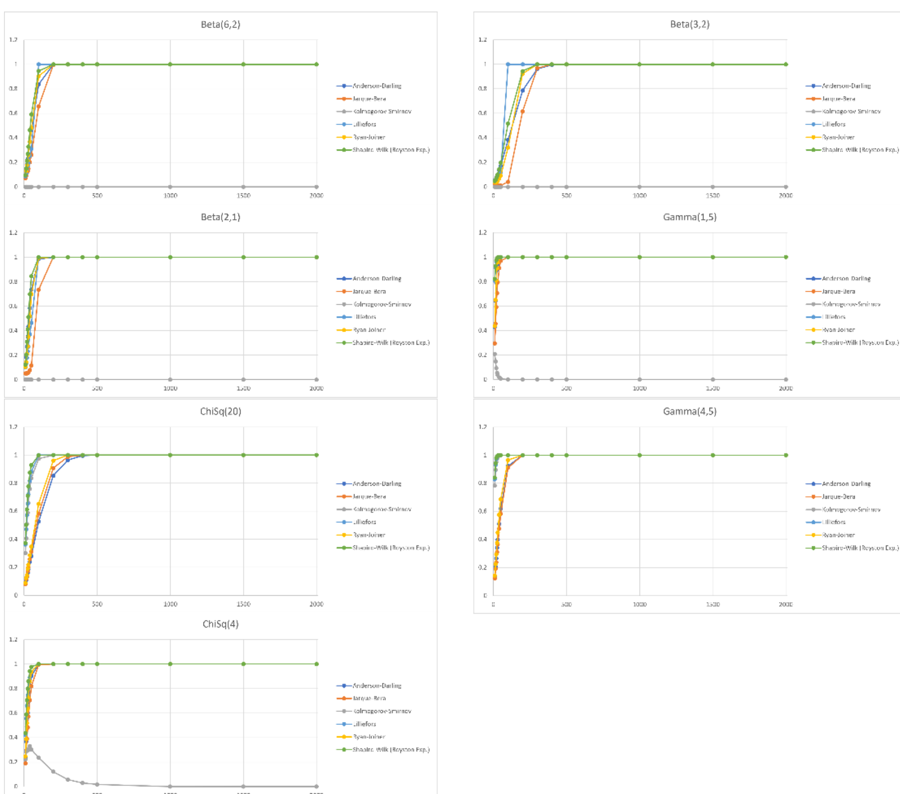

Fig. 3 Collection of Power Comparison Charts for Asymmetric Non-normal Distributions

DOI: https://doi.org/10.37255/jme.v19i2pp040-056

3.4. Power Comparison for General Asymmetric Non-normal Distributions

Table 15 represents the average power for each normality test at each sample size averaged across all asymmetric alternative distributions studied. Table 16 represents the rank of each test at each sample size, with the overall average across sample sizes at the bottom. Figure 4 is the chart generated to visualize the effect of sample size on the average power across all asymmetric distributions studied.

3.5. Average Power Comparison for All Selected Non-normal Distributions

Table 17 represents the average power for each normality test at each sample size averaged across all alternative distributions studied. Table 18 represents the rank of each test at each sample size, with the overall average across sample sizes at the bottom. Figure 5 is the chart generated to visualize the effect of sample size on the power averaged across all distributions studied.

Table 15 Average Power Across All Asymmetric Distributions

Average Power Across All Asymmetric Distributions						
N	**Anderson-Darling**	**Jarque-Bera**	**Kolmogorov-Smirnov**	**Lilliefors**	**Ryan-Joiner**	**Shapiro-Wilk (Royston Exp.)**
10	0.164428571	0.118928571	0.216271429	0.379128571	0.1644	0.3905
15	0.242728571	0.1745	0.2476	0.452157143	0.242042857	0.480714286
20	0.314542857	0.219171429	0.265714286	0.506528571	0.312314286	0.552057143
25	0.380285714	0.265842857	0.274085714	0.549042857	0.377057143	0.608871429
30	0.430728571	0.306657143	0.282857143	0.579771429	0.429757143	0.653357143
40	0.521671429	0.377057143	0.300228571	0.632928571	0.523742857	0.731742857
50	0.595614286	0.438428571	0.306257143	0.6722	0.6049	0.791014286
100	0.807185714	0.702971429	0.315542857	1	0.833571429	0.922614286
200	0.947657143	0.930842857	0.302985714	1	0.983342857	0.992085714
300	0.989628571	0.993742857	0.2939	1	0.999971429	0.999842857
400	0.998528571	0.999771429	0.290014286	1	1	1
500	0.999814286	0.999985714	0.288314286	1	1	1
1000	1	1	0.285728571	1	1	1
1500	1	1	0.285714286	1	1	1
2000	1	1	0.285714286	1	1	1

Table 16 Rank and Average Rank Across All Asymmetric Distributions

Rank and Average Rank Across All Asymmetric Distributions						
N	**Anderson-Darling**	**Jarque-Bera**	**Kolmogorov-Smirnov**	**Lilliefors**	**Ryan-Joiner**	**Shapiro-Wilk (Royston Exp.)**
10	4	6	3	2	5	1
15	4	6	3	2	5	1
20	3	6	5	2	4	1
25	3	6	5	2	4	1
30	3	5	6	2	4	1
40	4	5	6	2	3	1
50	4	5	6	2	3	1
100	4	5	6	1	3	2
200	4	5	6	1	3	2
300	5	4	6	1	2	3
400	5	4	6	1	1	1
500	5	4	6	1	1	1
1000	1	1	6	1	1	1
1500	1	1	6	1	1	1
2000	1	1	6	1	1	1
Avg	3.4	4.266666667	5.466666667	1.466666667	2.733333333	1.266666667
Avg Rank	4	5	6	2	3	1

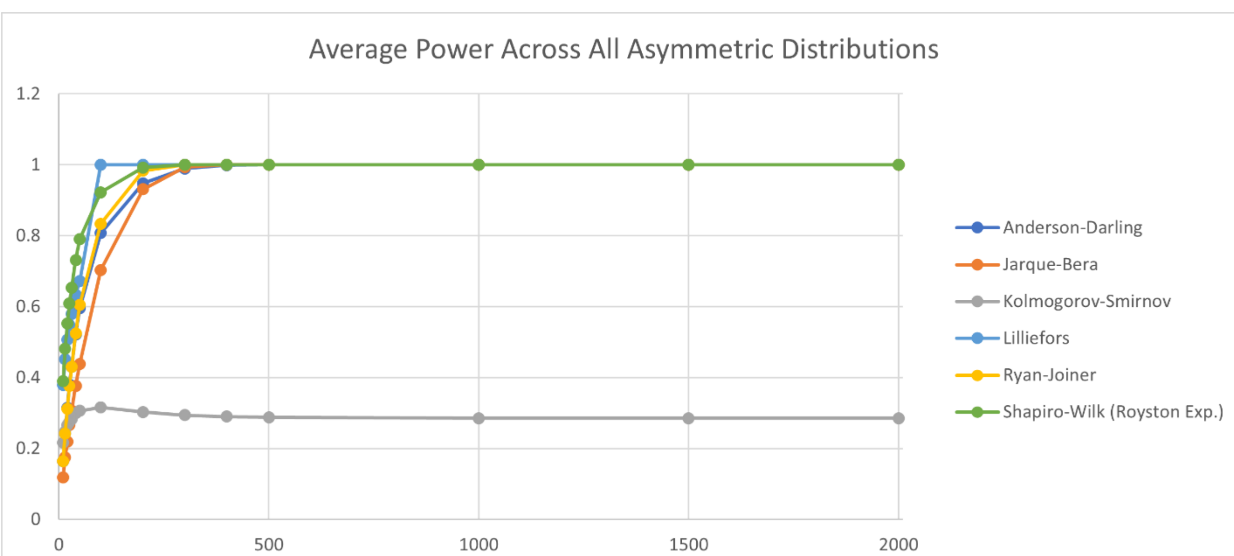

Fig. 4 Average Power Across All Asymmetric Distributions

Table 17 Average Power Across All Distributions Studied

Average Power Across All Distributions Studied						
N	**Anderson-Darling**	**Jarque-Bera**	**Kolmogorov-Smirnov**	**Lilliefors**	**Ryan-Joiner**	**Shapiro-Wilk (Royston Exp.)**
10	0.124875	0.099366667	0.14585	0.527275	0.126341667	0.5519
15	0.177158333	0.14225	0.154116667	0.595608333	0.181033333	0.616
20	0.22295	0.174908333	0.159033333	0.632083333	0.22755	0.659708333
25	0.266116667	0.210125	0.161633333	0.658366667	0.272208333	0.69415
30	0.2982	0.239908333	0.165675	0.676958333	0.307275	0.72105
40	0.360333333	0.292716667	0.175316667	0.7086	0.3737	0.769541667
50	0.411275	0.339975	0.17865	0.732633333	0.432	0.807666667
100	0.576083333	0.531908333	0.184066667	1	0.619183333	0.910566667
200	0.733708333	0.763208333	0.176741667	1	0.828358333	0.990116667
300	0.807433333	0.855166667	0.171441667	1	0.895883333	0.999783333
400	0.841958333	0.880258333	0.169175	1	0.9426	1
500	0.858533333	0.890416667	0.168183333	1	0.9931	1
1000	0.900625	0.913508333	0.166675	1	1	1
1500	0.913691667	0.917283333	0.166666667	1	1	1
2000	0.916616667	0.917125	0.166666667	1	1	1

Table 18 Rank and Average Rank Across All Distributions Studied

Rank and Average Rank Across All Distributions Studied						
N	**Anderson-Darling**	**Jarque-Bera**	**Kolmogorov-Smirnov**	**Lilliefors**	**Ryan-Joiner**	**Shapiro-Wilk (Royston Exp.)**
10	5	6	3	2	4	1
15	4	6	5	2	3	1
20	4	5	6	2	3	1
25	4	5	6	2	3	1
30	4	5	6	2	3	1
40	4	5	6	2	3	1
50	4	5	6	2	3	1
100	4	5	6	1	3	2
200	5	4	6	1	3	2
300	5	4	6	1	3	2
400	5	4	6	1	3	1
500	5	4	6	1	3	1
1000	5	4	6	1	1	1
1500	5	4	6	1	1	1
2000	5	4	6	1	1	1
Avg	4.533333333	4.666666667	5.733333333	1.466666667	2.666666667	1.2
Avg Rank	4	5	6	2	3	1

Journal of Manufacturing Engineering, June 2024, Vol. 19, Issue. 2, pp 040-056
DOI: https://doi.org/10.37255/jme.v19i2pp040-056

4. Discussion

The effect of sample size showed a dramatic increase in the test's power as the sample size increased. The effect of the underlying distribution of the data impacted the power of the tests, with all tests exhibiting a higher power with distributions that are less similar to the normal distribution. In a manufacturing setting, the underlying distribution of data cannot truly be known. Therefore, it is generally more helpful to distinguish between symmetric data and non-symmetric data, which could allow organizations to have a separate sequence of tests based on whether the data is determined to be symmetric or asymmetric. If an organization decides to discriminate between symmetric and asymmetric data, a standard benchmark for symmetry must be determined.

When comparing the results of this study to those reported by Razali et al. [9], the study which this one was designed to replicate, the results are similar. In general, the Shapiro-Wilk (Royston) test is the most powerful, however for some sample sizes between 100 & 300-400, the Lilliefors will outperform the Shapiro-Wilk (Royston), but very shortly after (400 samples), the Shapiro-Wilk (Royston) converges to 100% rejection of the null hypothesis, and both are equally effective, at 100% power. Razali et al. [9] did not include the Ryan Joiner Test in their study, so no comparison can be made as to its placement. Using the results of this study, it can be concluded that the Ryan-Joiner is, in general, a good third option because it also converges to 1.0 around 400 to 500 samples. The Anderson-Darling and Jarque-Bera tests performed very similarly throughout all sample size levels, with the Anderson-Darling performing slightly better at smaller sample sizes (up to about 100-200) and the Jarque-Bera being more powerful at sample sizes larger than that. Neither the Anderson-Darling nor Jarque-Bera converged to 1.0 by the 2,000 sample trials, instead seeming to level off at a power of about 0.92. The Kolmogorov-Smirnov test performed the worst of all tests studied by a large margin. This is likely due to the significance level ($\alpha = 0.05$) because when looking at the raw output of the p-values, most were close to 0.10 – 0.15. However, some p-values were extremely high, and due to the evidence that other tests are extremely powerful, the KS test is not recommended for extensive use.

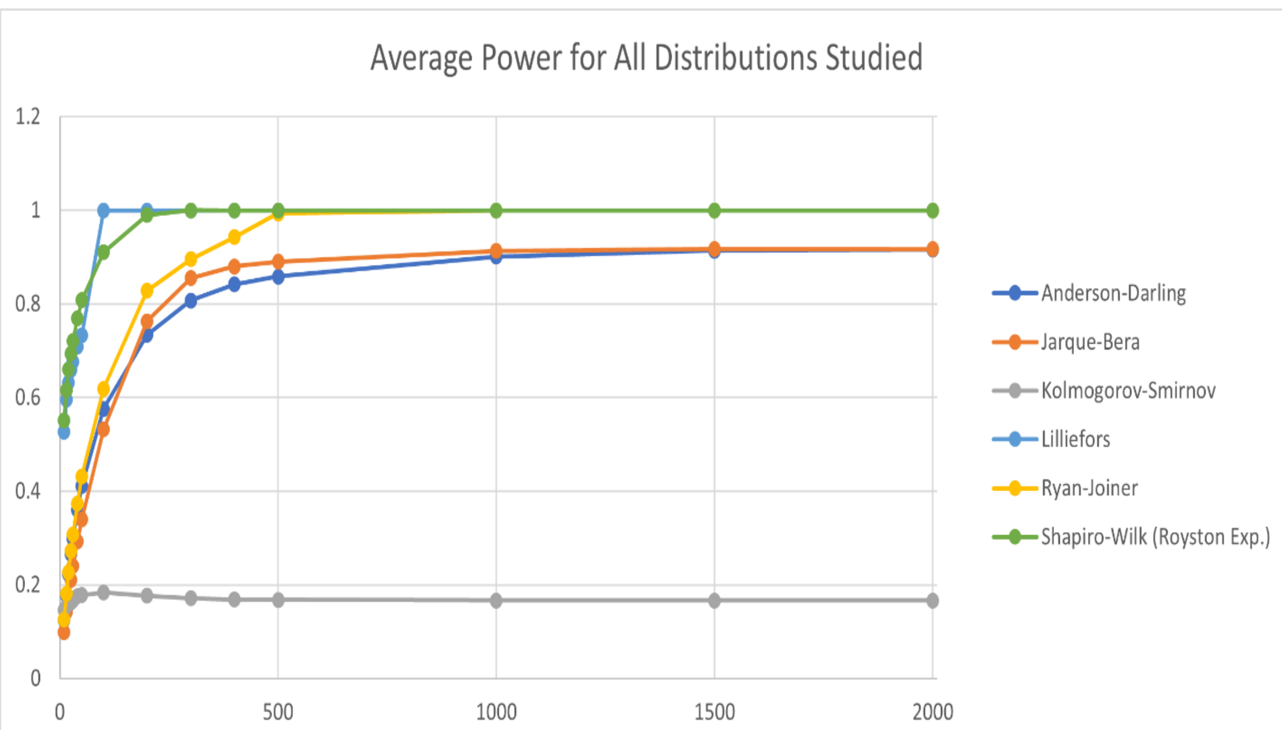

Fig. 5 Average Power for All Distributions Studied

5. Conclusion

Based on the average rankings across distributions and sample sizes, the following general orders, Table 19, have been proposed. These general orders are broken down into three columns: the suggested order for symmetric distributions, the suggested order for asymmetric distributions, and the suggested order for general distributions, disregarding symmetry. There is no difference in the testing order of the first three tests for all distributions studied. The recommended order of the first three tests, regardless of distribution or sample size, is: 1. Shapiro-Wilk(Royston), 2. Lilliefors, and 3. Ryan-Joiner.

Table 19 Proposed Testing Order of Normality Tests

Proposed series (general sample size)			
Rank	**Symmetric distributions**	**Asymmetric distributions**	**General distribution**
1	Shapiro-Wilk (Royston Exp.)	Shapiro-Wilk (Royston Exp.)	Shapiro-Wilk (Royston Exp.)
2	Lilliefors	Lilliefors	Lilliefors
3	Ryan-Joiner	Ryan-Joiner	Ryan-Joiner
4	Jarque-Bera	Anderson-Darling	Anderson-Darling
5	Anderson-Darling	Jarque-Bera	Jarque-Bera
6	Kolmogorov-Smirnov	Kolmogorov-Smirnov	Kolmogorov-Smirnov

References

1. *Amedie, W., Awaj, Y., and Singh, A., "Quality Improvement Using Statistical Process Control Tools in Glass Bottles Manufacturing Company," International Journal for Quality Research, vol. 7, no. 1, pp. 107-126, 2013.*

2. *Black, K., "Business statistics: for contemporary decision making," John Wiley & Sons, 2023.*

3. *Ghasemi, A., and Zahediasl, S., "Normality tests for statistical analysis: a guide for non-statisticians," International Journal of Endocrinol Metab, vol. 10, no. 2, pp. 486-9, 2012. doi: 10.5812/ijem.3505.*

4. *He, Q., and Wang, J., "Statistics pattern analysis: A new process monitoring framework and its application to semiconductor batch processes," AIChE Journal, vol. 57, no. 1, pp. 107-121, 2011. https://doi.org/10.1002/aic.12345.*

5. *Jarque, C. M., and Bera, A. K., "A Test for Normality of Observations and Regression Residuals," International Statistical Review, vol. 55, pp. 163-172, 1987. http://doi.org/10.2307/1403192.*

6. *Lee, C., Jeong, J., and Park, S., "Comprehensive comparison of normality tests: Empirical study using many different types of data," Journal of the Korean Data & Information Science Society, vol. 27, no. 5, pp. 1399-1412, 2016. https://doi.org/10.7465/jkdi.2016.27.5.1399.*

7. *Lieber, R. L., "Statistical significance and statistical power in hypothesis testing," Journal of Orthopaedic Research, vol. 8, no. 2, pp. 304-309, 1990. https://doi.org/10.1002/jor.1100080221.*

8. *Lilliefors, H., "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," Journal of the American Statistical Association, vol. 62, no. 318, pp. 399-402, 1967. https://doi.org/10.2307/2283970.*

9. *Razali, N., and Yap, B., "Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests," Journal of Statistical Modeling and Analytics, vol. 2, 2011.*

10. *Thode, H. J., "Testing for normality," Marcel Dekker, 2002.*

11. *Wijekularathna, D., Manage, A., and Scariano, S., "Power analysis of several normality tests: A Monte Carlo simulation study," Communications in Statistics - Simulation and Computation, vol. 51, no. 3, pp. 757-773, 2022. https://doi.org/10.1080/03610918.2019.1658780.*

12. *Yazici, B., and Asma, S., "A comparison of various tests of normality," Journal of Statistical Computation and Simulation, vol. 77, pp. 175-183, 2007. https://doi.org/10.1080/10629360600678310.*