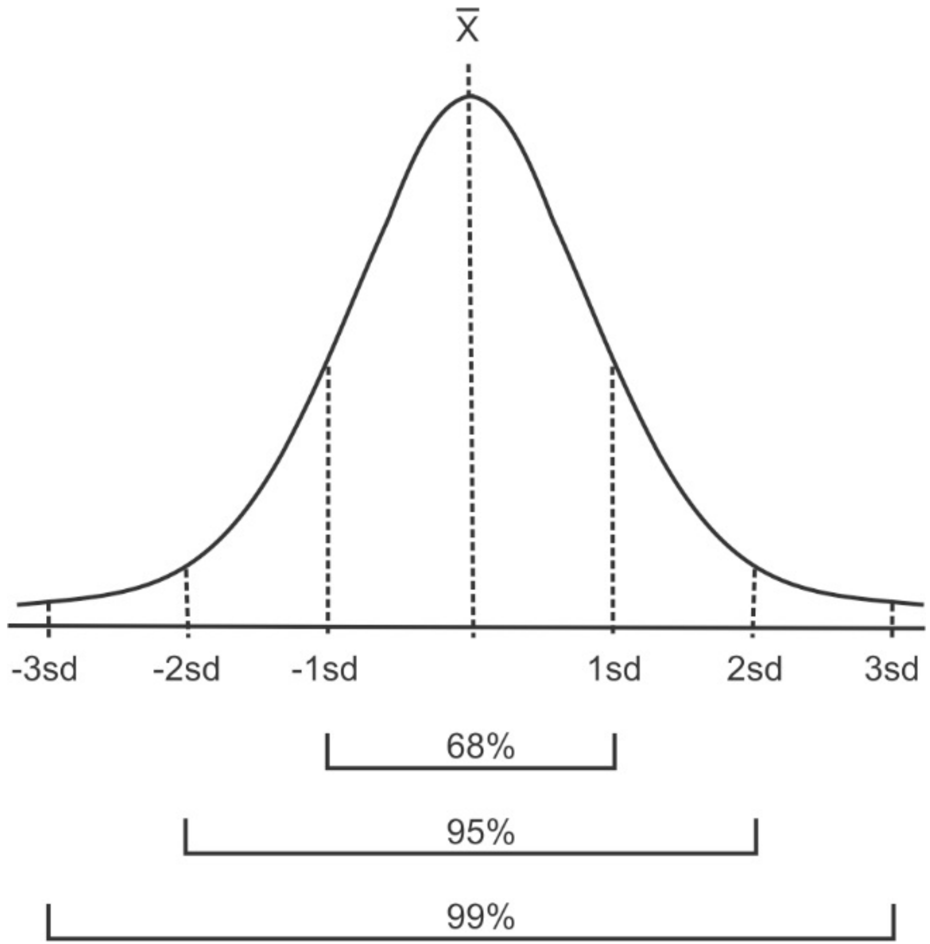
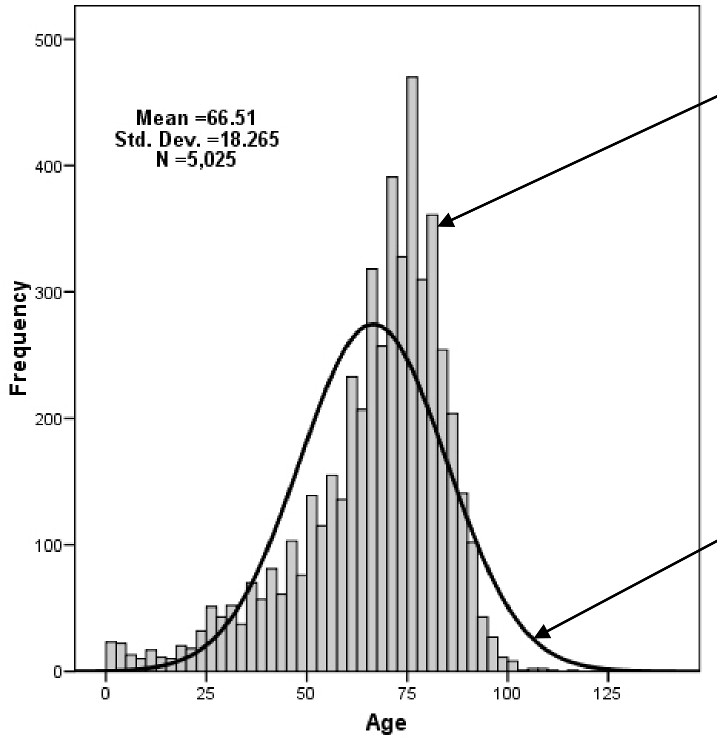
Testing for Normality

For each mean and standard deviation combination a theoretical normal distribution can be determined. This distribution is based on the proportions shown below.



This theoretical normal distribution can then be compared to the actual distribution of the data.



The ***actual*** data distribution that has a mean of 66.51 and a standard deviation of 18.265.

Theoretical normal distribution ***calculated*** from a mean of 66.51 and a standard deviation of 18.265.

Are the actual data statistically different than the computed normal curve?

There are several methods of assessing whether data are normally distributed or not. They fall into two broad categories: *graphical* and *statistical*. The some common techniques are:

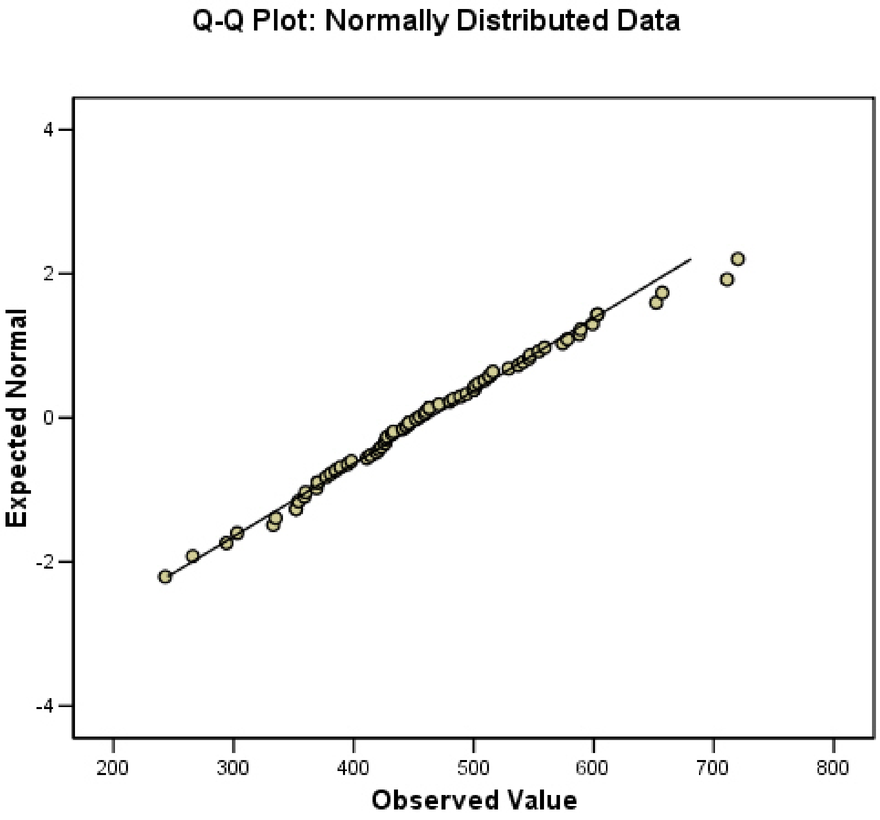
Graphical

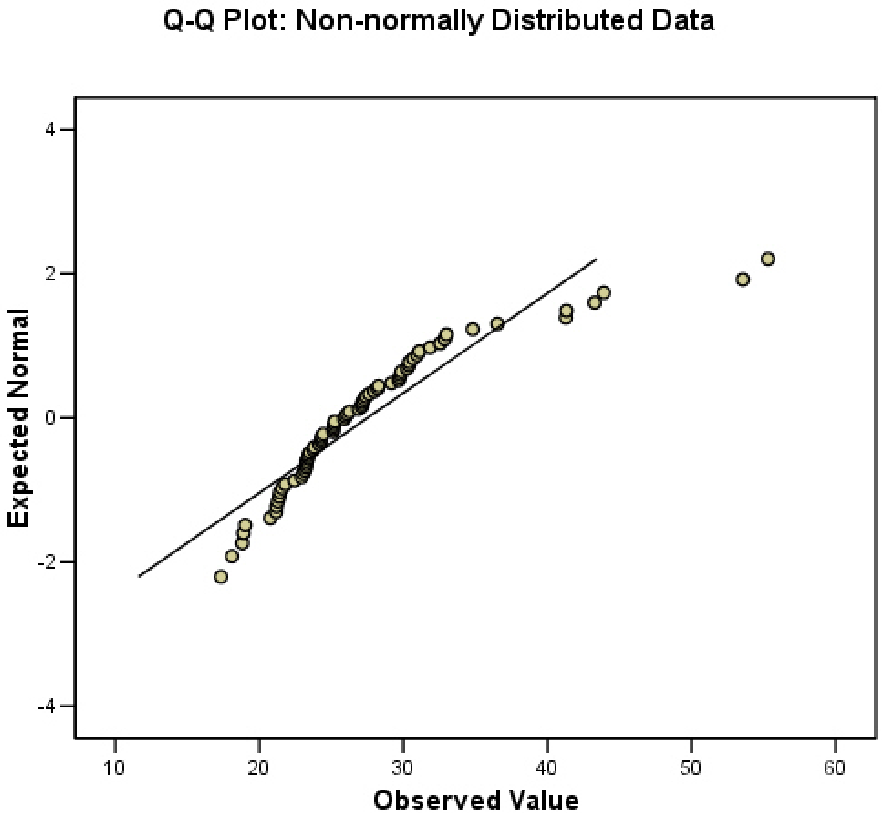
- Q-Q probability plots
- Cumulative frequency (P-P) plots

Statistical

- W/S test
- Jarque-Bera test
- Shapiro-Wilks test
- Kolmogorov-Smirnov test
- D’Agostino test

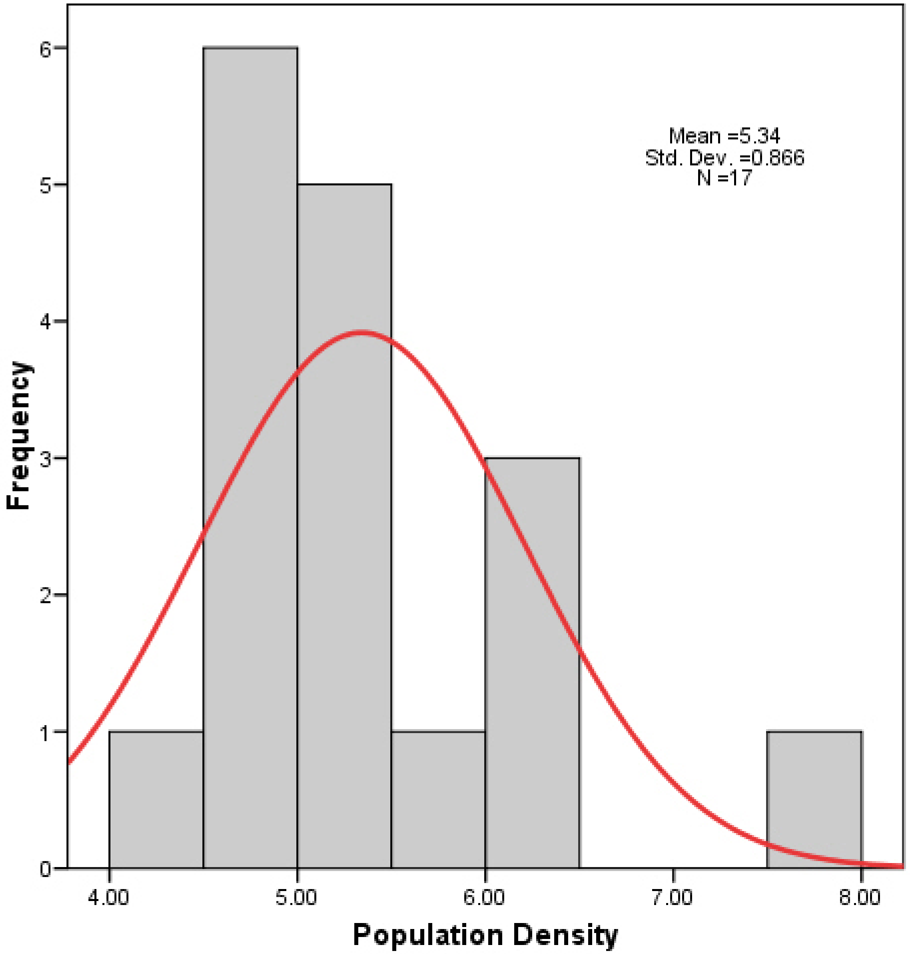
Q-Q plots display the observed values against normally distributed data (represented by the line).

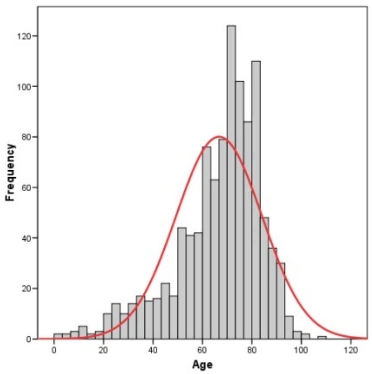


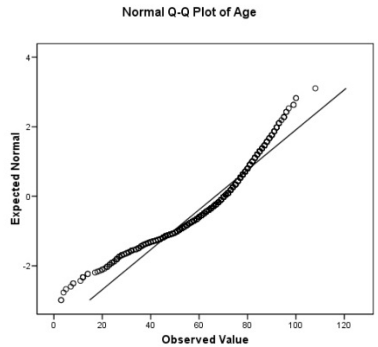


Normally distributed data fall along the line.

Graphical methods are typically not very useful when the sample size is small. This is a histogram of the last example. These data do not ‘look’ normal, but they are not statistically different than normal.



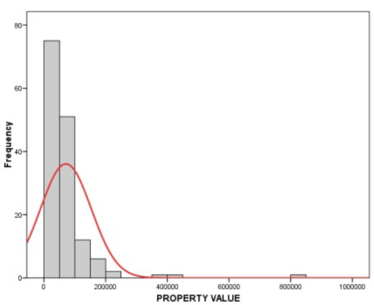


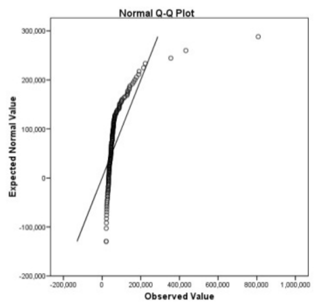


Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Age	.110	1048	.000	.931	1048	.000

a. Lilliefors Significance Correction

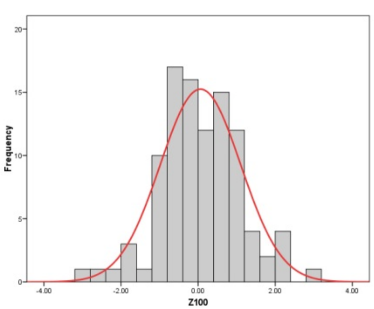


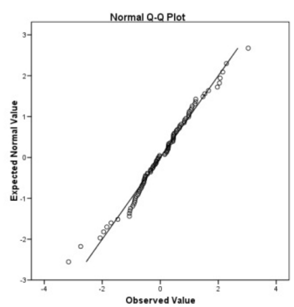


Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
TOTAL_VALU	.283	149	.000	.463	149	.000

a. Lilliefors Significance Correction





Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Z100	.071	100	.200*	.985	100	.333

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Statistical tests for normality are more precise since actual probabilities are calculated.

Tests for normality calculate the probability that the sample was drawn from a normal population.

The hypotheses used are:

Ho: The sample data are not significantly different than a normal population.

Ha: The sample data are significantly different than a normal population.

When testing for normality:

- Probabilities > 0.05 indicate that the data are normal.
- Probabilities < 0.05 indicate that the data are NOT normal.

Normally Distributed Data

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Asthma Cases	.069	72	.200*	.988	72	.721

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

In SPSS output above the probabilities are greater than 0.05 (the typical alpha level), so we accept H_o… these data are not different from normal.

Non-Normally Distributed Data

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Average PM10	.142	72	.001	.841	72	.000

a. Lilliefors Significance Correction

In the SPSS output above the probabilities are less than 0.05 (the typical alpha level), so we reject H_o… these data are significantly different from normal.

Simple Tests for Normality

W/S Test for Normality

- A fairly simple test that requires only the sample standard deviation and the data range.

- Should not be confused with the Shapiro-Wilk test.

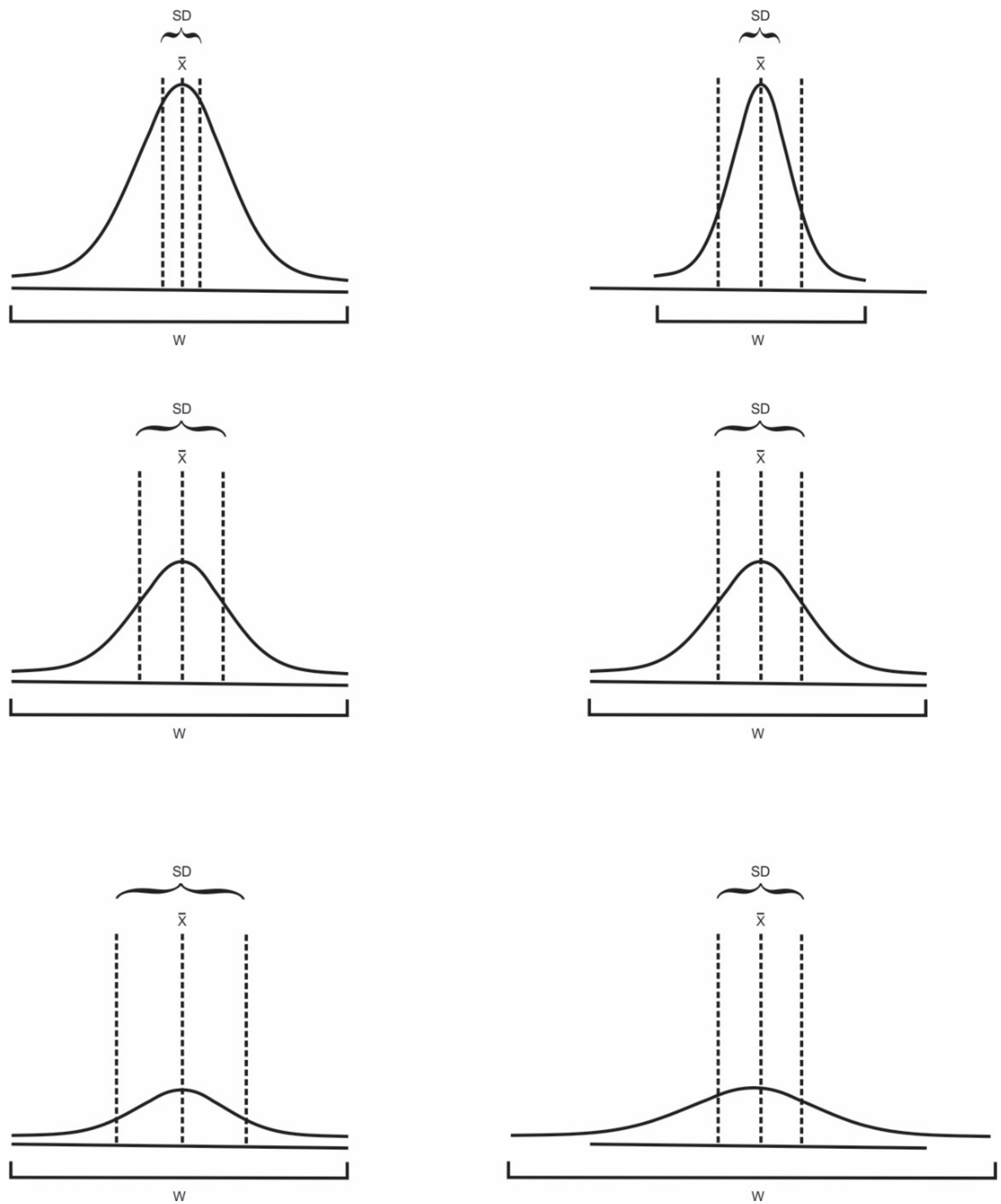
- Based on the q statistic, which is the ‘studentized’ (meaning t distribution) range, or the range expressed in standard deviation units.

$$q = \frac{w}{s}$$

where *q* is the test statistic, *w* is the range of the data and *s* is the standard deviation.

- The test statistic *q* (Kanji 1994, table 14) is often reported as *u* in the literature.

Range constant,
SD changes



Range changes,
SD constant

Village	Pop Density
Ajuno	5.11
Angahuan	5.15
Arantepacua	5.00
Aranza	4.13
Charapan	5.10
Cheran	5.22
Cocucho	5.04
Comachuen	5.25
Corupo	4.53
Ihuatzio	5.74
Janitzio	6.63
Jaracuaro	5.73
Nahuatzen	4.77
Nurio	6.06
Paracho	4.82
Patzcuaro	4.98
Pichataro	5.36
Pomacuaran	4.96
Quinceo	5.94
Quiroga	5.01
San Felipe	4.10
San Lorenzo	4.69
Sevina	4.97
Tingambato	5.01
Turicuaro	6.19
Tzintzuntzan	4.67
Urapicho	6.30

Standard deviation (s) = 0.624
Range (w) = 2.53
n = 27

$$q = \frac{w}{s}$$

$$q = \frac{2.53}{0.624} = 4.05$$

$$q_c = 3.34\ to\ 4.71$$

The W/S test uses a critical range. If the calculated value falls ***within*** the range, then accept H_o. If the calculated value falls ***outside*** the range then reject H_o.

Since 3.34 < *q=4.05* < 4.71, we accept H_o.

Since n = 27 is not on the table, we will use the next LOWER value.

Critical Values of q for the W/S Normality Test

Taken from Kanji, 1994 Table 14

Columns *a* denote the lower boundaries or the left-sided critical values.
Columns *b* denote the upper boundaries or the right-sided critical values.

n	Level of significance α											
	0.000		0.005		0.01		0.025		0.05		0.10	
	a	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*
3	1.732	2.000	1.735	2.000	1.737	2.000	1.745	2.000	1.758	1.999	1.782	1.997
4	1.732	2.449	1.82	2.447	1.87	2.445	1.93	2.439	1.98	2.429	2.04	2.409
5	1.826	2.828	1.98	2.813	2.02	2.803	2.09	2.782	2.15	2.753	2.22	2.712
6	1.826	3.162	2.11	3.115	2.15	3.095	2.22	3.056	2.28	3.012	2.37	2.949
7	1.871	3.464	2.22	3.369	2.26	3.338	2.33	3.282	2.40	3.222	2.49	3.143
8	1.871	3.742	2.31	3.585	2.35	3.543	2.43	3.471	2.50	3.399	2.59	3.308
9	1.897	4.000	2.39	3.772	2.44	3.720	2.51	3.634	2.59	3.552	2.68	3.449
10	1.897	4.243	2.46	3.935	2.51	3.875	2.59	3.777	2.67	3.685	2.76	3.57
11	1.915	4.472	2.53	4.079	2.58	4.012	2.66	3.903	2.74	3.80	2.84	3.68
12	1.915	4.690	2.59	4.208	2.64	4.134	2.72	4.02	2.80	3.91	2.90	3.78
13	1.927	4.899	2.64	4.325	2.70	4.244	2.78	4.12	2.86	4.00	2.96	3.87
14	1.927	5.099	2.70	4.431	2.75	4.34	2.83	4.21	2.92	4.09	3.02	3.95
15	1.936	5.292	2.74	4.53	2.80	4.44	2.88	4.29	2.97	4.17	3.07	4.02
16	1.936	5.477	2.79	4.62	2.84	4.52	2.93	4.37	3.01	4.24	3.12	4.09
17	1.944	5.657	2.83	4.70	2.88	4.60	2.97	4.44	3.06	4.31	3.17	4.15
18	1.944	5.831	2.87	4.78	2.92	4.67	3.01	4.51	3.10	4.37	3.21	4.21
19	1.949	6.000	2.90	4.85	2.96	4.74	3.05	4.56	3.14	4.43	3.25	4.27
20	1.949	6.164	2.94	4.91	2.99	4.80	3.09	4.63	3.18	4.49	3.29	4.32
25	1.961	6.93	3.09	5.19	3.15	5.06	3.24	4.87	3.34	4.71	3.45	4.53
30	1.966	7.62	3.21	5.40	3.27	5.26	3.37	5.06	3.47	4.89	3.59	4.70
35	1.972	8.25	3.32	5.57	3.38	5.42	3.48	5.21	3.58	5.04	3.70	4.84
40	1.975	8.83	3.41	5.71	3.47	5.56	3.57	5.34	3.67	5.16	3.79	4.96
45	1.978	9.38	3.49	5.83	3.55	5.67	3.66	5.45	3.75	5.26	3.88	5.06
50	1.980	9.90	3.56	5.93	3.62	5.77	3.73	5.54	3.83	5.35	3.95	5.14
55	1.982	10.39	3.62	6.02	3.69	5.86	3.80	5.63	3.90	5.43	4.02	5.22
60	1.983	10.86	3.68	6.10	3.75	5.94	3.86	5.70	3.96	5.51	4.08	5.29
65	1.985	11.31	3.74	6.17	3.80	6.01	3.91	5.77	4.01	5.57	4.14	5.35
70	1.986	11.75	3.79	6.24	3.85	6.07	3.96	5.83	4.06	5.63	4.19	5.41
75	1.987	12.17	3.83	6.30	3.90	6.13	4.01	5.88	4.11	5.68	4.24	5.46
80	1.987	12.57	3.88	6.35	3.94	6.18	4.05	5.93	4.16	5.73	4.28	5.51
85	1.988	12.96	3.92	6.40	3.99	6.23	4.09	5.98	4.20	5.78	4.33	5.56
90	1.989	13.34	3.96	6.45	4.02	6.27	4.13	6.03	4.24	5.82	4.36	5.60
95	1.990	13.71	3.99	6.49	4.06	6.32	4.17	6.07	4.27	5.86	4.40	5.64
100	1.990	14.07	4.03	6.53	4.10	6.36	4.21	6.11	4.31	5.90	4.44	5.68
150	1.993	17.26	4.32	6.82	4.38	6.64	4.48	6.39	4.59	6.18	4.72	5.96
200	1.995	19.95	4.53	7.01	4.59	6.84	4.68	6.60	4.78	6.39	4.90	6.15
500	1.998	31.59	5.06	7.60	5.13	7.42	5.25	7.15	5.47	6.94	5.49	6.72
1000	1.999	44.70	5.50	7.99	5.57	7.80	5.68	7.54	5.79	7.33	5.92	7.11

Source: Sachs, 1972

Since we have a critical range, it is difficult to determine a probability range for our results. Therefore we simply state our alpha level.

The sample data set is not significantly different than normal ($q_{4.05}$, $p > 0.05$).

D'Agostino Test

- A very powerful test for departures from normality.
- Based on the D statistic, which gives an upper and lower critical value.

$$D = \frac{T}{\sqrt{n^3 SS}} \qquad T = \sum \left(i - \frac{n+1}{2} \right) x_i$$

where D is the test statistic, SS is the sum of squares of the data and n is the sample size, and *i* is the order or rank of observation x. The df for this test is n (sample size).

- First the data are ordered from smallest to largest or largest to smallest.

Village	Pop Density	i	Deviates²
San Felipe	4.10	1	1.2218
Aranza	4.13	2	1.1505
Corupo	4.53	3	0.4582
Tzintzuntzan	4.67	4	0.2871
San Lorenzo	4.69	5	0.2583
Nahuatzen	4.77	6	0.1858
Paracho	4.82	7	0.1441
Pomacuaran	4.96	8	0.0604
Sevina	4.97	9	0.0538
Patzcuaro	4.98	10	0.0509
Arantepacua	5.00	11	0.0401
Tingambato	5.01	12	0.0359
Quiroga	5.01	13	0.0354
Cocucho	5.04	14	0.0250
Charapan	5.10	15	0.0111
Ajuno	5.11	16	0.0090
Angahuan	5.15	17	0.0026
Cheran	5.22	18	0.0003
Comachuen	5.25	19	0.0027
Pichataro	5.36	20	0.0253
Jaracuaro	5.73	21	0.2825
Ihuatzio	5.74	22	0.2874
Quinceo	5.94	23	0.5456
Nurio	6.06	24	0.7398
Turicuaro	6.19	25	0.9697
Urapicho	6.30	26	1.2062
Janitzio	6.63	27	2.0269
	Mean = 5.2		**SS = 10.12**

$\bar{x} = 5.2$ $\quad$ SS $= 10.12$ $\quad$ df $= 27$

← $(4.13 - 5.2)^2 = 1.1505$

$$\frac{n+1}{2} = \frac{27+1}{2} = 14$$

$$T = \sum (i - 14) x_i$$

$$T = (1 - 14)4.10 + (2 - 14)4.13 \ldots + (27 - 14)6.63$$

$$T = 122.04$$

$$D = \frac{122.04}{\sqrt{(27^3)(10.12)}} = \frac{122.04}{446.31} = 0.2734$$

$$D_c = 0.2647 \; to \; 0.2866$$

$0.2647 > D = 0.2734 > 0.2866$ Accept Ho.

The village population density is not significantly different than normal ($D_{0.2243}$, $p > 0.05$).

Use the next lower *n* on the table if the sample size is NOT listed.

Critical Values for the D'Agostino D Normality Test
Taken from Zar, 1981 Table B.22

n	$\alpha =$ 0.20	0.10	0.05	0.02	0.01
10	0.2632, 0.2835	0.2573, 0.2843	0.2513, 0.2849	0.2436, 0.2855	0.2379, 0.2857
12	0.2653, 0.2841	0.2598, 0.2849	0.2544, 0.2854	0.2473, 0.2859	0.2420, 0.2862
14	0.2669, 0.2846	0.2618, 0.2853	0.2568, 0.2858	0.2503, 0.2862	0.2455, 0.2865
16	0.2681, 0.2848	0.2634, 0.2855	0.2587, 0.2860	0.2527, 0.2865	0.2482, 0.2867
18	0.2690, 0.2850	0.2646, 0.2855	0.2603, 0.2862	0.2547, 0.2866	0.2505, 0.2868
20	0.2699, 0.2852	0.2657, 0.2857	0.2617, 0.2863	0.2564, 0.2867	0.2525, 0.2869
22	0.2705, 0.2853	0.2670, 0.2859	0.2629, 0.2864	0.2579, 0.2869	0.2542, 0.2870
24	0.2711, 0.2853	0.2675, 0.2860	0.2638, 0.2865	0.2591, 0.2870	0.2557, 0.2871
26	0.2717, 0.2854	0.2682, 0.2861	0.2647, 0.2866	0.2603, 0.2870	0.2570, 0.2872
28	0.2721, 0.2854	0.2688, 0.2861	0.2655, 0.2866	0.2612, 0.2870	0.2581, 0.2873
30	0.2725, 0.2854	0.2693, 0.2861	0.2662, 0.2866	0.2622, 0.2871	0.2592, 0.2872
32	0.2729, 0.2854	0.2698, 0.2862	0.2668, 0.2867	0.2630, 0.2871	0.2600, 0.2873
34	0.2732, 0.2854	0.2703, 0.2862	0.2674, 0.2867	0.2636, 0.2871	0.2609, 0.2873
36	0.2735, 0.2854	0.2707, 0.2862	0.2679, 0.2867	0.2643, 0.2871	0.2617, 0.2873
38	0.2738, 0.2854	0.2710, 0.2862	0.2683, 0.2867	0.2649, 0.2871	0.2623, 0.2873
40	0.2740, 0.2854	0.2714, 0.2862	0.2688, 0.2867	0.2655, 0.2871	0.2630, 0.2874
42	0.2743, 0.2854	0.2717, 0.2861	0.2691, 0.2867	0.2659, 0.2871	0.2636, 0.2874
44	0.2745, 0.2854	0.2720, 0.2861	0.2695, 0.2867	0.2664, 0.2871	0.2641, 0.2874
46	0.2747, 0.2854	0.2722, 0.2861	0.2698, 0.2866	0.2668, 0.2871	0.2646, 0.2874
48	0.2749, 0.2854	0.2725, 0.2861	0.2702, 0.2866	0.2672, 0.2871	0.2651, 0.2874
50	0.2751, 0.2853	0.2727, 0.2861	0.2705, 0.2866	0.2676, 0.2871	0.2655, 0.2874
60	0.2757, 0.2852	0.2737, 0.2860	0.2717, 0.2865	0.2692, 0.2870	0.2673, 0.2873
70	0.2763, 0.2851	0.2744, 0.2859	0.2726, 0.2864	0.2708, 0.2869	0.2687, 0.2872
80	0.2768, 0.2850	0.2750, 0.2857	0.2734, 0.2863	0.2713, 0.2868	0.2698, 0.2871
90	0.2771, 0.2849	0.2755, 0.2856	0.2740, 0.2862	0.2721, 0.2866	0.2707, 0.2870
100	0.2774, 0.2849	0.2759, 0.2855	0.2745, 0.2860	0.2727, 0.2865	0.2714, 0.2869
120	0.2779, 0.2847	0.2765, 0.2853	0.2752, 0.2858	0.2737, 0.2863	0.2725, 0.2866
140	0.2782, 0.2846	0.2770, 0.2852	0.2758, 0.2856	0.2744, 0.2862	0.2734, 0.2865
160	0.2785, 0.2845	0.2774, 0.2851	0.2763, 0.2855	0.2750, 0.2860	0.2741, 0.2863
180	0.2787, 0.2844	0.2777, 0.2850	0.2767, 0.2854	0.2755, 0.2859	0.2746, 0.2862
200	0.2789, 0.2843	0.2779, 0.2848	0.2770, 0.2853	0.2759, 0.2857	0.2751, 0.2860
250	0.2793, 0.2841	0.2784, 0.2846	0.2776, 0.2850	0.2767, 0.2855	0.2760, 0.2858
300	0.2796, 0.2840	0.2788, 0.2844	0.2781, 0.2848	0.2772, 0.2853	0.2766, 0.2855
350	0.2798, 0.2839	0.2791, 0.2843	0.2784, 0.2847	0.2776, 0.2851	0.2771, 0.2853
400	0.2799, 0.2838	0.2793, 0.2842	0.2787, 0.2845	0.2780, 0.2849	0.2775, 0.2852
450	0.2801, 0.2837	0.2795, 0.2841	0.2789, 0.2844	0.2782, 0.2848	0.2778, 0.2851
500	0.2802, 0.2836	0.2796, 0.2840	0.2791, 0.2843	0.2785, 0.2847	0.2780, 0.2849
600	0.2804, 0.2835	0.2799, 0.2839	0.2794, 0.2842	0.2788, 0.2845	0.2784, 0.2847
700	0.2805, 0.2834	0.2800, 0.2838	0.2796, 0.2840	0.2791, 0.2844	0.2787, 0.2846
800	0.2806, 0.2833	0.2802, 0.2837	0.2798, 0.2839	0.2793, 0.2842	0.2790, 0.2844
900	0.2807, 0.2833	0.2803, 0.2836	0.2799, 0.2838	0.2795, 0.2841	0.2792, 0.2843
1000	0.2808, 0.2832	0.2804, 0.2835	0.2800, 0.2838	0.2796, 0.2840	0.2793, 0.2842
1250	0.2809, 0.2831	0.2806, 0.2834	0.2803, 0.2836	0.2799, 0.2839	0.2797, 0.2840
1500	0.2810, 0.2830	0.2807, 0.2833	0.2805, 0.2835	0.2801, 0.2837	0.2799, 0.2839
1750	0.2811, 0.2830	0.2808, 0.2832	0.2806, 0.2834	0.2803, 0.2836	0.2801, 0.2838
2000	0.2812, 0.2829	0.2809, 0.2831	0.2807, 0.2833	0.2804, 0.2835	0.2802, 0.2837

For each significance level, α is given a pair of critical values. If the calculated D is $\leq$ the first member of the pair, or $\geq$ the second, then, the null hypothesis of population normality is rejected.

Breaking down the equations:

$$\frac{n+1}{2} = \frac{17+1}{2} = 9$$

← This is the 'middle' of the data set.

$$T = \sum (i - 9) X_i$$

$(i - 9)$ ← This is the observation's distance from the middle.

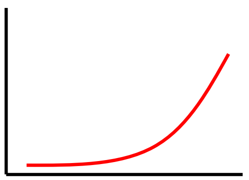
X_i ← This is the observation, and is used to 'weight' the result based on the size of the observation and its distance.

$$D = \frac{T}{\sqrt{n^3 SS}}$$

T ← This represents which tail is more pronounced (- for left, + for right).

SS ← This is the dataset's total squared variation.

n^3 ← This adjusts for sample size like this:



$\sqrt{\;}$ ← This transforms the squared values from SS.

Village	Pop Density	i	Deviates^2	*T*
San Felipe	4.10	1	1.2218	-53.26
Aranza	4.13	2	1.1505	-49.56
Corupo	4.53	3	0.4582	-49.78
Tzintzuntzan	4.67	4	0.2871	-46.67
San Lorenzo	4.69	5	0.2583	-42.25
Nahuatzen	4.77	6	0.1858	-38.17
Paracho	4.82	7	0.1441	-33.76
Pomacuaran	4.96	8	0.0604	-29.74
Sevina	4.97	9	0.0538	-24.85
Patzcuaro	4.98	10	0.0509	-19.91
Arantepacua	5.00	11	0.0401	-15.01
Tingambato	5.01	12	0.0359	-10.03
Quiroga	5.01	13	0.0354	-5.01
Cocucho	5.04	14	0.0250	0.00
Charapan	5.10	15	0.0111	5.10
Ajuno	5.11	16	0.0090	10.21
Angahuan	5.15	17	0.0026	15.45
Cheran	5.22	18	0.0003	20.88
Comachuen	5.25	19	0.0027	26.27
Pichataro	5.36	20	0.0253	32.17
Jaracuaro	5.73	21	0.2825	40.14
Ihuatzio	5.74	22	0.2874	45.91
Quinceo	5.94	23	0.5456	53.47
Nurio	6.06	24	0.7398	60.63
Turicuaro	6.19	25	0.9697	68.06
Urapicho	6.30	26	1.2062	75.61
Janitzio	6.63	27	2.0269	86.14

→ -418.00

→ 540.04

These data are more heavily weighted in the positive (right) tail…

$$540.04 - 418.00 = 122.04$$

but not enough to conclude the data are different than normal.

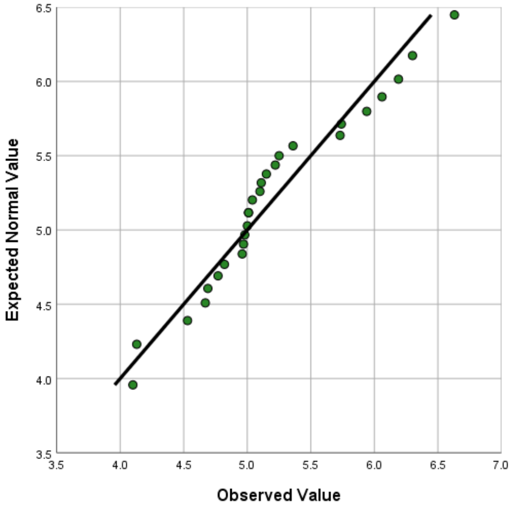
Normality tests using various random normal sample sizes:

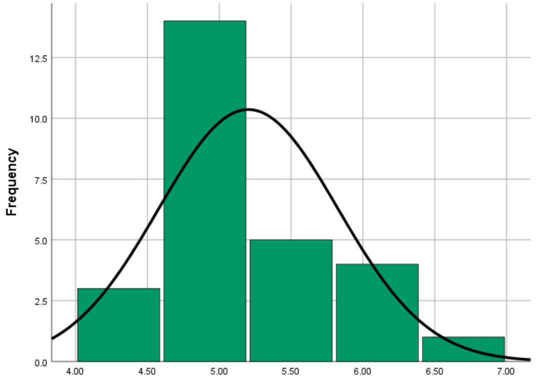
Sample Size	JB Prob
10	0.6667
50	0.5649
100	0.5357
200	0.5106
500	0.4942
1000	0.4898
2000	0.4823
5000	0.4534
7000	0.3973
10000	0.2948

Notice that as the sample size increases, the probabilities decrease. In other words, ***it gets harder*** to meet the normality assumption as the sample size increases since even small departures from normality are detected.

Different normality tests produce different probabilities. This is due to where in the distribution (central, tails) or what moment (skewness, kurtosis) they are examining.

Normality Test	Statistic	Probability	Results
W/S	4.05	> 0.05	Normal
Jarque-Bera	1.209	0.5463	Normal
D'Agostino	0.2734	> 0.05	Normal
Shapiro-Wilk	0.9428	0.1429	Normal
Kolmogorov-Smirnov	1.73	0.0367	Not-normal
Anderson-Darling	0.7636	0.0412	Not-normal
Lilliefors	0.1732	0.0367	Not-normal





W/S or studentized range (q):

- Simple, very good for symmetrical distributions and short tails.
- Very bad with asymmetry.

Shapiro Wilk (W):

- Fairly powerful omnibus test. Not good with small samples or discrete data.
- Good power with symmetrical, short and long tails. Good with asymmetry.

Jarque-Bera (JB):

- Good with symmetric and long-tailed distributions.
- Less powerful with asymmetry, and poor power with bimodal data.

D'Agostino (D or Y):

- Good with symmetric and very good with long-tailed distributions.
- Less powerful with asymmetry.

Anderson-Darling (A):

- Similar in power to Shapiro-Wilk but has less power with asymmetry.
- Works well with discrete data.

Distance tests (Kolmogorov-Smirnov, Lillifors, Chi^2):

- All tend to have lower power. Data have to be very non-normal to reject Ho.
- These tests can outperform other tests when using discrete or grouped data.

When is non-normality a problem?

- Normality can be a problem when the sample size is small (< 50).

- Highly skewed data create problems.

- Highly leptokurtic data are problematic, but not as much as skewed data.

- Normality becomes a serious concern when there is “activity” in the tails of the data set.

  - Outliers are a problem.

  - “Clumps” of data in the tails are worse.

SPSS Normality Tests

Analyze > Descriptive Statistics > Explore, then *Plots > Normality Tests with Plots*.

Available tests: Kolmogorov-Smirnov and Shapiro-Wilk.

PAST Normality Tests

Univariate > Normality Tests

Available tests: Shapiro-Wilk, Anderson-Darling, Lilliefors, Jarque-Bera.

Final Words Concerning Normality Testing:

1. Since it IS a test, state a null and alternate hypothesis.

2. If you perform a normality test, do not ignore the results.

3. If the data are not normal, use non-parametric tests.

4. If the data are normal, use parametric tests.

AND MOST IMPORTANTLY:

5. **If you have groups of data, you MUST test each group for normality.**

Testing for Outliers

Grubbs Test

$$G_{Max} = \frac{x_n - \bar{x}}{s} \quad or \quad G_{Min} = \frac{\bar{x} - x_n}{s}$$

$df = n$

where x_n is the suspected outlier, $\bar{x}$ is the mean, and s is the standard deviation. G_{Max} is used when the suspect observation is greater than the mean and G_{Min} is used when it is less than the mean.

Ho: The suspected outlier is not different than the sample distribution.
Ha: The suspected outlier is different than the sample distribution.

Obs
15
7
6
6
5
5
5
4
4
3

$$G_{Max} = \frac{15 - 6}{3.37} = 2.671$$

The critical value for an n = 10 from Grubbs modified t table (G table) at an α = 0.05 is 2.18.

Since 2.671 > 2.18, reject Ho.

The suspected outlier is from a significantly different sample population (G_{Max}, 2.671, $p < 0.01$).

Critical Values of Grubb's Outlier (G) Test
Taken from Grubb 1969, Table 1

N	α=0.05	α=0.025	α=0.01
3	1.15	1.15	1.15
4	1.46	1.48	1.49
5	1.67	1.71	1.75
6	1.82	1.89	1.94
7	1.94	2.02	2.10
8	2.03	2.13	2.22
9	2.11	2.21	2.32
10	2.18	2.29	2.41
11	2.23	2.36	2.48

Dixon Test

$$Q = \frac{x_n - x_{n-1}}{x_n - x_1}$$

$df = n$

where x_n is the suspected outlier, x_{n-1} is the next ranked observation, and x_1 is the last ranked observation.

Ho: The suspected outlier is not different than the sample distribution.
Ha: The suspected outlier is different than the sample distribution.

Obs
15
7
6
6
5
5
5
4
4
3

$$Q = \frac{15 - 7}{15 - 3} = 0.6667$$

The critical value for an n = 10 from Verma and Quiroz-Ruiz expanded Dixon table at an $\alpha = 0.05$ is 0.4122. Since $0.6667 > 0.4122$, reject H_o.

The suspected outlier is from a significantly different sample population ($Q_{0.6667}$, $p < 0.005$).

Critical Values of Expanded Dixon Outlier Test

Taken from Verma and Quiroz-Ruiz, Table 2

n	CL	70%	80%	90%	**95%**	98%	**99%**	99.5%
	SL	30%	20%	10%	**5%**	2%	**1%**	0.5%
	α	0.30	0.20	0.10	**0.05**	0.02	**0.01**	0.005
3		0.6836	0.7808	0.8850	0.9411	0.9763	0.9881	0.9940
4		0.4704	0.5603	0.6789	0.7651	0.8457	0.8886	0.9201
5		0.3730	0.4508	0.5578	0.6423	0.7291	0.7819	0.8234
6		0.3173	0.3868	0.4840	0.5624	0.6458	0.6987	0.7437
7		0.2811	0.3444	0.4340	0.5077	0.5864	0.6371	0.6809
8		0.2550	0.3138	0.3979	0.4673	0.5432	0.5914	0.6336
9		0.2361	0.2915	0.3704	0.4363	0.5091	0.5554	0.5952
10		0.2208	0.2735	0.3492	0.4122	0.4813	0.5260	0.5658
11		0.2086	0.2586	0.3312	0.3922	0.4591	0.5028	0.5416
12		0.1983	0.2467	0.3170	0.3755	0.4405	0.4831	0.5208

These tests have several requirements:

1) The data are from a normal distribution
2) There are not multiple outliers (3+),
3) The data are sorted with the suspected outlier first.

If 2 observations are suspected as being outliers and both lie on the same side of the mean, this test can be performed again after removing the first outlier from the data set.

Caution must be used when removing outliers. Only remove outliers if you suspect the value was caused by an error of some sort, or if you have evidence that the value truly belongs to a different population.

If you have a small sample size, extreme caution should be used when removing any data.