Minitab®

SPC

LESSON: Standardized Control Charts for Measurement Data with Varying Sample Sizes

This lesson includes an overview of the subject, instructor notes, and example exercises using Minitab.

Build $\bar{x}$ and R charts given Variable Sample Sizes, such as $n_1 = 4$, $n_2 = 3$, $n_3 = 3$, $n_4 = 5$, etc.

What if the sample size n changes from sample to sample? How do we compute control chart limits when n is changing with each sample?

Let's first investigate this problem with Minitab. We'll go back to the cardboard thickness problem.

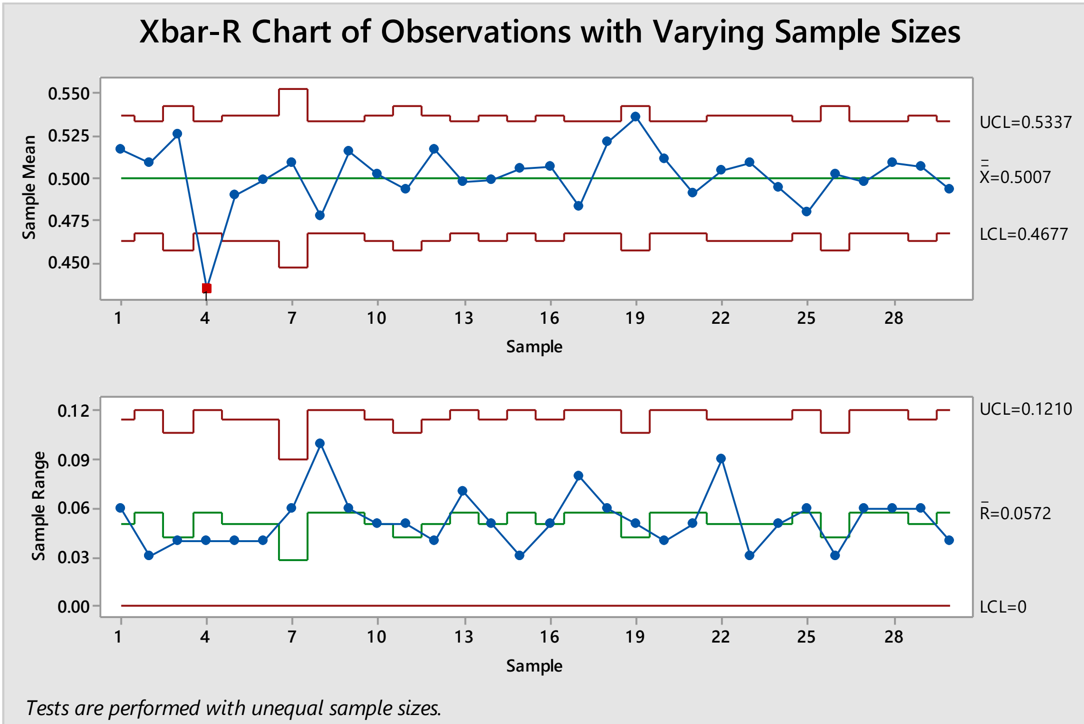
A paper manufacturing company has to control the thickness of cardboard sheets produced. Random samples of size ***n* = 2, 3, 4, and 5** are selected and the thickness is recorded. The k = 30 samples and corresponding sample sizes are in the Minitab worksheet (Minitab-Lesson11-DATA_VariableSampleSize) for this assignment: Columns C10-C14

Obs 1	Obs 2	Obs 3	Obs 4	Obs 5	Means	Sample Sizes
0.52	0.55	0.49	0.51	*	0.517500	4
0.53	0.50	0.51	0.51	0.50	0.510000	5
0.52	0.51	0.55	*	*	0.526667	3
0.42	0.45	0.43	0.42	0.46	0.436000	5
...	...	...	...	...	...	...

(a) What could cause the sample sizes to be different at each sampling time?

- Cost of sampling at each sample time
- Workers available at each sample to take measurements
- Some of the cardboard at a sampling may be damaged
- Not enough time at some collection interval to take the maximum number of samples
- Sample how many easily are available; for convenience
- Workers are told to take samples, but size isn't indicated
- Supplier sends you varying numbers of items

(b) Construct an Xbar-R chart and copy and paste it in below. Select the “Xbar-*R* Options” menu and use $\bar{R}$ to estimate the process standard deviation. Select the columns containing the observations: Obs 1 through Obs 5,



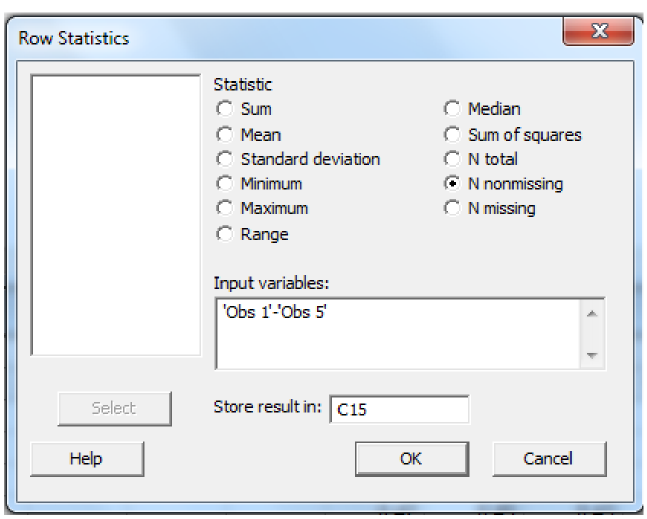
To compute $\bar{\bar{X}}$ for variable sample sizes; we need to take a weighted average of the Xbar’s:

$$\bar{\bar{X}} = \frac{\sum_{i=1}^{k} n_i \bar{X}_i}{\sum_{i=1}^{k} n_i}$$

Note that *k* represents the number of SUBGROUPS we are using, usually 20, 25, or 30.

Note: To make a column of sample sizes in **Minitab**, use

Calc > **Row Statistics**, and select **N non-missing**



(c) Compute $\overline{R}$, which is the weighted average of the ranges. Minitab uses d_2 for $n = 5$ (which is the maximum of all of the sample sizes) to determine an estimate of $\hat{\sigma}$.

$$\overline{R} = \frac{\sum_{i=1}^{k} n_i R_i}{\sum_{i=1}^{k} n_i}$$

Build Xbar and R charts by converting sample means to z-scores

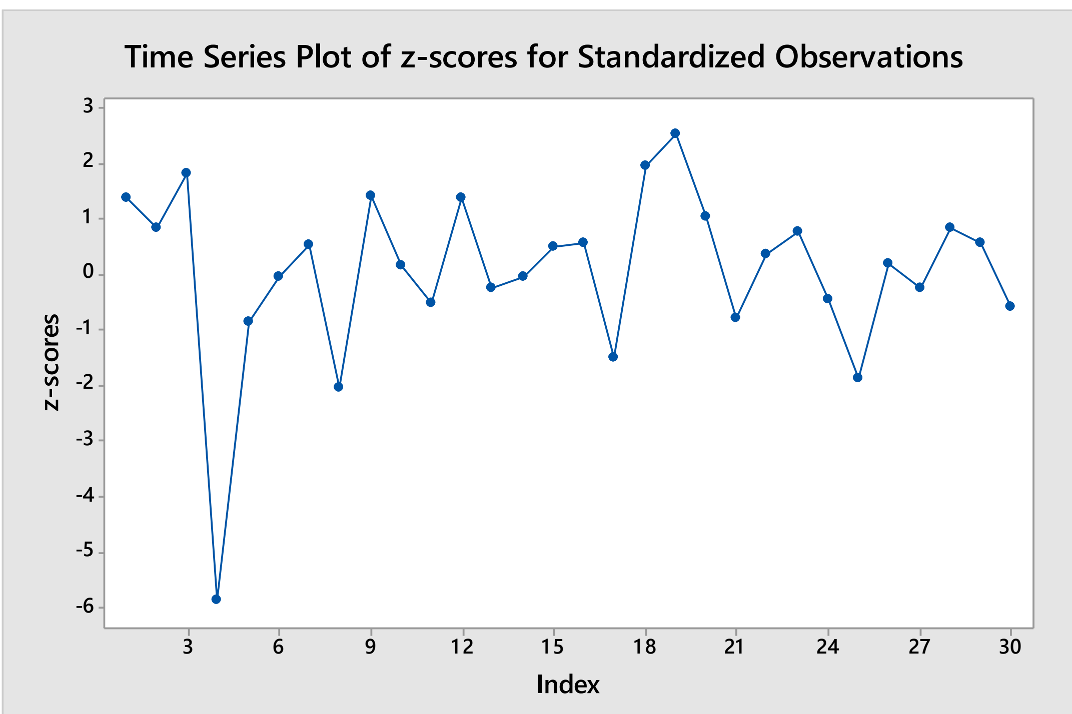
(a) Let's start all over again, but this time we'll make a control chart that's easy to read! First, we'll turn our sample means into z-scores.

We already have the mean of the means, an estimate of true process mean $\hat{\mu}$, and an estimate of the true process standard deviation $\hat{\sigma}$ (from above). So, in order to standardize our sample means, we just ed the sample size of each trial, which is in column "Sample Sizes."
So, use Minitab's calculator feature to make a column of z-scores by standardizing each sample mean.

First 5 **z-scores: 1.3659, 0.8453, 1.8283, -5.8810, -0.8699**

(b) Plot your z-scores on a basic time series plot in **Minitab**.

Graph > **Time Series Plot** > **Simple**



(c) Because you are working with z-scores on a control chart, what should the values of the upper and lower control limits be?

Solution: 3 and -3

(d) Are there any out of control points according to Rule 1? If so, how many standard deviations is it above or below the target value?

Solution: Yes, one out of control value; **observation 4 is -5.88 standard deviations below the target**.

(e) The formula for converting ranges to standardized ranges is stated as:

First, the range R_i for sample i is divided by the approximate process standard deviation, which was estimated by Rbar and d_2 for n = 5. That is,

$$ri = R_i / \hat{\sigma}.$$

The values of ri are standardized by subtracting its mean d2 and then dividing that entire quantity by its standard deviation d_3. That is, standardized range = $(r - d_2) / d_3$.

Compute the first two standardized ranges for trials 1 and 2.

Solution: $r_1 = R_1/\hat{\sigma} = 0.06/0.024 = 2.5$; standardized $r_1 = (2.5 - 2.059)/0.879 \cong$ **0.502**

$r_2 = R_2/\hat{\sigma} = 0.03/0.024 = 1.25$; standardized $r_2 = (1.25 - 2.326)/0.864 \cong$ **-1.245**

A standardized control chart makes it possible to calculate Type I error for other rules besides Rule 1.

With fluctuating (variable) control limits, the rules for identifying out of control conditions that we discussed in earlier in the quarter become difficult to apply – except Rule 1

If your process is deemed out of control by Rule 1 on the standardized control chart, Rule 1 will also show up on the varying sample size control chart for Xbar and *R*. So, if you're just looking for Type I error by Rule 1, then go ahead with your control chart for variable sample sizes.

We would use a standardized chart so that we could STILL detect the OTHER out-of-control situations (Rules 2 – 8), which we can't do with charts with varying subgroups. (When control limits are fluctuating, how do we know if 4 of 5 points are beyond 2σ?)

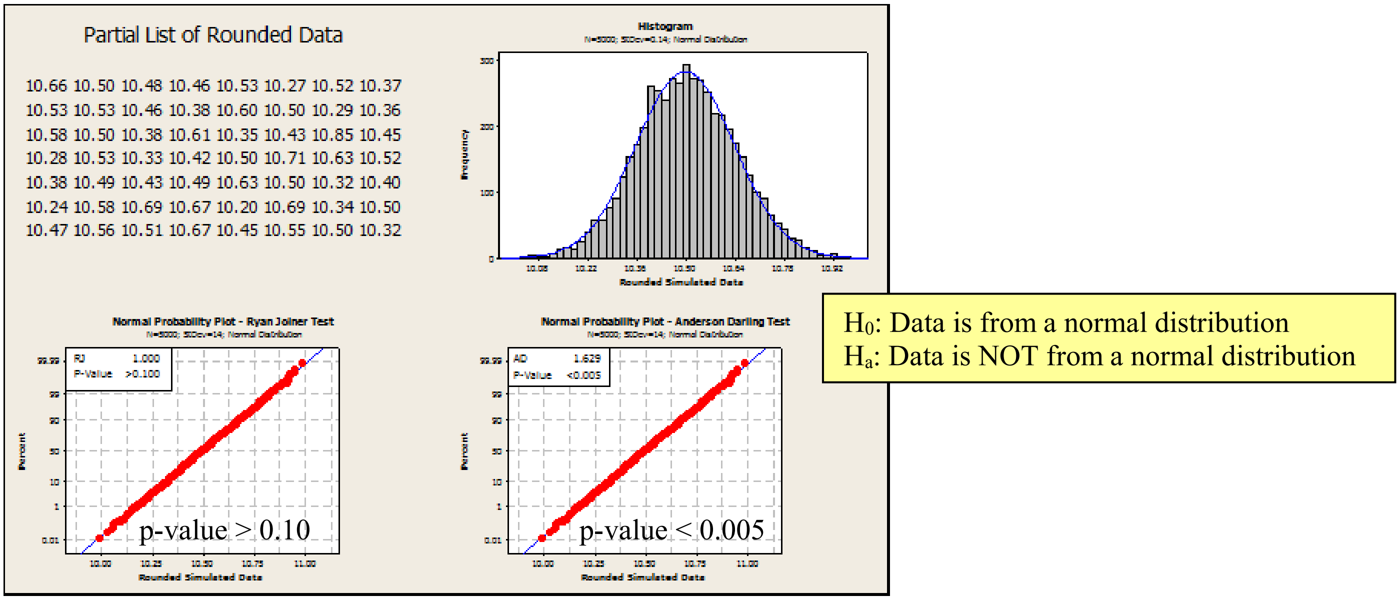
Normality Tests and Rounding; Jim Colton (Minitab), 12/4/2013

All measurements are rounded to some degree. In most cases, you would not want to reject normality just because the data are rounded. In fact, the normal distribution would be a quite desirable model for the data if the underlying distribution is normal since it would smooth out the discreteness in the rounded measurements.]

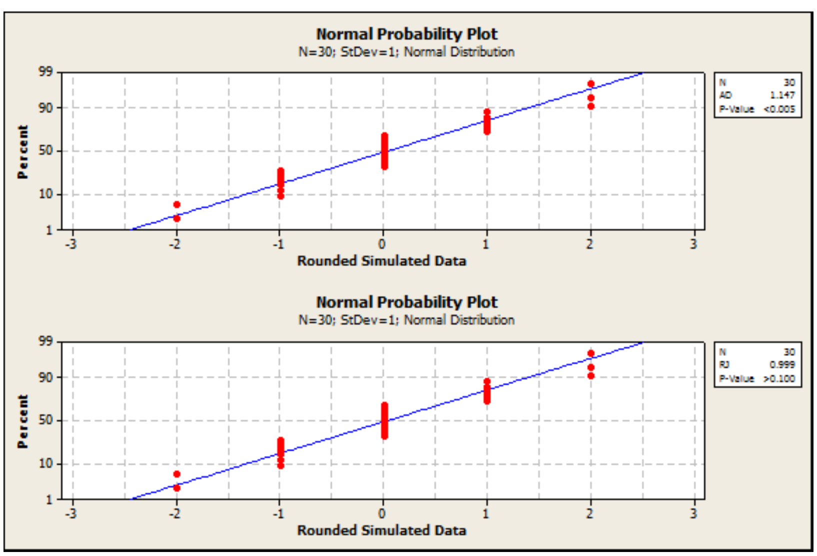
Some normality tests reject a very high percentage of time due to rounding when the underlying distribution is normal (Anderson-Darling and Kolmogorov-Smirnov), **while others seem to ignore the rounding** (Ryan-Joiner and Chi-Square).

As an extreme example of how data that is very well modeled by a normal distribution can get rejected, consider a sample size of 5000 from a normal distribution with mean 10 and standard deviation of 0.14.

The display below shows a partial list of the data rounded to the nearest 100th, a histogram, and probability plots of that same data. The histogram and probability plots look great. The Ryan-Joiner Test passes Normality with a p-value above 0.10 (probability plot on the left). However, the Anderson-Darling p-value is below 0.005 (probability plot on the right). Clearly, rejecting Normality in a case like this is inappropriate.

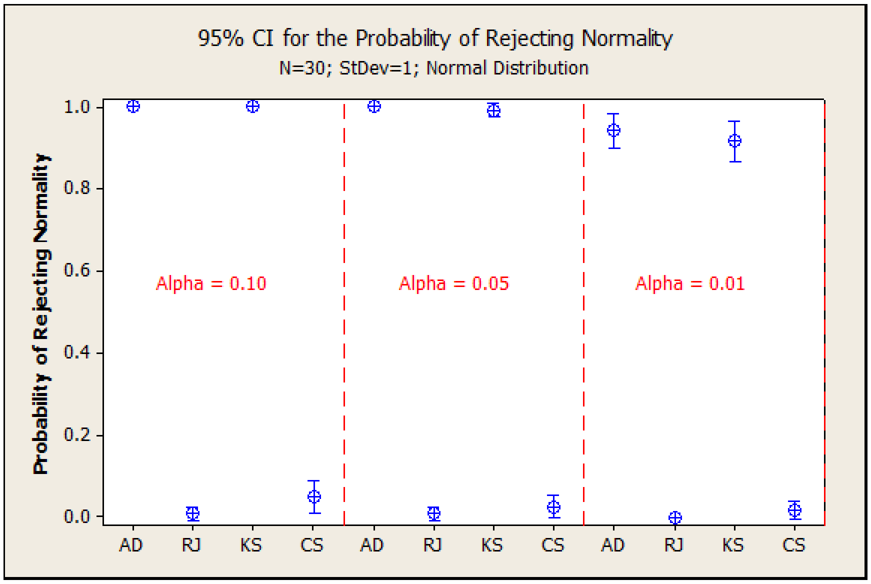


A simulation was conducted to address a more common sample size, $n = 30$. Data were simulated from a normal distribution with mean 0 and standard deviation 1, then rounded to the nearest integer. An example of a probability plot from this simulation appears below.

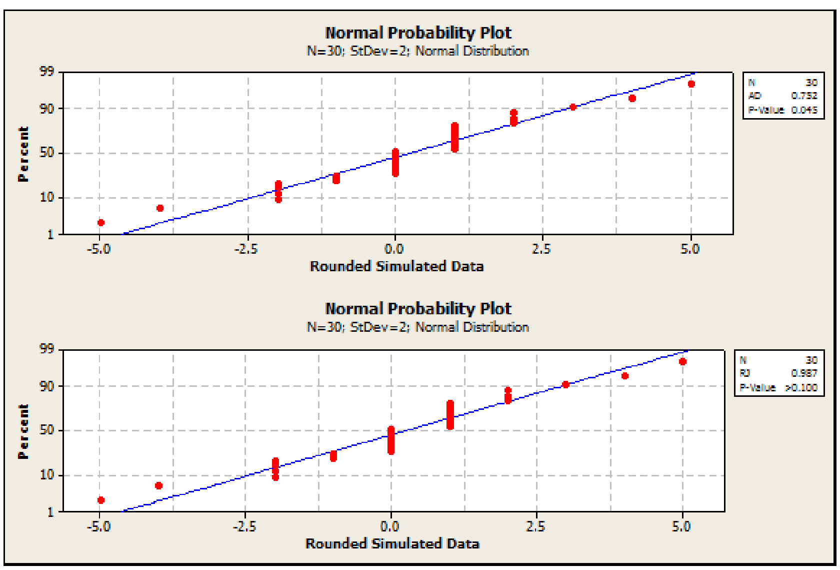


In this iteration of the simulation, the Anderson-Darling P-value was less than 0.005 while the Ryan-Joiner P-value was greater than 0.10.

The simulation results were remarkably consistent, with the Anderson-Darling (AD) test almost always rejecting normality and the Ryan-Joiner (RJ) test almost always failing to reject normality. The Kolmogorov-Smirnov (KS) and Chi-square (CS) tests were included in the simulation too. The CS test was almost as good as the RJ test at avoiding rejecting normality due to rounding.



A second simulation was conducted with less extreme rounding*. Data were simulated from a normal distribution with mean 0 and standard deviation 2, then rounded to the nearest integer. An example of a probability plot from this simulation is below. In this iteration of the simulation, the Anderson-Darling P-value was less than 0.05 while the Ryan-Joiner P-value was greater than 0.10.



In this second simulation with less extreme rounding, the AD and KS tests did not reject as often.

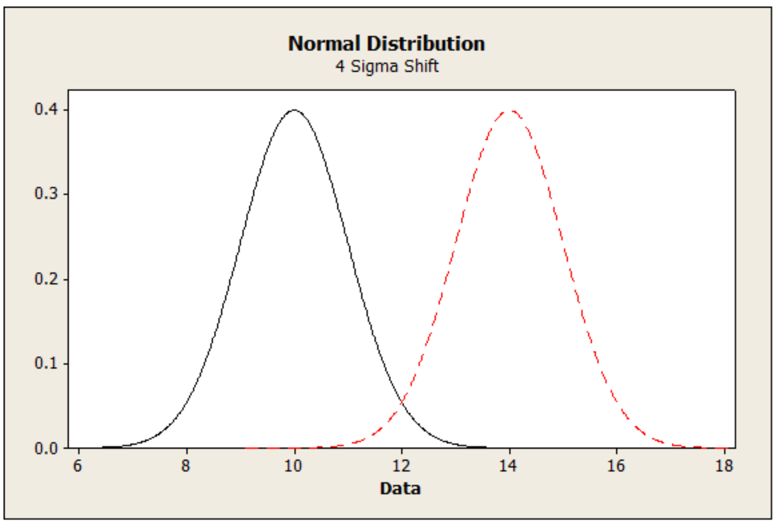
Anderson-Darling, Ryan-Joiner, or Kolmogorov-Smirnov:

Which Normality Test Is the Best? Jim Colton (Minitab), 10/10/13

Minitab Statistical Software offers three tests for Normality: Anderson-Darling (AD), Ryan-Joiner (RJ), and Kolmogorov-Smirnov (KS). The AD test is the default, but is it the best test at detecting Non-Normality? Let's compare the ability of each of these normality tests to detect non-normal data under three different scenarios. We'll use simulated data for each, but they reflect common situations you're likely to encounter if you're analyzing data for quality improvement.

Scenario 1: The manufacturing process produces large outliers from time-to-time. In this simulation, 29 values are simulated from a Normal ($\mu = 0$, $\sigma = 1$), and 1 value is simulated from a Normal ($\mu = 0$, $\sigma = 4$).

Scenario 2: The manufacturing process has a process change that results in a shift in the distribution. In this simulation, 15 values are simulated from a Normal ($\mu = 0$, $\sigma = 1$), and 15 values are simulated from a Normal ($\mu = 0$, $\sigma = 4$).



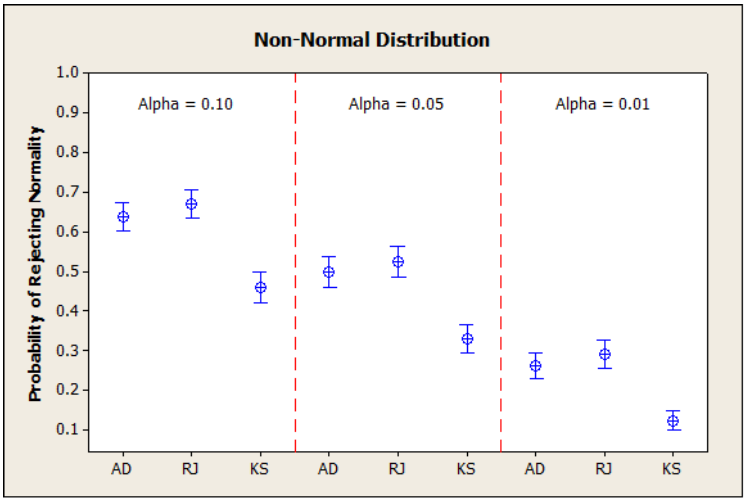
Scenario 3: – The measurements naturally follow a Non-normal distribution, as we'd typically see with time-to-failure data or strength measurements. For this scenario, 30 values are simulated from a Weibull($\alpha = 1$, $\beta = 1.5$) distribution.

I should note that the three scenarios evaluated in this blog are not designed to assess the validity of the Normality assumption for tests that benefit from the Central Limit Theorem, such as such as 1-sample, 2-sample, and paired t-tests. Our focus here is detecting Non-Normality when using a distribution to estimate the probability of manufacturing defective (out-of-spec) unit.

In scenario 1, the Ryan-Joiner test was a clear winner.

In scenario 2, the Anderson-Darling test was the best.

In scenario 3, there was not much difference between the AD and RJ test. Both were more effective at detecting Non-Normality than the Kolmogorov-Smirnov test. The simulation results are below.



In summary, the Anderson-Darling test was never the worst test, but it was not nearly as effective as the RJ test at detecting a 4-sigma outlier. ***If you're analyzing data from a manufacturing process tends to produce individual outliers, the Ryan-Joiner test is the most appropriate***.

The RJ test performed very well in two of the scenarios, but was poor at detecting Non-Normality when there was a shift in the data. ***If you're analyzing data from a manufacturing process that tends to shift due to unexpected changes, the AD test is the most appropriate.***

The KS test did not perform well in any of the scenarios.

Here's what a former colleague told me about the KS test: T***he Kolmogorov-Smirnov test is good for drinking, but not for thinking***.