Minitab® 16
Statistical Software

MINITAB ASSISTANT WHITE PAPER

This is one of a series of papers that explains the research conducted by Minitab statisticians to develop the methods and data checks used in the Assistant in Minitab 16 Statistical Software.

BINOMIAL CAPABILITY AND POISSON CAPABILITY

Overview

Capability analysis is used to evaluate whether a process is capable of producing output that meets customer requirements. When it is not possible to represent the quality of a product or service with continuous data, attribute data is often collected to assess its quality. The Minitab Assistant includes two analyses to examine the capability of a process with attribute data:

- Binomial Capability: This analysis is used when a product or service is characterized as defective or not defective. Binomial capability evaluates the chance (p) that a selected item from a process is defective. The data collected are the number of defective items in individual subgroups, which is assumed to follow a binomial distribution with parameter p.
- Poisson Capability: This analysis is used when a product or service can have multiple defects and the number of defects on each item is counted. Poisson capability evaluates the number of defects per unit. The data collected are the total number of defects in k units contained in individual subgroups, which is assumed to follow a Poisson distribution with an unknown mean number of defects per unit (u).

To adequately estimate the capability of the current process and to reliably predict the capability of the process in the future, the data for these analyses should come from a stable process (Bothe, 1991; Kotz and Johnson, 2002). In addition, there should be enough subgroups collected over time to ensure that the capability estimates represent the process capability over a long period of time. Even if a process is in control, it may experience input and environmental changes over time. Therefore, using an adequate number of subgroups can better enable you to capture the different sources of variation over time (Bothe, 1997; AIAG, 1995). Finally, there should be enough data to ensure that the capability statistics have good precision, as indicated by the width of the confidence interval for the key capability measure reported by both analyses.

Based on these requirements, the Assistant Report Card automatically performs the following checks on your data:

- Stability of process
  - Tests for special causes
  - Subgroup size
- Number of subgroups
- Amount of data

In this paper, we investigate how these requirements relate to capability analysis in practice and we describe how we established the guidelines to check for these requirements in the Assistant.

Note: Binomial and Poisson capability analyses include the P and U attribute control charts, respectively, to check process stability. These two charts depend on additional assumptions that either cannot be checked or are difficult to check. See Appendix A for details.

Data checks

Stability (Part I) – Test for special causes

To estimate process capability accurately, your data should come from a stable process. You should verify the stability of your process before you evaluate its capability. If the process is not stable, you should identify and eliminate the causes of the instability.

The P chart and the U chart are the most widely used attribute control charts to evaluate the stability of a process. The P chart plots the proportion of defective items per subgroup and is used with data that follow a binomial distribution. The U chart plots the number of defects per unit and is used with data that follow a Poisson distribution. Four tests can be performed on these charts to evaluate the stability of the process. Using these tests simultaneously increases the sensitivity of the control chart. However, it is important to determine the purpose and added value of each test because the false alarm rate increases as more tests are added to the control chart.

Objective

We wanted to determine which of the four tests for stability to include with the attribute control charts in the Assistant. Our first goal was to identify the tests that significantly increased sensitivity to out-of-control conditions without significantly raising the false alarm rate. Our second goal was to ensure the simplicity and practicality of the charts.

Method

The four tests for stability for attribute charts correspond with tests 1-4 for special causes for variables control charts. With an adequate subgroup size, the proportion of defective items (P chart) or the number of defects per unit (U chart) follow a normal distribution. As a result, simulations for the variables control charts that are also based on the normal distribution will yield identical results for the sensitivity and false alarm rate of the tests. Therefore, we used the results of a simulation and a review of the literature performed for variables control charts to evaluate how the four tests for stability affect the sensitivity and the false alarm rate of the attribute charts. In addition, we evaluated the prevalence of special causes associated with the test. For details on the method(s) used for each test, see the Results section below and Appendix B.

Results

Of the four tests used to evaluate stability in attribute charts, we found that tests 1 and 2 are the most useful:

Test 1: Identifies points outside of the control limits

Test 1 identifies points > 3 standard deviations from the center line. Test 1 is universally recognized as necessary for detecting out-of-control situations. It has a false alarm rate of only 0.27%.

Test 2: Identifies shifts in the proportion of defective items (P chart) or the mean number of defects per unit (U chart)

Test 2 signals when 9 points in a row fall on the same side of the center line. We performed a simulation to determine the number of subgroups needed to detect a signal for a shift in the proportion of defective items (P chart) or a shift in the mean number of defects per unit (U chart). We found that adding test 2 significantly increases the sensitivity of the chart to detect small shifts in the proportion of defective items or the mean number of defects per unit. When test 1 and test 2 are used together, significantly fewer subgroups are needed to detect a small shift compared to when test 1 is used alone. Therefore, adding test 2 helps to detect common out-of-control situations and increases sensitivity enough to warrant a slight increase in the false alarm rate.

Tests not included in the Assistant

Test 3: k points in a row, all increasing or all decreasing

Test 3 is designed to detect drifts in the proportion of defective items or in the mean number of defects per unit (Davis and Woodall, 1988). However, when test 3 is used in addition to test 1 and test 2, it does not significantly increase the sensitivity of the chart. Because we already decided to use tests 1 and 2 based on our simulation results, including test 3 would not add any significant value to the chart.

Test 4: k points in a row, alternating up and down

Although this pattern can occur in practice, we recommend that you look for any unusual trends or patterns rather than test for one specific pattern.

Stability (Part II) - Subgroup size

Although the P chart and the U chart monitor the stability of the process with attribute data, the normal distribution is used to approximate the distribution of the proportion of defective items ($\hat{p}$) in the P chart and the distribution of the number of defects per unit ($\hat{u}$) in the U chart. As the subgroup size increases, the accuracy of this approximation improves. Because the criteria for the tests used in each control chart are based on the normal distribution, increasing the subgroup size to obtain a better normal approximation improves the chart's ability to accurately identify out-of-control situations and reduces the false alarm rate. When the proportion of defective items or the number of defects per unit is low, you need larger subgroups to ensure accurate results.

Objective

We investigated the subgroup size that is needed to ensure that the normal approximation is adequate enough to obtain accurate results for the P chart and the U chart.

Method

We performed simulations to evaluate the false alarm rates for various subgroup sizes and for various proportions (p) for the P chart and for various mean numbers of defects per subgroup (c) for the U chart. To determine whether the subgroup size was large enough to obtain an adequate normal approximation and thus, a low enough false alarm rate, we compared the results with expected false alarm rate under the normal assumption (0.27% for Test 1 and 0.39% for test 2). See Appendix C for more details.

Results

P chart

Our research showed that the required subgroup size for the P chart depends on the proportion of defective items (p). The smaller the value of p, the larger the subgroup size (n) that is required. When the product np is greater than or equal to 0.5, the combined false alarm rate for both test 1 and test 2 is below approximately

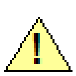
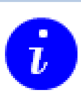
2.5%. However, when the product np is less than 0.5, the combined false alarm rate for tests 1 and 2 can be much higher, reaching levels well above 10%. Therefore, based on this criterion, the performance of the P chart is adequate when the value of np ≥ 0.5.

U chart

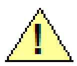
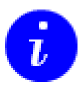
Our research showed that the required subgroup size for the U chart depends on the number of defects per subgroup (c), which equals the subgroup size (n) times the number of defects per unit (u). The percentage of false alarms is highest when the number of defects c is small. When c = nu is greater than or equal to 0.5, the combined false alarm rate for both test 1 and test 2 is below approximately 2.5%. However, for values of c less than 0.5, the combined false alarm rate for tests 1 and 2 can be much higher, reaching levels well above 10%. Therefore, based on this criterion, the performance of the U chart is adequate when the value of c = nu ≥ 0.5.

Based on the above results for the tests for special causes (Part I) and for the subgroup size (Part II), the Assistant Report Card displays the following status indicators when checking stability in the attribute control charts that are used in binomial and Poisson capability:

P chart – Binomial capability

Status	Condition
	No test 1 or test 2 failures on the chart and $n_i \bar{p} \geq 0.5$ for all i where $n_i =$ subgroup size for the ith subgroup $\bar{p}$ = mean proportion of defective items
	Test 1 or test 2 reveals one or more out-of-control points that may be due to special causes.
	The subgroup size may be too small. $n_i \bar{p} < 0.5 \; for \; at \; least \; one \; i$

U chart - Poisson capability

Status	Condition
	No test 1 or test 2 failures on the chart and a$n_i \bar{u} \geq 0.5$ for all i where $n_i =$ subgroup size for the i[th] subgroup $\bar{u}$ = mean number of defects per unit
	Test 1 or test 2 reveals one or more out-of-control points that may be due to special causes.
	The subgroup size may be too small. a$n_i \bar{\bar{u}} < 0.5$ *for at least one i*

Number of subgroups

To ensure that the capability estimates accurately reflect your entire process, you should try to capture all the likely sources of variation in your process over time. If you increase the number of subgroups you collect, you are likely to increase the chance that you are capturing the different sources of variation. Collecting an adequate number of subgroups also helps to improve the precision of the limits of the control charts that are used to evaluate the stability of your process. However, collecting more subgroups requires more time and resources; therefore, it is important to know how the number of subgroups affects the reliability of the capability estimates.

Objective

We investigated how many subgroups are needed to adequately represent the process and provide a reliable estimate of process capability.

Method

We reviewed the literature to find out the number of subgroups that is generally considered adequate for estimating process capability.

Results

According to the Statistical Process Control (SPC) manual, the number of subgroups you collect should be based on how long it takes to collect data that is likely to reflect the different sources of variation in your process (AIAG, 1995). That is, you should collect as many subgroups as is necessary to adequately represent your entire process. In general, to provide accurate tests of stability and a reliable estimate of process performance, AIAG (1995) recommends that you collect at least 25 subgroups.

Based on these recommendations, the Assistant Report Card displays the following status indicator when checking the number of subgroups for binomial or Poisson capability analysis:

Status	Condition
	Number of subgroups ≥ 25 The number of subgroups should be enough to capture different sources of process variation when collected over an adequate period of time. **Number of subgroups < 25** Generally, you should collect at least 25 subgroups over an adequate period of time to capture different sources of process variation.

Amount of data

The Assistant reports for binomial and Poisson capability analyses also include a 95% confidence interval for the percentage of defective items or the number of defects per unit, respectively. This interval is calculated using standard statistical methodology and did not require any special research or simulations.

The Assistant Report Card displays the following status indicator when checking the amount of data:

Status	Condition
	Binomial capability The 95% confidence interval for % defective is (a, b). If this interval is too wide for your application, you can gather more data to increase the precision. **Poisson capability** The 95% confidence interval for the number of defects per unit is (a, b). If this interval is too wide for your application, you can gather more data to increase the precision.

References

AIAG (1995). *Statistical process control (SPC) reference manual.* Automotive Industry Action Group.

Bischak, D.P., & Trietsch, D. (2007). The rate of false signals in $\bar{X}$ control charts with estimated limits. *Journal of Quality Technology*, *39*, 55–65.

Bothe, D.R. (1997). *Measuring process capability: Techniques and calculations for quality and manufacturing engineers*. New York: McGraw-Hill.

Bowerman, B.L., & O' Connell, R.T. (1979). *Forecasting and time series: An applied approach*. Belmont, CA: Duxbury Press.

Chan, L. K., Hapuarachchi K. P., & Macpherson, B.D. (1988). Robustness of $\bar{X}$ and *R* charts. *IEEE Transactions on Reliability, 37*, 117–123.

Davis, R.B., & Woodall, W.H. (1988). Performance of the control chart trend rule under linear shift. *Journal of Quality Technology*, *20*, 260–262.

Montgomery, D.C. (2001). *Introduction to statistical quality control*, 4th edition. New York: John Wiley & Sons, Inc.

Schilling, E.G., & Nelson, P.R. (1976). The effect of non-normality on the control limits of $\bar{X}$ charts. *Journal of Quality Technology*, *8*, 183–188.

Trietsch, D. (1999). *Statistical quality control: A loss minimization approach*. Singapore: World Scientific Publishing Co.

Wheeler, D.J. (2004). *Advanced topics in statistical process control. The power of Shewhart's charts*, 2nd edition. Knoxville, TN: SPC Press.

Yourstone, S.A., & Zimmer, W.J. (1992). Non-normality and the design of control charts for averages. *Decision Sciences*, *23*, 1099–1113.

Appendix A: Additional assumptions for attribute control charts

The P chart and the U chart require additional assumptions that are not evaluated by data checks:

P Chart	U Chart
• The data consists of n distinct items, with each item classified as either defective or not defective. • The probability of an item being defective is the same for each item within a subgroup. • The likelihood of an item being defective is not affected by whether the preceding item is defective or not.	• The counts are counts of discrete events. • The discrete events occur within some well-defined finite region of space, time, or product. • The events occur independently of each other and the likelihood of an event is proportional to the size of area of opportunity.

For each chart, the first two assumptions are an inherent part of the data collection process; the data itself cannot be used to check whether these assumptions are satisfied. The third assumption can be verified only with a detailed and advanced analysis of data, which is not performed by the Assistant.

Appendix B:
Stability - Tests for special causes

Simulation B1: How adding test 2 to test 1 affects sensitivity

Test 1 detects out-of-control points by signaling when a point is greater than 3 standard deviations from the center line. Test 2 detects shifts in the proportion of defective items or the number of defects per unit by signaling when 9 points in a row fall on the same side of the center line.

To evaluate whether using Test 2 with Test 1 improves the sensitivity of the attribute charts, we established control limits based on a normal (p, $\sqrt{\frac{p(1-p)}{n}}$) (p is the proportion of defective items and n is the subgroup size) distribution for the P chart and on a normal (u $\sqrt{u}$) (u is the mean number of defects per unit) distribution for the U chart. We shifted the location (p or u) of each distribution by a multiple of the standard deviation (SD) and then recorded the number of subgroups needed to detect a signal for each of 10,000 iterations. The results are shown in Table 1.

Table 1: Average number of subgroups until a test 1 failure (Test 1), test 2 failure (Test 2) or test 1 or test 2 failure (Test 1 or 2). The shift equals a multiple of the standard deviation (SD).

Shift	**Test 1**	**Test 2**	**Test 1 or 2**
0.5 SD	154	84	57
1 SD	44	24	17
1.5 SD	15	13	9
2 SD	6	10	5

As shown in the table, when both tests are used (*Test 1 or 2* column) an average of 57 subgroups are needed to detect a 0.5 standard deviation shift in the location, compared to an average of 154 subgroups needed to detect a 0.5 standard deviation shift when test 1 is used alone. Therefore, using both tests significantly increases sensitivity to detect small shifts in the proportion of defective items or the mean number of defects per unit. However, as the size of the shift increases, adding test 2 does not increase the sensitivity as significantly.

Appendix C: Stability - Subgroup size

The central limit theorem states that the normal distribution can approximate the distribution of the average of an independent, identically distributed random variable. For the P chart, $\hat{p}$ (subgroup proportion) is the average of an independent, identically distributed Bernoulli random variable. For the U chart, $\hat{u}$ (subgroup rate) is the average of an independent, identically distributed Poisson random variable. Therefore, the normal distribution can be used as an approximation in both cases.

The accuracy of the approximation improves as the subgroup size increases. The approximation also improves with a higher proportion of defective items (P chart) or a higher number of defects per unit (U chart). When either the subgroup size is small or the values of p (P chart) or u (U chart) are small, the distributions for $\hat{p}$ and $\hat{u}$ are right skewed, which increases the false alarm rate. Therefore, we can evaluate the accuracy of the normal approximation by looking at the false alarm rate and we can also determine the minimum subgroup size necessary to obtain an adequate normal approximation.

To do this, we performed simulations to evaluate the false alarm rates for various subgroup sizes for the P chart and the U chart and compared the results with the expected false alarm rate under the normal assumption (0.27% for Test 1 and 0.39% for test 2).

Simulation C1: Relationship between subgroup size, proportion, and false alarm rate of the P chart

Using an initial set of 10,000 subgroups, we established the control limits for various subgroup sizes (n) and proportions (p). We also recorded the percentage of false alarms for an additional 2,500 subgroups. We then performed 10,000 iterations and calculated the average percentage of false alarms from test 1 and test 2, as shown in Table 2.

Table 2: % of false alarms due to test 1, test 2 (np) for various subgroup sizes (n) and proportions (p)

	p				
Subgroup Size (n)	**0.001**	**0.005**	**0.01**	**0.05**	**0.1**
10	0.99, 87.37 (0.01)	4.89, 62.97 (0.05)	0.43, 40.14 (0.1)	1.15, 1.01 (0.5)	1.28, 0.42 (1)

	p				
Subgroup Size (n)	**0.001**	**0.005**	**0.01**	**0.05**	**0.1**
50	4.88, 63.00 (0.05)	2.61, 10.41 (0.25)	1.38, 1.10 (0.5)	0.32, 0.49 (2.5)	0.32, 0.36 (5)
100	0.47, 40.33 (0.10)	1.41, 1.12 (0.5)	1.84, 0.49 (1)	0.43, 0.36 (5)	0.20, 0.36 (10)
150	1.01, 25.72 (0.15)	0.71, 0.43 (0.75)	0.42, 0.58 (1.5)	0.36, 0.42 (7.5)	0.20, 0.36 (15)
200	1.74, 16.43 (0.2)	1.86, 0.50 (1.00)	0.43, 0.41 (2)	0.27, 0.36 (10)	0.34, 0.36 (20)
500	1.43, 1.12 (0.5)	0.42, 0.50 (2.5)	0.52, 0.37 (5)	0.32, 0.37 (25)	0.23, 0.36 (50)

The results in Table 2 show that the percentage of false alarms is generally highest when the proportion (p) is small, such as 0.001 or 0.005, or when the sample size is small (n = 10). Therefore, the percentage of false alarms is highest when the value of the product np is small, and lowest when np is large. When np is greater or equal to 0.5, the combined false alarm rate for both test 1 and test 2 is below approximately 2.5%. However, for values of np less than 0.5, the combined false alarm rate for tests 1 and 2 is much higher, reaching levels well above 10%. Therefore, based on this criterion, the performance of the P chart is adequate when the value of np ≥ 0.5. Thus, the subgroup size should be at least $\frac{0.5}{\bar{p}}$.

Simulation C2: Relationship between subgroup size, number of defects per unit, and false alarm rate of the U chart

Using an initial set of 10,000 subgroups, we established the control limits for various subgroup sizes (n) and number of defects per subgroup (c). We also recorded the percentage of false alarms for an additional 2,500 subgroups. We then performed 10,000 iterations and calculated the average percent of false alarms from test 1 and test 2, as shown in Table 3.

Table 3: % of false alarms due to test 1, test 2 for various number of defects per subgroup (c = nu)

c	0.1	0.3	0.5	0.7	1.0	3.0	5.0	10.0	30.0	50
% False alarms	0.47, 40.40	3.70, 6.67	1.44, 1.13	0.57, 0.39	0.36, 0.51	0.38, 0.40	0.54, 0.38	0.35, 0.37	0.29, 0.37	0.25, 0.37

The results in Table 3 show that the percentage of false alarms is highest when the product of the subgroup size (n) times the number of defects per unit (u), which equals the number of defects per subgroup (c), is small. When c is greater or equal to 0.5, the combined false alarm rate for both test 1 and test 2 is below approximately 2.5%. However, for values of c less than 0.5, the combined false alarm rate for tests 1 and 2 is much higher, reaching levels well above 10%. Therefore, based on this criterion, the performance of the U chart is adequate when the value of c = nu ≥ 0.5. Thus, the subgroup size should be at least $\frac{0.5}{\bar{u}}$.