Statistical Significance
Testing, Control and Comparison Charts

Contents
Introduction .............................................................. 3
Statistical Significance Testing ..................................... 3
  Interpreting Statistical Significance ............................ 3
  Influencing Statistical Significance ............................ 3
  Confidence Levels .................................................... 4
Control and Comparison Charts .................................. 5
  Control Charts ...................................................... 5
  Control Chart Errors ............................................... 5
  Special Cause Variation ............................................ 5
  Other Analysis Modules ............................................ 6
Comparison Charts .................................................... 6
  Data Distribution .................................................... 6
  Comparative Norm .................................................. 7
  Special Cause Variation ............................................ 7
Definition of Related Terms ......................................... 8
Appendix: Formulas Used ............................................. 9
  Proportion or Rate Measures ...................................... 9
  Statistical Significance ............................................ 9
    Calculating the z statistic ...................................... 9
    Calculating the t statistic ..................................... 9
    Control Charts ................................................... 9
Center Line Calculation 10
Upper and Lower Control Limits for each period 10
Comparison Charts 10
  Continuous Variable Measures (charge, cost, LOS) ........ 11
  Control Charts ..................................................... 11
M and MR Charts 11
X-bar S Charts and XmR Charts 13
  Comparison Charts ................................................ 14
  ................................................................. 16
Introduction

Quantros’ web-based software includes a comprehensive statistical package within the system. This package includes statistical significance testing at 95% and 75% confidence intervals for continuous variables (e.g., charges, costs, and length of stay) and rate-based measures (e.g., mortality, complications, and readmissions). A 99% confidence interval is used for National Hospital Quality Measures as specified by The Joint Commission. The package also includes the ability to produce presentation-ready control and comparison charts.

This training manual is in the form of frequently asked questions and should cover most of the questions pertaining to the application and interpretation of statistical significance testing. If you should have additional questions please feel free to contact our client support department at 1-800-711-8363.

Statistical Significance Testing

What is Statistical Significance Testing

Statistical significance testing relies on a mathematical formula for determining whether the difference between the mean of one value when compared to the mean of another like variable can be attributed simply to random or “common causes” associated with normal variation in the data or whether the variation (or difference) is due to assignable or “special” causes not explained by normal variation in the data. For instance, one might want to determine if the difference in a hospital’s length of stay in a particular DRG is statistically significant when compared to a market norm (or peer group) for the same DRG in order to direct and prioritize process improvement initiatives.

Interpreting Statistical Significance

When a value, such as a hospital’s mortality rate, is compared to a national norm and found to be statistically significant at a confidence level of 95%, one can conclude there is only a 5% probability that the variation is due to random chance, or said differently, one can be 95% confident that the variation is due to special causes. Statistical significance indicates there are discoverable reasons that explain why the values are different and in so doing appropriately identifies opportunities for improvement (if a higher mortality rate) or validates superior performance (if a lower mortality rate).

Influencing Statistical Significance

There are four (4) important factors that affect the likelihood of statistical significance being identified:

1. The extent of the difference between the mean of the two values being assessed (i.e., the greater the difference between the mean of two values the greater the chance of statistical significance existing).

2. The volume associated with the value being assessed (i.e., the more volume associated with a value, the less variation is required for statistical significance to exist).
3. The degree to which values in a distribution deviate from the mean of the distribution as measured by the "standard deviation" (i.e., as the standard deviation of the value becomes smaller there is an increased likelihood that the value will be statistically significant).

4. The confidence level (or interval) used to determine statistical significance (i.e., the lower the confidence level, the greater the probability that the value will be statistically significant).

5. In short, there is an increased likelihood that the difference between the means of two values will be statistically significant when: (a) the difference between the means becomes larger, (b) the dispersion (standard deviation) of the distribution for the value becomes smaller, (c) the volume associated with the value increases, or (d) the confidence level is lowered.

Statistical Significance Tests

There are two (2) common tests that can be used to determine statistical significance: the z-test (or z-value) and the t-test (or t-score). The t-test is appropriate to use when small sample sizes are associated with the value.

Note:

1. As the sample size increases the t-test approximates the z-test in which case the z-test is used
2. For rate-based measures the z-test (or z-value) is used.
3. For continuous variables both tests are used depending on the sample size:
   - If the number of cases is ≥ 25 the “z-test” is used
   - If the number of cases is < 25 the “t-test” is used.

Confidence Levels

Hospital Core Compliance Analysis

A 99% confidence level is used based on The Joint Commission survey specifications which are designed to red flag a manageable number of hospitals for performance improvement activities. The higher the confidence level the less likely the variation will be due to chance which reduces the probability that a hospital will be red flagged.

All Other Analysis Modules

Both a 95% and 75% confidence level is used to determine statistical significance. A 95% confidence level is the most common interval used in published research and is considered the standard for assessing statistical significance. A 75% confidence level is used to identify additional improvement opportunities beyond those identified at a 95% confidence level. Opportunities found at a 75% confidence level are less likely to be due to special (or assignable) causes. With limited time and resources, the use of two confidence levels allows for the prioritization of improvement objectives (those found at a 95% confidence level would be a higher priority than those at 75%).

Statistical Significance within Quality Solutions

Quantros’ software uses the following statistical significance designations:

- ↑** High at a 95% confidence level
- ↓** Low at a 95% confidence level
- ↑* High at a 75% confidence level
• ↓ Low at a 75% confidence level

Note: Statistical significance is identified as high or low using arrows and color coding, where “orange up” arrows indicate “high” and “blue down” arrows indicate “low”. The confidence level is identified using a “double asterisk” (**) for “95%” and a “single asterisk” (*) for “75%.”

Control and Comparison Charts

Control charts measure the HCO’s observed outcomes over time in order to analyze the type of variation “within” the HCO--common cause or special cause. The control chart enables the organization to determine if the process is stable, or “under control”. Once the stability of the process is confirmed then comparison charts can be used to compare the HCO’s performance to other HCOs. Unlike control charts, comparison charts provide a static comparison to an external norm for a point in time for the purpose of evaluating performance improvement opportunities.

Control Charts

Control charts evaluate the stability of a given process by determining whether process is under statistical control or not. Statistical control is achieved when only random variation exists in the data (referred to as common cause variation). However, common cause variation does not imply that the process is functioning at a desirable level or not, it only describes the nature of the variation, namely that it is stable and predictable within given limits - it is in statistical control. Note that even if the process is in control one can introduce a program for improvement in order to achieve a better result in outcomes (e.g., the implementation of evidenced-based guidelines or clinical pathways). Special cause variation indicates that the process is out of statistical control and unstable. It is no longer predictable within limits. The control chart can detect a special cause variation by identifying points outside the upper and lower control limits.

A control chart is a line graph with four lines: the upper control limit (UCL), lower control limit (LCL), a center line representing the overall process average or mean, and a line representing the process over time. Control limits describe the natural variability of a process over time and are usually set at three standard deviations (or “three sigma”).

Control Chart Errors

When using a control chart to identify special cause variation there are two types of errors that can be made:

• Type I: concludes there is a special cause when it is not present
• Type II: concludes there is no special cause, when in fact it is present

Using three sigma control limits offers the best balance for reducing the probability of making either type of error. Note that “sigma” is used synonymously with “standard deviation”.

Special Cause Variation

There are three tests that determine if special cause variation exists. They are as follows:

• One data point above or below the UCL or LCL
Testing, Control and Comparison Charts

- A run of eight consecutive data points on one side of the center line
- A trend of six consecutive data points steadily increasing or decreasing

Note: *Special cause variation can exist even if there are no points above or below the control limits.*

Other Analysis Modules

There are several different types of control charts that can be used depending on the kind of data that is evaluated. Quantros’ software will automatically generate the appropriate chart. Refer to the section below for a summary of the type of control charts used for each type of measure.

<table>
<thead>
<tr>
<th>Core Analysis Reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of Measure</strong></td>
</tr>
<tr>
<td>Proportion or Rate</td>
</tr>
<tr>
<td>Continuous Variable</td>
</tr>
<tr>
<td>Continuous Variable (low sample size)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quality Outcomes &amp; Ratings and Quality &amp; Utilization Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of Measure</strong></td>
</tr>
<tr>
<td>Proportion or Rate</td>
</tr>
<tr>
<td>Continuous Variable</td>
</tr>
<tr>
<td>Continuous Variable (low sample size)</td>
</tr>
</tbody>
</table>

Note: *Control charts can also be used for skewed data, because they are “averages of samples” and they tend to be normally distributed regardless of the shape of the distribution from which the samples were drawn.*

Using The Joint Commission criteria, in the Hospital Core Compliance Analysis, control charts are created only when there are at least 12 data points for a given measure. The 12 points can include missing data points. The charts will depict a center line (overall process average), +3 and -3 sigma lines (UCL and LCL) and the HCO’s observed rate. Because the sample size for monthly data points varies, the control limits will be calculated monthly often resulting in uneven lines.

Comparison Charts

While control charts evaluate the stability of an HCO’s process by comparing its own data to itself over time, comparison charts compare an HCO’s data to an external norm derived from the average performance of multiple HCOs. Comparison charts recognize that an HCO’s process may be in statistical control but still be performing poorly when compared to the outcomes of other HCOs (i.e., it is possible to have a poor process that is statistically stable). Comparison charts also allow HCOs to document where they are performing better (or worse) than their peers. Typically, a comparison chart analysis is generated after evaluating an HCO’s control chart because it is important to know that the existing process is stable before embarking on additional process improvement activities.
Data Distribution

- For rate-based measures, such as mortality rates, a binomial distribution is assumed;
- For continuous measures, such as charges, a normal distribution is assumed.

Comparative Norm

The comparative norms used are the:

- Predicted or expected rate (weighted median predicted rate) if the measure is risk-adjusted;
- Comparison group mean (weighted group mean) if the measure is not risk-adjusted;
- Weighted median rate if the measure is continuous and not risk-adjusted.

Special Cause Variation

Special cause variation is displayed in comparison charts by using a:

- P-value for tabular or numerical display, or an
- Expected range for graphical chart display

The expected range, also known as the acceptance interval, is an interval having upper and lower limits that represents the set of values for which the null hypothesis is accepted. A comparison chart consists of actual (or observed) rates, expected rates, and expected ranges for a given time period. The confidence limits (or levels) describe the degree of certainty that a given point is different from the average. For NHQM core measures, a 99% confidence level is applied. For all other measures, a 95% level is applied.
Definition of Related Terms

- **Average** or mean is the most common expression of the centering of a distribution. It is signified by \( \bar{x} \) and is calculated by totaling the observed values and dividing by the number of observations.

- **Common Cause** is a source of variation that is always present; part of the random variation inherent in the process itself. Its origin can usually be traced to an element of the system which only management can correct.

- **Control Chart** is a graphic representation of a characteristic of a process, showing plotted values of some statistic gathered for that characteristic, and one or two control limits. It has two basic uses: as a judgment to determine if a process is in control, and as an aid in achieving and maintaining statistical process control.

- **Control Limit** is a line (or lines) on a control chart used as a basis for judging the significance of the variation from subgroup to subgroup. Variation beyond a control limit is evidence that special causes are affecting the process. Control limits are calculated from process data and are not to be confused with engineering specifications.

- **Sigma** is the Greek letter used to designate the estimated standard deviation. In Quantros' software, three (3) sigma is used for all control charts (balances the risk of Type I and II errors).

- **Special Cause** is a source of variation that is intermittent, unpredictable, or unstable; sometimes referred to as an “assignable cause.” It is signaled by a point beyond the control limits.

- **Standard Deviation** is a standard unit of measure for describing the degree to which values in a distribution deviate from the mean. The standard deviation and variance are considered the two most important measures of variability (or dispersion) since they are commonly used in calculations for determining statistical inferences. For instance, in a normal distribution 68% of the values in a distribution always fall within one (1) standard deviation of the mean, 95% within two (2) standard deviations, and 99% within three (3) standard deviations. The standard deviation is calculated by taking the square root of the variance (refer to Appendix for formulas).

- **Statistical Control** is the condition describing a process from which all special causes have been removed, evidenced on a control chart by the absence of points beyond the control limits and by the absence of non-random patterns or trends within the control limits.

- **Trends** are the patterns in a run chart or control chart that feature the continued rise or fall of a series of points. A trend consists of six (6) consecutive data points steadily increasing or decreasing.
Appendix: Formulas Used

The statistical package is designed to follow closely the formulas and approach used by The Joint Commission for the National Hospital Quality Measures program in order to be consistent with industry standards.

Proportion or Rate Measures

Statistical Significance
To determine statistical significance, we utilize either the Z-statistic or the t-statistic and then compare to values on either the z table or the t table.

Symbols:
- \( n \) = number of cases for a period
- \( x_o \) = mean of observed values for a period
- \( s_o \) = standard deviation of observed values
- \( x_e \) = mean of risk adjusted value

Calculating the z statistic

\[
z = \frac{x_o - x_e}{s_o / \sqrt{n}}
\]

This value is assumed to follow a normal distribution when the sample size is > 25. The z value is compared to the values on the z table. If it is outside the specific confidence interval it is statistically significant.

Calculating the t statistic

\[
t = \frac{x_o - x_e}{S_o / \sqrt{n}}
\]

The t-statistic is assumed to follow a t-distribution. Unlike a normal distribution, the t- distribution depends on sample size. The t-value is compared to the values on the t table. If it is outside the specific confidence interval it is statistically significant.

Control Charts
Symbols:
- \( p \)-Chart – all proportion or rate measures are analyzed using the p-chart.
- \( n_i \) = number of denominator cases for the period
- \( x_i \) = number of numerator cases for the period
- \( p_i \) = observed rate for a month

**Center Line Calculation**

\[
p = \frac{\sum x_i}{\sum n_i} = \frac{x_1 + x_2 + \ldots + x_m}{n_1 + n_2 + \ldots + n_m}
\]

where \( m \) = number of data points

**Upper and Lower Control Limits for each period**

\[
p \pm 3 \sqrt{\frac{p*(1-p)}{n_i}}
\]

**Comparison Charts**

Symbols:
- \( n \) = number of denominator cases for the period
- \( p_o \) = observed rate for the period
- \( p_e \) = risk adjusted rate

1. Calculate the Z statistic or Z score. This value follows a normal distribution when the sample size is not very small. Any value less than -2.576 or greater than 2.576 signals statistically significant difference between the two rates at 1% significance level.

\[
Z = \frac{p_o - p_e}{\sqrt{p_e*(1-p_e)/n}}
\]

2. Calculate the Upper and Lower Control Limits of the Confidence Interval

\[
U_o = \left( p_o + \frac{Z^2}{2\cdot n} \right) + Z \cdot \sqrt{\frac{Z^2}{4\cdot n^2} + \frac{p_o*(1-P_o)}{n}}
\]

\[
U_o = 1 + \frac{Z^2}{n} \cdot \frac{1}{\alpha / 2}, \text{ where } Z \cdot \frac{1}{\alpha / 2} = 2.576
\]
Testing, Control and Comparison Charts

3. Calculate the upper limit and lower limit of the expected range

\[ U_e = p_e + (p_o - L_o) \]  \( \text{if } U_e > 1 \text{ then } U_e = 1 \)

\[ L_e = p_e + (p_o - U_o) \]  \( \text{if } L_e < 0 \text{ then } L_e = 0 \)

4. Sample Size Warning

(a) When the sample size (n) for a proportion measure is less than 30 or

(b) When the number of HCOs in the comparison group is less than 10, then an appropriate warning message is issued to indicate that the data may not be valid due to a small sample size and that such data should be interpreted and used with caution.

Continuous Variable Measures (charge, cost, LOS)

Control Charts

M and MR Charts

- \( \mu \) = process mean or the expected value of the population of median observed values
- \( \sigma \) = process standard deviation or standard deviation of the population of median observed values
- \( n_i \) = number of cases for the \( i \)th period
- \( M_i \) = median observed value in the \( i \)th period
- \( M \) = the overall average of the period median observed values
- \( R_i \) = range of measurements in the \( i \)th period
- \( N \) = number of periods being evaluated

\[ \sqrt{VAR(M_{\text{all } - 1/2})} \] = standard error of the median of \( n \) independent, normally distributed variables with unit standard deviation

- \( d_2(n_i) \) = the expected value of the range of \( n \) independent normally distributed variables with unit standard deviation
- \( d_3(n_i) \) = the standard error of the range of \( n \) independent observations from a normal population with unit standard deviation
1. Calculate the Center Line
   a. M-Chart
      \[ \bar{M} = \frac{(n_1M_1 + \ldots + n_kM_k)}{(n_1 + \ldots + n_k)} \]
   b. MR-Chart
      Median Chart:
      \[ \bar{M} = \frac{(n_1M_1 + \ldots + n_kM_k)}{(n_1 + \ldots + n_k)} \]
      Range Chart:
      \[ \bar{R} = d_2(n_i)\sigma \]

2. Calculate the control limits
   a. M-Chart
      \[ UCL/LCL = \left( M \pm 3\sigma \sqrt{VAR(M_{(n_1 + 1)/2})} \right) \]
   b. MR-Chart
      Median Chart:
      \[ UCL/LCL = \left( M \pm 3\sigma \sqrt{VAR(M_{(n_1 + 1)/2})} \right) \]
      Range Chart:
      \[ LCL = \max(d_4(n_i)\sigma - 3d_2(n_i)\sigma, 0) \]
      \[ UCL = (d_4(n_i)\sigma + 3d_2(n_i)\sigma) \]

Note: months where the number of cases = 1 are excluded from the control chart

X-bar S Charts and XmR Charts
- \( n_i \) = number of cases for the period
- \( x_i \) = mean of observed values for the period
- \( s_i \) = standard deviation of observed values for the period

1. Calculate the Center Line
   a. X-bar chart
Testing, Control and Comparison Charts

\[
\bar{x} = \frac{\sum n_i \cdot x_i}{\sum n_i}
\]

b. S-Chart – calculate the minimum variance linear unbiased estimate

\[
\bar{s} = c_4 \cdot \frac{\sum h_i \cdot s_i}{\sum h_i}, \quad h_i = \frac{c_4^2}{1 - c_4^2}
\]

\(c_4\) is a constant that depends on the sample size. As the sample size increases, \(c_4\) approaches to 1. The exact formula for \(c_4\) is:

\[
c_4 = \sqrt{\frac{2}{n_i - 1} \cdot \frac{\Gamma\left(\frac{n_i}{2}\right)}{\Gamma\left(\frac{n_i - 1}{2}\right)}}
\]

2. Calculate the control limits
   a. X-bar Chart

\[
\bar{x} \pm 3 \cdot \frac{\bar{s}}{c_4 \sqrt{n_i}}
\]

b. S-Chart

\[
\bar{s} \pm \left(1 + \frac{3}{c_4} \cdot \sqrt{1 - c_4^2}\right)
\]

If the average sample size is less than 10, an XmR chart is to be used instead of an X-bar and S-chart, assuming the observed mean value as a single observation for the period.

**Note:** The X-bar S chart is a paired chart, that is, the X-bar chart will reveal whether there is a special cause across time periods, while the S chart will reveal whether there are special causes within time periods. First interpret the S chart and if it is in control, then evaluate the X-bar chart.

**Comparison Charts**

Symbols:
- \(n\) = number of cases for a period
- \(x_o\) = median of observed values for a period
- \(s_o\) = standard deviation of observed values
- \(x_e\) = median of risk adjusted value
1. Calculate the z statistic
   
   a. \( n \geq 30 \)
   
   \[
   z = \frac{X_o - X_e}{s_o / \sqrt{n}}
   \]

   This value is assumed to follow a normal distribution when the sample size is not very small. Any value less than -2.576 or greater than 2.576 signals statistically significant difference between the two rates at 1% significance level.

   b. \( n < 30 \)

   \[
   t = \frac{X_o - X_e}{s_o / \sqrt{n}}
   \]

   The t-statistic is assumed to follow a t distribution. Unlike a normal distribution, the t distribution depends on sample size. For example, if the sample size is 15, any value less than -2.977 or greater than 2.977 signals statistically significant difference between the two rates at 1% significance level.

2. Calculate an expected range based on the test statistic

   Expected upper limit:
   
   \[
   U_e = x_e + (x_o - L_o)
   \]

   Expected lower limit:
   
   \[
   L_e = x_e + (x_o - U_o)
   \]

   \[
   U_o = x_o + Z_{\frac{1-\alpha}{2}} \times s_o / \sqrt{n} \quad \text{and} \quad L_o = x_o - Z_{\frac{1-\alpha}{2}} \times s_o / \sqrt{n} \quad \text{if} \quad n \geq 30
   \]

   Where;

   \[
   U_o = x_o + t_{\frac{1-\alpha}{2}} \times s_o / \sqrt{n} \quad \text{and} \quad L_o = x_o - t_{\frac{1-\alpha}{2}} \times s_o / \sqrt{n} \quad \text{if} \quad n \leq 30
   \]

   If the observed value is within the expected range, it is not a statistical outlier at 1% significance level. If the observed value is outside the expected range, the observed rate is a statistical outlier.

   **Note:** When the sample size \( n \) is very small, an appropriate warning message is issued to indicate that the data may not be valid due to a small sample size and that such data should be interpreted and used with caution. A warning message is issued (a) when the sample size is less than 10; or (b) when the number of HCOs in the comparison group is less than 10 for the risk adjusted measures whose risk adjusted data are not available.