Lean Six Sigma Statistical Tools in Healthcare

Richard W. Maclin

1/16/2017
Richard has over 23 years of continuous improvement experience. He has held roles in Quality Assurance, Supplier Development, Continuous Improvement, Product Realization, and Operations. His career has focused on reducing non-value added activities and process variation using CI approaches including Lean Six Sigma, VA/NVA, Value Stream Mapping, Theory of Constraints, 5S, and Kaizen. Richard has held positions with Toyoda Gosei, United Technologies, ArvinMeritor, Trelleborg, Superior Industries, and Gates Corporation and currently serves clients in the Automotive, Aerospace, Industrial, Nuclear, Food Service, and Consumer Products industries. He is an Adjunct Instructor at Northwest Arkansas Community College, Tulsa Technology Center, Francis Tuttle Technology Center, and is a Master Black Belt for CI Solutions, LLC. He is an ASQ member, ASQ Certified Six Sigma Black Belt, and ASQ Six Sigma Forum Advisory Council Member. He and his wife, Sheila, have four children and live near Bentonville, Arkansas.
Healthcare Finance Management

• Vitally important to organizational success!
• Healthcare is a “data-rich” environment.
• Analytical opportunities include both simple and complex methods (practices and tools)
Organizational Success

• Organizational success depends largely on effective strategic decision making and problem solving.

• Too often, organizations make decisions based on limited or anecdotal information, incomplete analysis, small data sets, and averages.
Why Use Six Sigma Stats Tools?

- Averages and small data sets do not tell the whole story! What about variation?
- Finding true causes can be confusing and hard!
- Organizations using Lean Six Sigma Statistical Tools to support decision making and problem solving have avoided costly mistakes.
- Many have experienced reduced costs and increased profitability as well as expanded capacity.
Disclaimer

The background story, information, and data presented here is for demonstration of Six Sigma statistical tools only and IS NOT an actual case study.

Any similarity to actual organizations or individuals is unintentional and purely accidental.

All statistical analysis presented here was completed using Minitab 17.3.1. This is not an endorsement or advertisement for Minitab.
Alpha Risk Medical Management (ARMM) is a healthcare management and investment firm. ARMM has purchased several local practices and hospitals to form 12 regional hospitals around the U.S. With hospitals, labs, and offices around the country, financial management is a top organizational function. The CFO estimates that ARMM left over $12 Million on the table in fiscal 2016 due to rejected, delayed, or unpaid claims.

A recent audit of 5000 medical bills indicated that 76% contained errors. As we know, insurance companies have very strict requirements for billing and coding, and claims are frequently rejected due to simple errors.
When claims are rejected, a long and costly process begins. The provider must correct the error, resubmit the claim, and wait for the claim to be accepted and processed. This may delay payment for several weeks or months.

The delay in payment, investigation costs, correction costs, and unpaid claims add up throughout a year. Additional costs include lost opportunity because some amount of our capacity is consumed with managing billing errors instead of treating patients.

ARMM’s leadership team has asked the CFO to lead a team to investigate and address the causes of lost revenue due to billing errors throughout the organization. She has chosen to build a team that includes a Lean Six Sigma Black Belt and members from all ARMM hospitals.

The background story, information, and data presented here is for demonstration of Six Sigma statistical tools only and IS NOT an actual case study. Any similarity to actual organizations or individuals is unintentional and purely accidental.
Lean Six Sigma Overview

Most Lean Six Sigma projects will follow the DMAIC structure.

- Standard approach
  - Five (5) phases
  - Team-based
  - Data-focused problem solving

Define > Measure > Analyze > Improve > Control
What are the DMAIC phases?

**Define** the problem with your product or process.

**Measure** your current process and collect data.

**Analyze** your data to find the problem’s root causes.

**Improve** your process by implementing and verifying corrective actions.

**Control** your new process and monitor it over time to hold the gains.
Define Measure Analyze Improve Control

Define Measure Analyze Improve Control
Define

What is the problem with your product or process?
Make the business case to justify the project.
Define Phase

The Lean Six Sigma journey:

✓ Starts with a problem
✓ Defines the customer
✓ Explains how the problem impacts the business
✓ Defines the process scope
✓ Identifies when the problem is solved
✓ Answers how success is measured & maintained

The primary output of Define is a complete and approved Project Charter to begin the project.
Types of Problems

 ✓ A donated liver is thrown away.
 ✓ A paid-off house is foreclosed on.
 ✓ Vaccine recalled for contamination.
 ✓ An automobile is recalled for accelerator issue.
 ✓ Automotive parts are rejected as off color.
 ✓ Shirts are rejected for being too small.
 ✓ Cell phones catching fire while charging.
 ✓ Billing error rate too high.

(Possible DMAIC Project Motivators)
Problem Statement

The CFO estimates that ARMM left over $12 Million on the table in fiscal 2016 due to rejected, delayed, or unpaid claims.

A recent audit of 5000 medical bills indicated that 76% contained errors.

Note: “Bills” are the statements received by the patient and includes one or more transactions (records).
Measure

Measure your current process and collect data.
Validate the measurement system.
Establish initial process capability.
The Lean Six Sigma journey continues:
- What do we need to measure?
- How do we measure them?
- Are we using suitable measurement systems?
- What is our performance specification?
  - What performance is expected?
- What is the baseline process performance?
Data Collection Tools

1. Records and Secondary Data
2. Observations (Tally sheets, Measurements)
3. Surveys and Interviews
4. Focus Groups
5. Diaries, Journals, Checklists, Check sheets
6. Expert Judgment
7. Integrated Systems (Databases)
Data Types

Process data falls into one of 3 major types:

- **Numeric**
  - Measurements
  - Continuous
  - Discrete

- **Categorical**
  - Ordinal
  - Nominal

- **Electronic**
  - Large record sets
  - Hard to change methods

Validated by Gage R&R Study

Validated by Attribute Agreement Analysis

Validated by review for missing entries, datatype mismatch, etc.
Billing Records Data

The team pulled a random sample of patient records from all 12 regional hospitals for June through August for review...

The background story, information, and data presented here is for demonstration of Six Sigma statistical tools only and IS NOT an actual case study. Any similarity to actual organizations or individuals is unintentional and purely accidental.
Analyze

Analyze your data to find the problem’s root cause(s).
Identifying the ‘vital few’ inputs.

This is where we use the most statistical tools!
Analyze Phase

The third Lean Six Sigma phase:

✓ What are the possible causes for the problem?
✓ How do each of those possible causes affect process performance?
✓ What are the vital few x’s (causes, inputs, etc.)?
✓ How can we identify the vital few?
Where do we start?
Before we know what to fix, we must identify possible causes that lead to the problem.
How can we do that?
• Brainstorming
• Review records & data
• Ask experts; Go to Gemba!!!
• Cause and Effect Diagrams
• Affinity / Tree Diagrams
Silent Brainstorming is a simple method to collect a large amount of information from a group of team members quickly. Here’s how:

- State the ‘Trigger Question’
- Set a 3-5 minute time limit
- Write 1 idea per sticky note
- No talking
- Be prolific..... Quantity!
- Put all post-its on the board
- Arrange post-its into groups
- Label the groups
- Discuss the groups and assign priority
Review Records

- Review any available records, both electronic and hardcopy if necessary
- It is important to define the time frame to review
- Recall the process scope! Avoid scope creep!
Go to Gemba!

- The people who do the work (where the value is created) are a great source of root cause information.
- Their inputs often come through expressing their pain points and the conditions present when the problem occurs.
- When visiting Gemba, it is crucial that we take lots of notes and ask open-ended questions.
Sometimes referred to as Ishikawa Diagrams or Fishbone Diagrams.

Billing Error Cause and Effect Diagram

- **Measurements**
  - New ACA Code List
  - In Patient/Out Patient ratio
  - Error detection false positives
  - July is worst (1)
  - Day of Week (2)
  - Accessibility
  - Union (3)
  - Profit type (4)

- **Material**
  - Old Record Errors
  - No internal code review
  - Manual coding
  - Region (6)
  - Hospital (5)

- **People**
  - Under Staffed (8)
  - Volunteers (7)
  - Poor Training
  - Intentional miscoding
  - Update to ICD-10 (9)
  - Defective Barcode Scanner
  - Out Dated Software

**Rejected Claims Due To Billing Errors**
Affinity / Tree Diagrams

- Similar to Cause and Effect Diagrams, Affinity / Tree Diagrams organize possible causes along branches representing major logical categories.
- There is no single ‘correct’ way to organize possible causes.
- Each possible cause must be reviewed and either ruled out or kept in for further consideration.
As we organize ideas of possible causes, our focus is to reduce the list of possible causes to the vital few process inputs (x’s) that contribute the most to the occurrence of the problem.

Hypothesis testing can be used to eliminate or confirm x’s, define correlations, determine differences in groups of data, explain relationships between x’s and y.
Identifying the Vital Few

Rejected Claims Due To Billing Errors

Measurements
- New ACA Code List
- In-patient / Out-patient Rate
- Error detection false positives

Material
- Old Record Errors

People
- Under Staffed (8)
- Volunteers (7)
- Poor Training
- Intentional miscoding

Environment
- Profit type (4)
- Accessible
- Day of Week (2)
- July is worst month (1)

Methods
- Hospital (5)
- Union (3)
- Accessibility
- Manual Coding
- No internal code review

Machines
- Outdated Software
- Defective Barcode Scanner
- Update to ICD-10 (9)

The Vital Few Possible Causes
Common Hypothesis Tests:
- Graphical Summary
- 1 Sample T test
- 2 Sample T test
- ANOVA test
- Test for two % Defective (not shown here)
- Chi-Square % Defective (not shown here)
- Correlation test
- Regression Analysis
Basic Steps in Hypothesis Testing

1. Define the problem and determine the objectives.
   - What question are we trying to answer?
2. Establish the hypothesis. (Null & Alternative)
3. State the risk levels (alpha & beta)
4. Decide on appropriate statistical test (assume distribution Z, t, F).
5. Establish the effect size (delta).
6. Determine critical statistic from the appropriate table.
7. Calculate test statistic (Z, t, or F) from the data.
8. State the Conclusion

Thankfully, we can use software like Minitab to help with steps 2-7!

*The Certified Six Sigma Black Belt Handbook* - Donald W. Benbow and T. M. Kubiak
Graphical Summary

- Not technically a Hypothesis Test, it is a collection of analytical information:
  - Descriptive statistics
  - Confidence intervals
  - Normality test
Graphical Summary of Value
Summary Report

Distribution of Data
Examine the center, shape, and variability.

Descriptive Statistics
- N: 21287
- Mean: 185.27
- StdDev: 91.153
- Minimum: -86.627
- 5th percentile: 51.691
- 25th percentile: 118.86
- Median: 172.26
- 75th percentile: 249.34
- 95th percentile: 349.99
- Maximum: 533.33

95% Confidence Intervals
- Mean: (184.05, 186.50)
- Median: (170.65, 173.95)
- StdDev: (90.295, 92.027)

Normality Test
- Decision: Fail
- P-value: <0.005

Data in Worksheet Order
Investigate any outliers (marked in red).
This test compares the mean (average) and confidence interval of the values recorded in the record set to a known historical mean or, perhaps, a target value.

Minitab’s Assistant feature makes understanding this tool simple.

- Descriptive statistics
- Graphical display
- Simple question
- Comments

Answers a question like:

Does our process performance differ from the target of $50 per rejected claim?
Based on this study, there is a statistically significant difference between the mean rejected claim value for all hospitals in this dataset and the target of $50.
Since our process is not achieving the desired performance, we should review possible causes from the earlier brainstorming activity.

Can we test any to either eliminate or confirm their importance?

- Organization Type (FP / NFP)
- Union vs. Nonunion
- Use Volunteers or No Volunteers
This test compares the means (average) and confidence intervals of the values recorded in the record set for two groups (or subsets) within the whole record set.

Again, Minitab’s Assistant feature makes understanding this tool simple.

Answers questions like:

*Is there a difference in the values of rejected claims from For Profit Hospitals compared to Not For Profit Hospitals?*
Based on this study, there is a statistically significant difference in the mean value of rejected claims between FP and NFP hospitals.
Possible Causes

We would use a 2 Sample t Test to check each possible cause that has 2 possible conditions?

- Organization Type (FP / NFP)
- Union vs. Nonunion
- Use Volunteers or No Volunteers
- Others…
Based on this study, there is a statistically significant difference in the mean value of rejected claims between Union & Non Union hospitals.
Possible Causes

- We would use the 2 Sample t Test to either eliminate or confirm the importance of each possible cause that has only 2 possible conditions (FP vs NFP; U vs. NU; etc.)
- But what about other possible causes that have more than 2 conditions?
  - Day of the Week
  - Month of the Year

ANOVA can test more than 2 groups!
ANOVA

✓ This test compares the means and confidence intervals of the values recorded in the record set for more than two groups within the whole record set. Minitab can compare up to 12 groups.

✓ Again, Minitab’s Assistant feature makes understanding this tool simple.

✓ Answers questions like:
  ✓ Are the rejected claims amounts different for some days compared to other days?
  ✓ Are the rejected claims amounts different for some Regions or Hospitals compared to others?
Based on this study, there is no statistically significant difference in the mean value of rejected claims regardless of week day.
Possible Causes

Since our initial test did not indicate a statistically significant difference between claims rejected based on the week day, it can be eliminated from further study. For now...

Continue testing possible causes with more than 2 conditions using ANOVA.
Based on this study, there is no statistically significant difference in the mean value of rejected claims regardless of month.
More ANOVA Tests

Based on this study, there is a statistically significant difference in the mean value of rejected claims from one or more hospital compared to other hospitals.

Based on this study, there is a statistically significant difference in the mean value of rejected claims from one or more regions compared to other regions.
Month is also eliminated…
We would use the ANOVA Test to either eliminate or confirm the importance of each possible cause that has more than 2 possible conditions (Weekday; Month; Hospital; Region; etc.)
But what about other possible causes that do not have discrete conditions, but are variable?
Staffing to Patient Ratio

Let’s check for any relationships!
Correlation – Regression

✓ Correlation analysis is used to quantify the degree of linear association between continuous variables.

✓ Correlation DOES NOT imply causation!

✓ Regression analysis is used to create an equation that defines the functional relationship between one or more continuous y’s and at least one continuous x.
Correlation Study

**Correlation: Value, StafftoPatientRatio**
Pearson correlation of Value and StafftoPatientRatio = -0.479
P-Value = 0.000

Based on this study, there is a correlation between Staffing and Rejected Claim Value, but it is fairly weak. Additionally, correlation does not imply causation, so do not jump to conclusions yet!

**Correlation: Value, Software**
Pearson correlation of Value and Software = -0.616
P-Value = 0.000

Based on this study, there is a correlation between Software Version and Rejected Claim Value, but it too is fairly weak. Recall, correlation does not imply causation, so do not jump to conclusions yet!
Possible Causes

- The team has tested or investigated most of the possible causes that resulted from the earlier Brainstorming Activity.
  - Many were eliminated.
  - Eight were confirmed or inconclusive (so leave them in).
  - Two could not be tested due to lack of data.
  - New measurement system may be required.

Let’s look at the Regression model!
Regression

Regression transitions from descriptive to inferential statistics.

Regression provides a mathematical model (formula) of a process, $Y = f(x)$.

Regressions can be relatively simple cases with one continuous input variable and an output.

Regressions can also be more complex with multiple (continuous or discrete) input variables and an output.
Regression Analysis results in a lot of information including:

<table>
<thead>
<tr>
<th>S</th>
<th>R-sq</th>
<th>R-sq(adj)</th>
<th>R-sq(pred)</th>
</tr>
</thead>
<tbody>
<tr>
<td>59.4112</td>
<td>57.54%</td>
<td>57.52%</td>
<td>57.49%</td>
</tr>
</tbody>
</table>

R-sq indicates that about 57.5% of the observed variation can be explained by the current model. This is not necessarily conclusive, but understanding some variation is better than understanding none.

Regression also constructs a formula that can be used to predict the output based on input data.

\[
\text{Value} = 27399 - 4354 \text{ StafftoPatientRatio} - 13.192 \text{ Software} \\
0.0 \text{ Region_Central} + 32.97 \text{ Region_East Coast} - 27.70 \text{ Region_Mid West} - \\
78.29 \text{ Region_Northwest} - 185.26 \text{ Region_Southeast} - 25.42 \text{ Region_West Coast} \\
+ 0.0 \text{ Codes_ICD-10-CM} - 278.53 \text{ Codes_ICD-9-CM} + 0.0 \text{ Type_FP} + 48.19 \text{ Type_NFP} \\
0.0 \text{ Union_NonUnion} - 29.06 \text{ Union_Union} + 0.0 \text{ Volunteer_No} - 25.75 \text{ Volunteer_Yes}
\]

- Main Effects Plots show how much each factor contributes to the process variation.
- The larger the range, the higher the effect.
- Nearly flat lines may be removed from the model, then re-run the regression.
Regression

Regression Analysis results in a lot of information including:

<table>
<thead>
<tr>
<th>Region</th>
<th>Codes</th>
<th>Value</th>
<th>StafftoPatientRatio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central</td>
<td>ICD-10-CM</td>
<td>23883</td>
<td>2219.7</td>
</tr>
<tr>
<td>Central</td>
<td>ICD-9-CM</td>
<td>23715</td>
<td>2219.7</td>
</tr>
<tr>
<td>East Coast</td>
<td>ICD-10-CM</td>
<td>23905</td>
<td>2219.7</td>
</tr>
<tr>
<td>East Coast</td>
<td>ICD-9-CM</td>
<td>23736</td>
<td>2219.7</td>
</tr>
<tr>
<td>Midwest</td>
<td>ICD-10-CM</td>
<td>23859</td>
<td>2219.7</td>
</tr>
<tr>
<td>Midwest</td>
<td>ICD-9-CM</td>
<td>23690</td>
<td>2219.7</td>
</tr>
<tr>
<td>Northwest</td>
<td>ICD-10-CM</td>
<td>23710</td>
<td>2219.7</td>
</tr>
<tr>
<td>Northwest</td>
<td>ICD-9-CM</td>
<td>23542</td>
<td>2219.7</td>
</tr>
<tr>
<td>Southeast</td>
<td>ICD-10-CM</td>
<td>23542</td>
<td>2219.7</td>
</tr>
<tr>
<td>Southeast</td>
<td>ICD-9-CM</td>
<td>23542</td>
<td>2219.7</td>
</tr>
</tbody>
</table>

R-sq indicates that about 54.5% of the observed variation can be explained by the current model. This is lower than the initial study but the model is much more simple now.

Multiple formulae based on Region and Codes but again, simplified.

Based on this regression, we may suspect that, in order to minimize the rejected claims mean value, we need a higher staff to patient ratio, newer software, replicate the methods used in the Southeast region, and go back to ICD-9.
Based on this regression, we may suspect that, in order to minimize the rejected claims mean value, we need a higher staff to patient ratio, newer software, replicate the methods used in the Southeast region, and go back to ICD-9.
Based on this regression, we may suspect that, in order to minimize the rejected claims mean value, we need a higher staff to patient ratio, newer software, replicate the methods used in the Southeast region, and go back to ICD-9.

But those conditions may not be practical!

So let’s use the Optimizer tool to look at possible scenarios that are more realistic.
• The optimizer shows the best combination of factors to achieve the target that we asked for…in this case, $0.
• But we can’t go back to ICD-9…thankfully, the optimizer allows us to change settings and see a predicted result.
Regression Optimizer

- Knowing that we cannot change from ICD-10, just grab the red line in the codes panel and move it.
- But now the model predicts a mean rejected claim value of $168!
- Can we improve the prediction?
With ICD-10, replicating the practices of the Southeast Region, keeping the newest software, and staffing to a ratio of 0.1371, the model predicts the mean rejected claims value to be $0.06.

How do we make all this happen?
Possible Causes

- Based on the Regression, we eliminated 4 additional factors and are left with the Vital Few process factors / causes!
- Now on to the Improve Phase…
Other Situations

- All the tools presented are designed for datasets that fit the Standard Normal distribution (symmetrical, balanced).

- Many processes result in non-normal data (and that’s ok!), non-parametric tests are available to analyze non-normal processes.

- Often, we only have summarized data instead of raw data. In these cases, use proportions tests.
Improve

Implement and verify corrective actions using Pilot Trials and Designed Experiments (DOE).
Improve

✓ Activity now becomes basic project management.
✓ Make a Plan – Execute the Plan – Evaluate the Results
✓ Additional tools may include using a Designed Experiment (DOE) to improve the Piloted process established in the Analyze phase.
An Action Item List is a common tool that aids teams in completion of tasks and follow up / follow through.

### Action Item Register

<table>
<thead>
<tr>
<th>Action Item</th>
<th>Owner</th>
<th>Due Date</th>
<th>Progress</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Team travel to all Southeast Region hospitals to understand and document their practices and procedures.</td>
<td>Lidio (LSSBB)</td>
<td>10/1/2015</td>
<td>50% 75% 100%</td>
<td>Travel plans scheduled.</td>
</tr>
<tr>
<td>2 Document practice / procedural differences across organization.</td>
<td>Team</td>
<td>12/1/2015</td>
<td>25% 50% 75% 100%</td>
<td></td>
</tr>
<tr>
<td>3 Develop strategy to standardize practices / procedures across the organization.</td>
<td>Mindy (CFO)</td>
<td>1/15/2016</td>
<td>25% 50% 75% 100%</td>
<td></td>
</tr>
<tr>
<td>4 Update and go live all hospitals to newest software version.</td>
<td>Info. Systems</td>
<td>1/15/2016</td>
<td>25% 50% 75% 100%</td>
<td></td>
</tr>
<tr>
<td>5 Train all hospitals on new software version.</td>
<td>Info. Systems</td>
<td>1/15/2016</td>
<td>25% 50% 75% 100%</td>
<td></td>
</tr>
<tr>
<td>6 Complete organization-wide assessment of Staff to Patient Ratio.</td>
<td>Dave (HR VP)</td>
<td>10/1/2015</td>
<td>25% 50% 75% 100%</td>
<td></td>
</tr>
<tr>
<td>7 Develop strategy to staff to optimized rate in East Coast Region to pilot the higher staffing level.</td>
<td>Dave (HR VP)</td>
<td>11/1/2015</td>
<td>25% 50% 75% 100%</td>
<td>Pilot plan to be developed not to exceed 90 days.</td>
</tr>
<tr>
<td>8 Develop staffing level Pilot Plan</td>
<td>Dave (HR VP)</td>
<td>11/15/2015</td>
<td>25% 50% 75% 100%</td>
<td></td>
</tr>
</tbody>
</table>
Control

Establish and control your new process and monitor it over time to hold the gains with SPC or other tools.
Control Phase includes developing:

- Control Plans
- Procedures
- Play Books
- Training For Everyone
- Control Charts
Summary

DMAIC problem solving includes:

• Defining the problem
• Assessing the measurement systems
• Collecting data
• Identifying possible causes
• Using statistical tools to evaluate possible causes
• Finding the Vital Few causes (factors)
• Implementing process changes
• Optimizing the new process
• Controlling the new process to maintain the gains

Define    Measure    Analyze    Improve    Control
What questions do you have for me?
Appendix
Real Example

- Prolonged wait times to be triaged
- LWBS (2015): 4.61%
- Avg 10 /day for April
- Not a clear understanding of Arrival-to-Doc time
- LOS (2015): 201.6
Patient Arrival by Day and Hour

- **Monday** is the highest volume day of the week.
- **9am – 7pm** are the highest volume hours of the day.
Real Example

MD / RN Staffing by Hour

- Patient-RN ratio is above 3.5 from 8am-10am

Avg Patient Load per RN/MD

- Patient-MD Ratio is above 12 from 6am-12pm

1 MD at 6am-7am
Real Example Results

Door to Room

- Before: 60.1
- After: 19.6

Door to Doc

- Before: 66.2
- After: 24.3

ED LOS

- Before: 201.6
- After: 82.4

LWBS

- Before: 4.60%
- After: 0%