Application of Statistical Methodology to Transactional Processes

September 2016
T.C. Simpson and M.E. Rusak
What is Different/The Same?

• It’s still Data

• Measurement System Analysis (MSA)
  - Might be non-traditional – still important

• Statistical Methods
  - Non-normal; need non-parametric methods
  - Hypothesis tests are relevant
Outline

Define
- Voice of the Customer (VOC)
- Failure Modes and Effects Analysis (FMEA) and t-tests

Measure
- Developing Measurement Systems
- Attribute Agreement Analysis

Analyze
- Moods median; t-test; Design of Experiments (DOE)

Improve
- Laney P; Cross tabulation; autocorrelation

Control
- Controlling and sustaining
Define

- VOC – surveys/data mining; statistical analysis
- FMEA and t-tests
Role of Minitab in Define Phase

• Data preparation

• Initial graphical analysis

• Hypothesis testing
**VOC – Surveys/Data Mining**

500 surveys. >100 respondents. >60 verbatims; affinitized

Affinitized responses used for initial graphical analysis

<table>
<thead>
<tr>
<th>RespondentID</th>
<th>Region</th>
<th>Business Area</th>
<th>Free Form Response</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>2821196988</td>
<td>Region 1</td>
<td>Electronics and Performance Materials</td>
<td>increase network bandwidth to each office so that cad, database, and other large files</td>
<td>Processing Speed/Bandwidth</td>
</tr>
<tr>
<td>2820561822</td>
<td>Region 2</td>
<td>Merchant Gases - Packaged Gases</td>
<td>access to major applications (APOnline Share, Workbooks) by mobile phone.</td>
<td>Mobile Access</td>
</tr>
<tr>
<td>2820480519</td>
<td>Region 1</td>
<td>Global Operations</td>
<td>It would be great to be able access the APCI intranet with a company mobile smartphone,</td>
<td>Mobile Access</td>
</tr>
<tr>
<td>28195938271</td>
<td>Region 1</td>
<td>Electronics and Performance Materials</td>
<td>Is there an application that runs on windows 7 that allows us to view Microstation files</td>
<td>Collaboration (Search, App/File Sharing)</td>
</tr>
<tr>
<td>2819697237</td>
<td>Region 1</td>
<td>Technology/R&amp;D</td>
<td>Better file sharing via internal version control service (i.e., &quot;cp dropbox&quot; or like &quot;repo github&quot;)</td>
<td>Collaboration (Search, App/File Sharing)</td>
</tr>
<tr>
<td>2819888490</td>
<td>Region 1</td>
<td>IHS &amp;Q</td>
<td>The speed of the Internet. A significant portion of my time is searching regulatory</td>
<td>Collaboration (Processing, Speed/Bandwidth)</td>
</tr>
<tr>
<td>2819884688</td>
<td>Region 1</td>
<td>Technology/R&amp;D</td>
<td>Sometimes the content filter on Explorer is too aggressive and blocks relevant searches.</td>
<td>Collaboration (Search, App/File Sharing)</td>
</tr>
<tr>
<td>2817089704</td>
<td>Region 2</td>
<td>Technology/R&amp;D</td>
<td>More simple the structure to find standard documents in AP-Shares. Support. We are being asked to do more and</td>
<td>Collaboration (Search, App/File Sharing)</td>
</tr>
<tr>
<td>2816597503</td>
<td>Region 1</td>
<td>Global Operations</td>
<td></td>
<td>Support</td>
</tr>
</tbody>
</table>

**Pareto Chart of Category 1**

**Pareto Chart of Category 1 by Region**
## Chi-Square Test for Association: Category_1, Region

<table>
<thead>
<tr>
<th></th>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaboration</td>
<td>10</td>
<td>5</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>9.967</td>
<td>6.230</td>
<td>2.803</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0001</td>
<td>0.2427</td>
<td>0.5109</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0001</td>
<td>0.2427</td>
<td>0.5109</td>
<td></td>
</tr>
<tr>
<td>Mobile Access</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>6.295</td>
<td>3.934</td>
<td>1.770</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.4617</td>
<td>0.2219</td>
<td>0.3353</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Processing Speed</td>
<td>4</td>
<td>9</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>8.393</td>
<td>5.246</td>
<td>2.361</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.2997</td>
<td>2.6865</td>
<td>0.1732</td>
<td></td>
</tr>
<tr>
<td>Support</td>
<td>10</td>
<td>3</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>7.344</td>
<td>4.590</td>
<td>2.066</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9603</td>
<td>0.5509</td>
<td>0.5497</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>32</td>
<td>20</td>
<td>9</td>
<td>61</td>
</tr>
</tbody>
</table>

**Cell Contents:**
- Count
- Expected count
- Contribution to Chi-square

Pearson Chi-Square = 8.993, DF = 6, P-Value = 0.174
Likelihood Ratio Chi-Square = 9.172, DF = 6, P-Value = 0.164

* NOTE * 6 cells with expected counts less than 5
Chi-Square Test for Association: Category_3, Region_3

Rows: Category_3  Columns: Region_3

<table>
<thead>
<tr>
<th></th>
<th>Region 1</th>
<th>Region 2</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaboration</td>
<td>10</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>Processing Speed</td>
<td>4</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>All</td>
<td>14</td>
<td>14</td>
<td>28</td>
</tr>
</tbody>
</table>

Cell Contents: Count

Expected count

Contribution to Chi-square

Pearson Chi-Square = 3.590, DF = 1, P-Value = 0.058
Likelihood Ratio Chi-Square = 3.673, DF = 1, P-Value = 0.055
The FMEA risk evaluation done by two teams as part of the initial Sources of Variation (SOV)/subjective analysis. Used paired t-test to determine if there were differences.
• Developing a Measurement System
• Attribute Agreement Analysis
How do you complete an MSA in a non-production environment?

• Can’t always evaluate all
• Evaluate those that you can
• Know your sample size and measurement needs
  - What is it going to be used for?
  - What level of precision do you need?
Developing and Refining a Measurement System

Measurement system needed for future DOE

Don’t settle for inconsistent measures....

Makes your future work much more difficult!
Refining a Measurement System

Selected Improved Test Environment
• Microsoft Visual Studio – Allows hundreds of runs per test! (compared to six runs total)

Develop Acceptable Test Protocol
• “Warmup” time, length of run, frequency of measures, etc.
Refining a Measurement System

Improved Test Procedure
Only needed four replicates per DOE Test instead of nine replicates!

Reliable Test System Developed!
Other Measurement System Analyses – Attribute Agreement

• Design is a key
  - Ensure adequate sample size
• Randomize and include replicates where possible
Ticket Analysis

Attribute MSA, 50 complete tickets; Used three operators; randomized the 50 sample tickets had operators categorize the tickets. Added replication and compared to standard.
Reproducibility. Only 67 out of 150 matched or 44.67%.
### MSA - Attribute Agreement Analysis

#### Fleiss' Kappa Statistics

<table>
<thead>
<tr>
<th>Response</th>
<th>Kappa</th>
<th>SE Kappa</th>
<th>Z</th>
<th>P (vs &gt; 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absences and attendances</td>
<td>1.00000</td>
<td>0.0471405</td>
<td>21.2132</td>
<td>0.0000</td>
</tr>
<tr>
<td>Advance Payment</td>
<td>0.73668</td>
<td>0.0471405</td>
<td>15.6273</td>
<td>0.0000</td>
</tr>
<tr>
<td>Agreement/Pact</td>
<td>0.75975</td>
<td>0.0471405</td>
<td>16.1167</td>
<td>0.0000</td>
</tr>
<tr>
<td>Basic Pay</td>
<td>0.75021</td>
<td>0.0471405</td>
<td>15.9143</td>
<td>0.0000</td>
</tr>
<tr>
<td>Contract</td>
<td>-0.01238</td>
<td>0.0471405</td>
<td>-0.2625</td>
<td>0.6035</td>
</tr>
<tr>
<td>password reset</td>
<td>0.37869</td>
<td>0.0471405</td>
<td>8.0332</td>
<td>0.0000</td>
</tr>
<tr>
<td>Pay Slip</td>
<td>0.46570</td>
<td>0.0471405</td>
<td>9.8790</td>
<td>0.0000</td>
</tr>
<tr>
<td>Pension Plan</td>
<td>1.00000</td>
<td>0.0471405</td>
<td>21.2132</td>
<td>0.0000</td>
</tr>
<tr>
<td>Position changes</td>
<td>0.86441</td>
<td>0.0471405</td>
<td>18.3368</td>
<td>0.0000</td>
</tr>
<tr>
<td>Report</td>
<td>0.65105</td>
<td>0.0471405</td>
<td>9.8790</td>
<td>0.0000</td>
</tr>
<tr>
<td>Social Security</td>
<td>0.77398</td>
<td>0.0471405</td>
<td>16.4187</td>
<td>0.0000</td>
</tr>
<tr>
<td>Taxes</td>
<td>0.51033</td>
<td>0.0471405</td>
<td>10.8258</td>
<td>0.0000</td>
</tr>
<tr>
<td>Overall</td>
<td>0.67462</td>
<td>0.0213520</td>
<td>31.5949</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

The Kappa Value* is below 0.7 for several categories being evaluated.

*Kappa Values: Values ≥0.7 are considered acceptable in most cases, and values ≥0.9 are excellent.

Measurement system needs to be improved.
**Key Questions**

- How will you ensure the data being collected is accurate and unbiased?
- What are the areas of possible bias or system errors?
- Have I assessed the Reproducibility & Repeatability aspects of my data collection process?
- Are data collectors properly trained?

**Alternate Process Analysis**

Attribute MSA, 30 scenarios; Used three operators; randomized the 30 sample tickets had operators evaluate them. Compared to standard. Replication not possible.
**MSA - Attribute Agreement Analysis**

**Key Questions**
- How will you ensure the data being collected is accurate and unbiased?
- What are the areas of possible bias or system errors?

**Measurement System Needs Improvement**
- Used results to identify Sources of Variability
  - Guidelines missing
  - Unusual
  - Maximum thresholds/exception process

*The Kappa Value is below 0.7 for several categories being evaluated*
Analyze

- Hypothesis testing; DOE
Hypothesis Testing

• Examine stability, shape and spread first
  - Many will be non-normal
  - Use appropriate distribution
• Use non-parametric tests when appropriate
Transform Data if Necessary

Box-Cox Plot of Cycle Time +1 for BC

Worksheet: Subset of Global Excluding Small Volumes, Long and Negative Cycle Times

Process Capability Report for Cycle Time +1 for BC
Using Box-Cox Transformation With $\lambda = -0.42$

Air Products Public
Alternate to Chi Square Tables

Often better for communication of effects
Mood Median Test: Cycle Time versus Customer Group

Mood median test for Cycle Time
Chi-Square = 2495.10    DF = 7
P = 0.000

<table>
<thead>
<tr>
<th>Customer Group</th>
<th>N≤</th>
<th>N&gt;</th>
<th>Median</th>
<th>Q3-Q1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemicals</td>
<td>445</td>
<td>827</td>
<td>7.00</td>
<td>29.00</td>
</tr>
<tr>
<td>Cryoease/Micro Bulk</td>
<td>1931</td>
<td>3040</td>
<td>3.00</td>
<td>10.00</td>
</tr>
<tr>
<td>Electronics</td>
<td>647</td>
<td>1004</td>
<td>4.00</td>
<td>18.00</td>
</tr>
<tr>
<td>Equipment</td>
<td>555</td>
<td>306</td>
<td>0.00</td>
<td>7.00</td>
</tr>
<tr>
<td>Liquid Bulk</td>
<td>5024</td>
<td>7231</td>
<td>4.00</td>
<td>26.00</td>
</tr>
<tr>
<td>PG</td>
<td>28157</td>
<td>18805</td>
<td>0.00</td>
<td>5.00</td>
</tr>
<tr>
<td>Sales Center Cust</td>
<td>473</td>
<td>742</td>
<td>3.00</td>
<td>7.00</td>
</tr>
<tr>
<td>Tonnage</td>
<td>71</td>
<td>165</td>
<td>5.50</td>
<td>15.00</td>
</tr>
</tbody>
</table>

Overall median = 1.00

Particularly good with non-normal data; when population has outliers
## DOE – Transactional Applications

- Two Factor, Two Level Model
- Four Replicates per Test (16 runs)

<table>
<thead>
<tr>
<th>Optimization</th>
<th>Utilization</th>
<th>Run ID</th>
<th>Run ID</th>
<th>Run ID</th>
<th>Run ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off</td>
<td>Low</td>
<td>3065</td>
<td>3067</td>
<td>3071</td>
<td>3075</td>
</tr>
<tr>
<td>Off</td>
<td>High</td>
<td>3069</td>
<td>3070</td>
<td>3077</td>
<td>3078</td>
</tr>
<tr>
<td>On</td>
<td>Low</td>
<td>3066</td>
<td>3072</td>
<td>3074</td>
<td>3080</td>
</tr>
<tr>
<td>On</td>
<td>High</td>
<td>3068</td>
<td>3076</td>
<td>3079</td>
<td>3081</td>
</tr>
</tbody>
</table>
## DOE – Results

### I-MR Chart of Elapsed Time

- **Run**: 1, **Average**: 2.21, **Standard Deviation**: 4.65
- **Run**: 2, **Average**: 2.00, **Standard Deviation**: 3.52
- **Run**: 3, **Average**: 2.10, **Standard Deviation**: 4.37
- **Run**: 4, **Average**: 1.96, **Standard Deviation**: 4.16

*Optimization Off; Utilization Low*
DOE – Results

Main Effects Plot

Main Effects Plot for Time (Sec)
Data Means

Optimization

Utilization

Mean

Off  On  High  Low

Worksheet: Worksheet 1; 6/13/2016 3:49:38 PM
DOE – Results

Do not forget to consider interactions!!

Results for: DoE Run

Factorial Fit: Mean Response (ms) versus Optimization, Utilization

Estimated Effects and Coefficients for Mean Response (ms) (coded units)

<table>
<thead>
<tr>
<th>Term</th>
<th>Effect</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td></td>
<td>6.1675</td>
<td>0.1622</td>
<td>38.02</td>
<td>0.000</td>
</tr>
<tr>
<td>Optimization</td>
<td>-0.2835</td>
<td>-0.1417</td>
<td>0.1622</td>
<td>-0.87</td>
<td>0.399</td>
</tr>
<tr>
<td>Utilization</td>
<td>9.4075</td>
<td>4.7037</td>
<td>0.1622</td>
<td>29.00</td>
<td>0.000</td>
</tr>
<tr>
<td>Optimization*Users</td>
<td>0.9235</td>
<td>0.4617</td>
<td>0.1622</td>
<td>2.85</td>
<td>0.015</td>
</tr>
</tbody>
</table>

S = 0.648854    PRESS = 8.98157
R-Sq = 98.61%    R-Sq(pred) = 97.52%    R-Sq(adj) = 98.26%

Optimization Effective at Low Line Utilization Levels, But Not at High Utilization Levels

Not Significant On Its Own
Model Represents Data Well
• Cross tabulation; autocorrelation
Evaluate Shape and Stability

Count of defects per day for about one year. Maximum defects 63!

Evidence of lack of stability; Team believed this was due to cyclic behavior.
Autocorrelation

Correlation between the elements of a series and others from the same series separated from them by a given interval.

Autocorrelation is also sometimes called “lagged correlation” or “serial correlation”, which refers to the correlation between members of a series of numbers arranged in time.
Defects were not related to weekly, monthly or other business cycles.
What Would be Different if There Were a Cyclic Effect?

Autocorrelation

What might we see if there were a timed effect? We added a random number to each count on every seventh day:

Worksheet: Pre Improve; 5/12/2016 10:56:13 AM
This is an example of what you would see with significant patterns on a seven day cycle.
Results of Improvement

Count of failures per day for about one year.
Average of about one failure/day
Only eight failures/day maximum!

Overall improvement of about 87%
Cross Correlation

No evidence of correlation --- but what if you suspect that the two variables are related? What is missing?
Cross Correlation

Worksheet: Worksheet 1

Computes and graphs correlations between two time series.
Cross Correlation

Use cross correlation to determine if there is a time-shift element in your data.
• Laney P’, Controlling and sustaining
When Subgroup Size is Large – Laney P’ to Avoid Overdispersion (This Example or Underdispersion)

Tests are performed with unequal sample sizes.
Worksheet: Worksheet 1
Evidence of Overdispersion

P Chart Diagnostic for Invoices Not Paid within Terms

Binomial Probability Plot

Ratio of observed variation to expected variation = 1373.4%
95% Upper Limit for ratio if process P is constant = 145.1%

Using a P chart may result in an elevated false alarm rate. Consider using a Laney P' chart instead.
The upper limit depends on the number of subgroups, the average subgroup size, and the overall process P.

Worksheet: Worksheet 1
MSA - Attribute Agreement Analysis

Results after process improvement
The Kappa Values are now above the threshold to indicate an acceptable Measurement System.

### Fleiss' Kappa Statistics

<table>
<thead>
<tr>
<th>Response</th>
<th>Kappa</th>
<th>SE Kappa</th>
<th>Z</th>
<th>P(vs &gt; 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absences and attendances</td>
<td>1.0000</td>
<td>0.1054</td>
<td>9.486</td>
<td>0.0000</td>
</tr>
<tr>
<td>Advance Payment</td>
<td>1.0000</td>
<td>0.1054</td>
<td>9.486</td>
<td>0.0000</td>
</tr>
<tr>
<td>Agreement/Pact</td>
<td>0.8830</td>
<td>0.1054</td>
<td>8.377</td>
<td>0.0000</td>
</tr>
<tr>
<td>Bank details</td>
<td>1.0000</td>
<td>0.1054</td>
<td>9.486</td>
<td>0.0000</td>
</tr>
<tr>
<td>Basic Pay</td>
<td>0.9273</td>
<td>0.1054</td>
<td>8.796</td>
<td>0.0000</td>
</tr>
<tr>
<td>Contract</td>
<td>0.8214</td>
<td>0.1054</td>
<td>7.792</td>
<td>0.0000</td>
</tr>
<tr>
<td>password reset</td>
<td>0.8830</td>
<td>0.1054</td>
<td>8.377</td>
<td>0.0000</td>
</tr>
<tr>
<td>Pay Slip</td>
<td>0.9273</td>
<td>0.1054</td>
<td>8.796</td>
<td>0.0000</td>
</tr>
<tr>
<td>Pension Plan</td>
<td>1.0000</td>
<td>0.1054</td>
<td>9.486</td>
<td>0.0000</td>
</tr>
<tr>
<td>Position changes</td>
<td>1.0000</td>
<td>0.1054</td>
<td>9.486</td>
<td>0.0000</td>
</tr>
<tr>
<td>Report</td>
<td>1.0000</td>
<td>0.1054</td>
<td>9.486</td>
<td>0.0000</td>
</tr>
<tr>
<td>Social Security</td>
<td>0.9407</td>
<td>0.1054</td>
<td>8.925</td>
<td>0.0000</td>
</tr>
<tr>
<td>Taxes</td>
<td>0.8545</td>
<td>0.1054</td>
<td>8.107</td>
<td>0.0000</td>
</tr>
<tr>
<td>Overall</td>
<td>0.9373</td>
<td>0.0355</td>
<td>26.368</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

**Measurement system has been improved**
Validation of Measurement System After Improvement

Measurement System Acceptable: Individual Kappa values were between 77% and 79% for each appraiser vs std; Overall Kappa between appraisers was 70%; Overall Kappa vs Std was also 70%
Summary

• Use of Minitab for graphical and statistical analysis is appropriate for many transactional applications

• Through this process, greater confidence in your decisions is achieved

• Evaluating sample size, shape of your population, and the basic characteristics of your population will lead to improved data analysis

• Many of the same methods used in production environments are appropriate for transactional applications

• Customizing your approach and sampling methodology will improve the validity of your analysis
Thank you
tell me more