Advancement of decision-making in Agile Projects by applying Logistic Regression on Estimates

Lakshminarayana Kompella
Engineering Program Director
CA Technologies
Agenda

* Introduction
* Challenges in Agile projects
* Group Estimation
* Suggested Methodology
* Results & Discussion
* Conclusion
**Introduction**

* Concept of Agility
  * “a priori characterization” [13] [10]
  * “emergent” [13] [10]

* Agility
  * apart from change & customer value
  * economy, quality and simplicity [10]
  * Individual techniques & principles or property that characterizes the whole method [13]

* Agile Methods and Organizational Culture
  * complex interplay [10]
  * Competent Value Method [13]
    * change vs. stability and internal vs. external focus

The readiness of an ISD method to rapidly or inherently create change, proactively or reactively embrace change and learn from change while contributing to perceived customer value (economy, quality and simplicity), through its collective components and relationships with its environment k. Conboy [10]
Challenges

* Agile constructs
  * Shared decision-making; team-work; self-aligning [4] [5][18] [10] [18]
* Dynamic changes
  * Sequence of “plan the iteration”; “commit to a goal”; “execute”; “demo”; “retrospective” [2][8][9]
* Team and Dynamic Changes
  * Tools that augment decision-making [17]
Estimation & Uncertainty

* Uncertainty
  * Break requirements into smaller components [8][9][16]
  * Frequent feedback [2]

* Estimation (no framework)
  * Group estimation ("planning-poker"; wide-band Delphi method) [15][14]
    * accuracy and solving complex problems – Literature has numerous examples [1][12][7][11]
  * Social interaction and group dynamics [7]
* Dependent (Y) and independent (X_i) variables
  * Known-unknowns and unknown-unknowns
  * Dependent variable is categorical & binary
  * Independent variables, not necessarily continuous
* Identification of variables (+ve & -ve influence)
  * Dialectical interplay
* Define Hypotheses
Model can be represented by equation

\[ \ln \left( \frac{p_i}{1 - p_i} \right) = A + \sum B_i X_i \]

where \( p_i \) is the odds, \( A \) is the constant, \( B_i \) are the coefficients and independent variables \( X_i \)

- Circulate Questionnaire
  - Number of cases = Questions * Members [25*17 = 425]
  - Respondents to answer Yes/No
- Run Logistic Regression Test
- Validate model
* Model’s appropriateness assessed by [19]
  * Regression coefficients are arrived at by maximum likelihood estimation, an iterative method.
  * Omnibus test of model coefficients
    * Chi-square test of the model. Whether to accept/reject the NULL hypothesis
  * Assessment of Goodness of fit:
    * Deviance is used and is given by \((-2 \ln \text{likelihood of (fitted/saturated model)})\)
    * \(R^2_L = (D – D_{\text{model}})/D\)
    * Cox & Snell \(R^2\) and Nagelkerke \(R^2\) are used for goodness of fit indexes
  * Hosmer and Lemeshow test
    * Test statistic that follows Chi-square
    * Probability that the observed value and the expected value are the same
  * Wald statistic to test the significance
    * The variables \(X_i\) and their significance
    * Analogous to t-test in linear regression
* Analyze Results
Example

- Telecommunications Domain
  - Product: Edge Router (ATM)
- Release cycle: 6 months
- 7 Epics/Themes
- Team Details
  - Geographies: Stockholm, Sweden & Hyderabad, India
  - Team size: 30
- Dependent variable [completing the project on schedule]
- Independent Variables
  - Voluntary requirement changes ($X_1$)
  - Involuntary requirement changes ($X_2$)
  - Quantum of unplanned additional re-work to meet quality criteria ($X_3$)
  - Impediments that result out of process delays due to cross-border factors ($X_4$)
  - Unavoidable Absenteeism including attrition and the possible self-aligning to compensate reduction in sprint velocity when project is in-flight ($X_5$)
### Questionnaire: Possible combinations of independent variables

Based on the following conditions do you believe that the project can be completed on-time?

<table>
<thead>
<tr>
<th>S. No</th>
<th>Description</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Changes to requirement identified by the product owner. ((X_1)).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>More than identified changes required to implement the requirements. ((X_2)).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>More than planned re-work is possible. ((X_3)).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Process delays due to distributed teams can be accommodated.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Absenteeism and resulting compensation in sprint velocity when project is in-flight. ((X_5)).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Effect due to ((X_1) &amp; (X_2))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Effect due to ((X_1) &amp; (X_3))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Effect due to ((X_1) &amp; (X_4))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Effect due to ((X_1) &amp; (X_5))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Effect due to ((X_2) &amp; (X_3))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Effect due to ((X_2) &amp; (X_4))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Effect due to ((X_2) &amp; (X_5))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>..... and so on</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Format to enter into Statistical analysis tool**

<table>
<thead>
<tr>
<th>Epics/Themes</th>
<th>(X_1)</th>
<th>(X_2)</th>
<th>(X_3)</th>
<th>(X_4)</th>
<th>(X_5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact on Project Schedule due to Epics/Themes ‘A’</td>
<td>1 0 0 0 0</td>
<td>0 1 0 0 0</td>
<td>0 0 1 0 0</td>
<td>0 0 0 1 0</td>
<td>(\ldots) (\ldots) (\ldots) (\ldots) (\ldots)</td>
</tr>
</tbody>
</table>
Results & Discussion (1/3)

Estimated Coefficients of independent variables & Constant

<table>
<thead>
<tr>
<th></th>
<th>Bi</th>
<th>S.E</th>
<th>Wald</th>
<th>Df</th>
<th>Sig</th>
<th>Exp (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>-0.855</td>
<td>0.192</td>
<td>0.192</td>
<td>1</td>
<td>0.000</td>
<td>0.425</td>
</tr>
<tr>
<td>X2</td>
<td>-3.322</td>
<td>0.790</td>
<td>17.683</td>
<td>1</td>
<td>0.000</td>
<td>0.036</td>
</tr>
<tr>
<td>X3</td>
<td>-2.281</td>
<td>0.48</td>
<td>22.579</td>
<td>1</td>
<td>0.000</td>
<td>0.102</td>
</tr>
<tr>
<td>X4</td>
<td>-0.541</td>
<td>0.376</td>
<td>1.87</td>
<td>1</td>
<td>0.171</td>
<td>0.598</td>
</tr>
<tr>
<td>X5</td>
<td>0.007</td>
<td>0.363</td>
<td>0.000</td>
<td>1</td>
<td>0.985</td>
<td>1.007</td>
</tr>
<tr>
<td>A</td>
<td>-1.148</td>
<td>0.202</td>
<td>32.293</td>
<td>1</td>
<td>0.000</td>
<td>0.317</td>
</tr>
</tbody>
</table>

Wald = B/SE² – analogous to t-test in linear regression

\[
\ln \left( \frac{p_i}{1 - p_i} \right) = -1.148 - 0.855 X1 - 3.322 X2 - 2.281 X3
\]

\[
\left[ \frac{p_i}{1 - p_i} \right] = e^{-1.148 - 0.855 X1 - 3.322 X2 - 2.281 X3}
\]
### Results & Discussion (2/3)

#### Omnibus Tests of Model Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Chi-Square</th>
<th>Df</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>135.225</td>
<td>6</td>
<td>.000</td>
</tr>
<tr>
<td>Block</td>
<td>135.225</td>
<td>6</td>
<td>.000</td>
</tr>
<tr>
<td>Model</td>
<td>135.225</td>
<td>6</td>
<td>.000</td>
</tr>
</tbody>
</table>

From Table-IV Omnibus Test of coefficients significance (Sig.) value is 0.00 (<0.05), Chi-Square test is significant for the model. Therefore, the selected model does explain the dependent variable and the NULL Hypothesis (model doesn’t explain the dependent variable) is rejected. There is also no significant change from Step to Block and to Model.

#### Hosmer and Lemeshow Test

<table>
<thead>
<tr>
<th>Step</th>
<th>Chi-Square</th>
<th>Df</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.481</td>
<td>7</td>
<td>1.00</td>
</tr>
</tbody>
</table>

From Table-V, Hosmer and Lemeshow Test significance (Sig.) value is 1.00 (>0.05) and therefore the probability of observed value and expected value (NULL Hypothesis) is not rejected.

#### Model Summary

<table>
<thead>
<tr>
<th>Step</th>
<th>-2 Log Likelihood</th>
<th>Cox &amp; Snell R square</th>
<th>Nagelkerke R Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>857.889</td>
<td>0.129</td>
<td>0.203</td>
</tr>
</tbody>
</table>

In Table-VI, Nagelkerke R Square value is 0.203 or 20.3% of the selected independent variables explain the dependent variable.
Results & Discussion (3/3)

* Significant variables are
  * Voluntary requirement changes ($X_1$)
  * Involuntary requirement changes ($X_2$)
  * Quantum of unplanned additional re-work to meet quality criteria ($X_3$)
* $X_2$ & $X_3$ have higher impact as compared to $X_1$
  * Focus on activities contributing towards emergent agility
* $X_4$ & $X_5$ not significant
  * Contrary to popular team belief
A scientific approach to decision-making that involves group activity and social interaction.

- Built on advantageous of Group activity (for e.g. Group Estimation) and also captures social interactions.
- The specific actions required for an organization to effect cultural changes so as to drive growth in products/services.
- The suggested method can also provide insights into how different techniques and principles of agile methods support emergent agility and the situations that are required. Moreover, cultural changes required to establish emergent agility can also be identified.
- Literature had numerous references on the relationships between organizational culture and Agile methods with theoretical constructs and suggested quantitative confirmatory tests, to be taken up later, that organizations’ could attempt to influence their culture(s) towards a specific goal or objective. [13]
- Deep incorporation of agility by leveraging both “a priori characterization” and “emergent” agility - a step towards next process of software development [3]
References


Questions
Thank You