Secrets of Agile Estimation: Myths, Math, and Methods

Agile2014
July 31, 2014
Stephen Vance
About Me

• Live near Boston, MA
• Originally from Detroit, MI
• Author of [book title]
• Independent Lean/Agile Software Development Coach currently at Zipcar
• Software Craftsman
• Senior Pet Care Engineer
How Long Will It Take?

- Opinions
- Measurements
- Models
- Forecasts
Opinions

- Human-generated values
- Also known as
  - Guesses
  - Guestimates
  - Educated guesses
  - SWAGs
  - Estimates
- Examples include
  - Story points
  - T-shirt sizes
  - Hour “estimates”
Measurements

• Observed, empirical values
• Also known as
  – Observations
  – Assessments
  – Estimates
• Examples include
  – Temperature
  – Length
  – Cycle time
Models

• Description of behavior

• Often mathematical

• Also known as
  – Equations
  – Frameworks
  – Behaviors

• Examples include
  – Targeting
  – Orbital mechanics
  – Linear extrapolation

\[ \sum (\bar{x} - \bar{x}_G)^2 n \]
Forecasts

• Statements about the future from applying models to opinions and measurements

• Also known as
  – Predictions
  – Estimates

• Examples include
  – Weather
  – Capacity planning
  – Delivery dates
Instantaneous Completion Likelihood
Cumulative Completion Likelihood
Practically Speaking ...
Management Multipliers

1X 50%

2X 91.7%

3X 98.6%
Data Gathering for Forecasting

- Time
- Relative sizing
  - Story points
  - T-shirt sizing
  - Confidence factors
- Cycle time
Time

• Opinion

• Examples
  – Hours
  – Days
  – Real days

• Resembles familiar units but unreliable and completely different

• Generally, people are horrible at estimating larger spans of time
Relative Sizing

• Opinion
• “Relative” is essential!
• Consensus is a form of normalization
• Only valid across short time intervals
• Only applies from point at which opinion is rendered
• You can’t compare different
  – Teams
  – Problem domains
  – Skills
  – Technologies
Why Relative Sizing Works

Points

1 Sprint

5 1 3 2
Story Points

• Don’t normalize!
• If it uses story point as a unit, don’t compare it.
• Be careful with story point math
  – Story points aren’t precise
  – Larger values have higher uncertainty
T-shirt Sizes

• “Buckets”
• Comparing adjacent sizes
• Prevents sizing math
  – For better and worse
• Great for epics!
Confidence Factors

- Normalize the risk perspective
- Ask for a 50% likelihood opinion
  - Reasonable happy case
- Vote on a confidence factor
  - 1 Completely
  - 2 Sorta
  - 3 Not at all
Cycle Time

- Measurement
- Applies when work is identified
- Applies across subsets of the process
- Normalizes for roll up forecasts
  - Don’t use for individual team comparisons
Velocity

- Not just for Scrum and story points
- Useful for predicting near-term capacity
- Any other use should be handled with care
- Sliding window velocity (related to valid reporting window)
Common Velocity Model

- **Velocity = S/P/D**
  - S is Size, e.g. story points, time
  - P is unit of Production, e.g. team, person, unit team strength
  - D is Duration, e.g. sprint, day, week

- Valid across estimation types
- Scrum
  - Points/team/sprint
- Cycle time
  - Items/system/unit time
# Team Strength

<table>
<thead>
<tr>
<th>Event</th>
<th>People</th>
<th>Days</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>5</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>Joe Vacation</td>
<td>1</td>
<td>5</td>
<td>-5</td>
</tr>
<tr>
<td>Team Training</td>
<td>5</td>
<td>.5</td>
<td>-2.5</td>
</tr>
<tr>
<td>Holiday</td>
<td>5</td>
<td>1</td>
<td>-5</td>
</tr>
<tr>
<td><strong>Sub-Total</strong></td>
<td></td>
<td></td>
<td><strong>37.5</strong></td>
</tr>
<tr>
<td><strong>Strength Coefficient</strong></td>
<td></td>
<td></td>
<td><strong>.75</strong></td>
</tr>
</tbody>
</table>
## Applying Team Strength

<table>
<thead>
<tr>
<th>Sprint</th>
<th>Predicted Strength</th>
<th>Strength Coefficient</th>
<th>Points Completed</th>
<th>Effective Velocity</th>
<th>Rolling Average</th>
<th>Predicted Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Opinion</td>
<td>Predicted/Nominal</td>
<td>Observation</td>
<td>Completed/Coefficient</td>
<td>3 Sprints</td>
<td>Rolling * Coefficient</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>.75</td>
<td>15</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td>1.00</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>9</td>
<td>35</td>
<td>.88</td>
<td>15</td>
<td>17</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>1.00</td>
<td>16</td>
<td>16</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>11</td>
<td>20</td>
<td>.50</td>
<td></td>
<td></td>
<td>18</td>
<td>9</td>
</tr>
</tbody>
</table>

https://github.com/srvance/AgileEstimation
Abuses of Velocity

• Evaluating productivity
• Evaluating process improvement
  – Except at the coarsest level
  – “After two consecutive sprints at 0 velocity we’ve managed to get to a consistent velocity of more than 20 for several sprints now!”
• Evaluating individuals
• Predicting individual task completion
• Forecasting precisely
Little’s Law

“The long-term average number of customers in a stable system $L$ is equal to the long-term average effective arrival rate, $\lambda$, multiplied by the (Palm-)average time a customer spends in the system, $W$; or expressed algebraically: $L = \lambda W$.”

$\lambda = L/W$ is a formulation of velocity
Translation

Throughput = WIP / Cycle Time
Cumulative Flow Diagrams (CFDs)

A–Team Cumulative Flow Diagram

Days

WIP

Throughput

Cycle Time

2013-06-05

Ready To Pull In Dev Code Review Accepted Merged Released

https://github.com/wrackzone/kanban-cumulative-flow-chart
Predictive vs. Analytical Use

A-Team Cumulative Flow Diagram

https://github.com/wrackzone/kanban-cumulative-flow-chart
The “Same Size” Misconception

Cycle Time (days)  Running Average
Am I Ready For Cycle Time?

<table>
<thead>
<tr>
<th>Week</th>
<th>Cycle Time</th>
<th>2-week</th>
<th>3-week</th>
<th>4-week</th>
<th>5-week</th>
<th>6-week</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>5.0</td>
<td>5.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>5.0</td>
<td>3.3</td>
<td>3.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0.0</td>
<td>3.3</td>
<td>2.5</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>7.5</td>
<td>5.0</td>
<td>6.3</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>10.0</td>
<td>6.7</td>
<td>5.0</td>
<td>6.0</td>
<td>5.0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>2.5</td>
<td>6.7</td>
<td>5.0</td>
<td>4.0</td>
<td>5.0</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>5.0</td>
<td>5.0</td>
<td>7.5</td>
<td>6.0</td>
<td>5.0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>5.0</td>
<td>3.3</td>
<td>3.8</td>
<td>6.0</td>
<td>5.0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0.0</td>
<td>3.3</td>
<td>2.5</td>
<td>3.0</td>
<td>5.0</td>
</tr>
<tr>
<td>12</td>
<td>15</td>
<td>7.5</td>
<td>5.0</td>
<td>6.3</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>Spread</td>
<td>15.0</td>
<td>10.0</td>
<td>3.3</td>
<td>5.0</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Std Dev</td>
<td>6.3</td>
<td>3.1</td>
<td>1.3</td>
<td>1.7</td>
<td>1.3</td>
</tr>
</tbody>
</table>
Statistical Process Control Chart

Quality characteristic

Sample

11.0

10.0

9.0

UCL = 10.860

Center line = 10.058

LCL = 9.256

http://upload.wikimedia.org/wikipedia/commons/thumb/f/f7/ControlChart.svg/1000px-ControlChart.svg.png
Real SPC Example
Zooming In
Predictions As Probabilities

• You can never give absolute commitments
• Sometimes you
  – Succeed
  – Fail
  – Work more hours
  – Add more people
  – Trim the scope
  – Cut corners
• Consciously choose and communicate your likelihoods
### Cycle Time Story Time

<table>
<thead>
<tr>
<th>Story Points</th>
<th>Cycle Time (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>17</td>
</tr>
</tbody>
</table>

5X spread fits in 17.5±2.5 days: 14% accuracy by assuming stories are all the same size!
Comparative Estimation Case Study

• For a release
  – 78 story points
  – 47 stories
  – 25 working days
  – 18 day average cycle time
  – 1.9 average points per story

• 78/25 =
  – 3.12 points/day/team

• 47/18 =
  – 2.61 stories/day/team

• 3.12/2.61 =
  – 1.19 points/story

• 1.19/1.9 = .63

  40% slack in story point velocity!
Contact Me

steve@vance.com
@StephenRVance
http://www.vance.com
LinkedIn: srvance
GitHub: srvance