

Using Monte Carlo Simulation to Build Trust

How I Learned to Stop Worrying and Love the Uncertainty

MCKESSON



LeanKanban
North America 2016

David Rosecrans
Director of Software Development
McKesson Health Solutions
[LinkedIn](#)



History of an Agile Group

That which does not kill us, makes us stronger

- Friedrich Nietzsche

History of an Agile Group

- Founded as an Internal Software Startup in Late 2008
- Founders Were Committed to Agile Development Practices
- Showing Progress and Adaptability to Change Was Valued
- Began with SCRUM, Switched to Lean/Kanban in 2010, as Customer Demand Grew
- In 2011, Became Part of an Established Business Unit
- Progress and Adaptability Became Less Important
- Long Range Predictability and Meeting Commitments Became More Important



Why Do Businesses Make Commitments?

“It’ll be done when it’s done!” they say, expecting that such a brave, funny zinger will reduce their boss to a fit of giggles, and in the ensuing joviality, the schedule will be forgotten. ⁽⁷⁾

Business Commitments

- *“A commitment refers to any action taken in the present that **binds** an organization to a future course of action”⁽¹⁾*
- Committing To Developing A Product Is An Investment
- No Point In Investing If There Isn't A Positive Return
- Understanding Return Requires Costs Over Time
- Expenses In Software Projects Are Scope and Productivity
- Assumptions Of Costs Built Are Built In To Business Plan
- Assumptions Become Expectations To Manage

If You Are Not Vigilant...

- Commitments Can Push You Away From Some Agile Practices
- Early Commitment
 - Change Is Discouraged Because We've Committed!
- Winging It (Increased Chance of Being Wrong)
 - Pushes Scope Decisions To Beginning of Product
 - Encourages Wrong People To Estimate
 - Superficial Requirements Or Knowledge Drives a Commitment
- Death By Planning (Increased Delays Prevent Early Learning)
 - Encourages Broader And Deeper Analysis Up Front
 - Increases Time Spent Estimating

The background of the slide is a soft, out-of-focus photograph of a landscape. It features a calm body of water, likely a lake, in the foreground, with rolling hills or mountains in the distance. The colors are muted and pastel-like, with shades of light blue, green, and white, creating a serene and contemplative atmosphere.

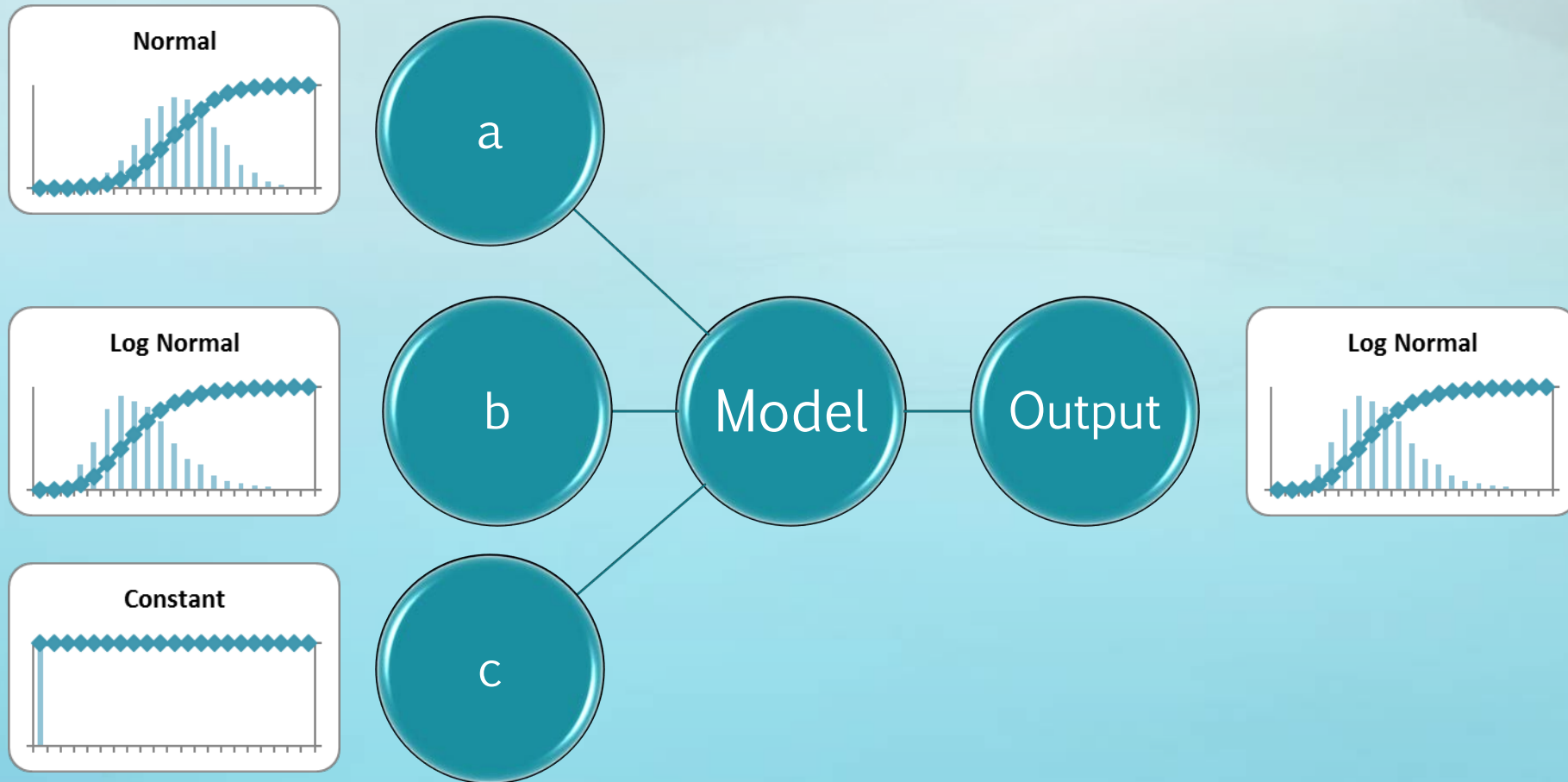
Why Monte Carlo Simulation?

“Maturity of mind is the capacity to endure uncertainty.”
- John Finley

Monte Carlo Simulation

- Leverages The Law Of Large Numbers
- Probability of Outcomes Calculated from Repeated Sampling of Probability of Inputs
- Inputs can Be Real, Simulated, Estimated or Even Constants
- Help's Visualize Risk and Uncertainty
- Overcomes Limitations Of Single Point Techniques

Monte Carlo Simulation



Example For Forecasting A Series of Tasks

- Single Developer
- 5 Tasks
- Historical Performance Data
 - Average 11 Days Per Task
- Simple Math $\# \text{Tasks} * \text{Average Days}$
 - Expect 55 Days To Complete

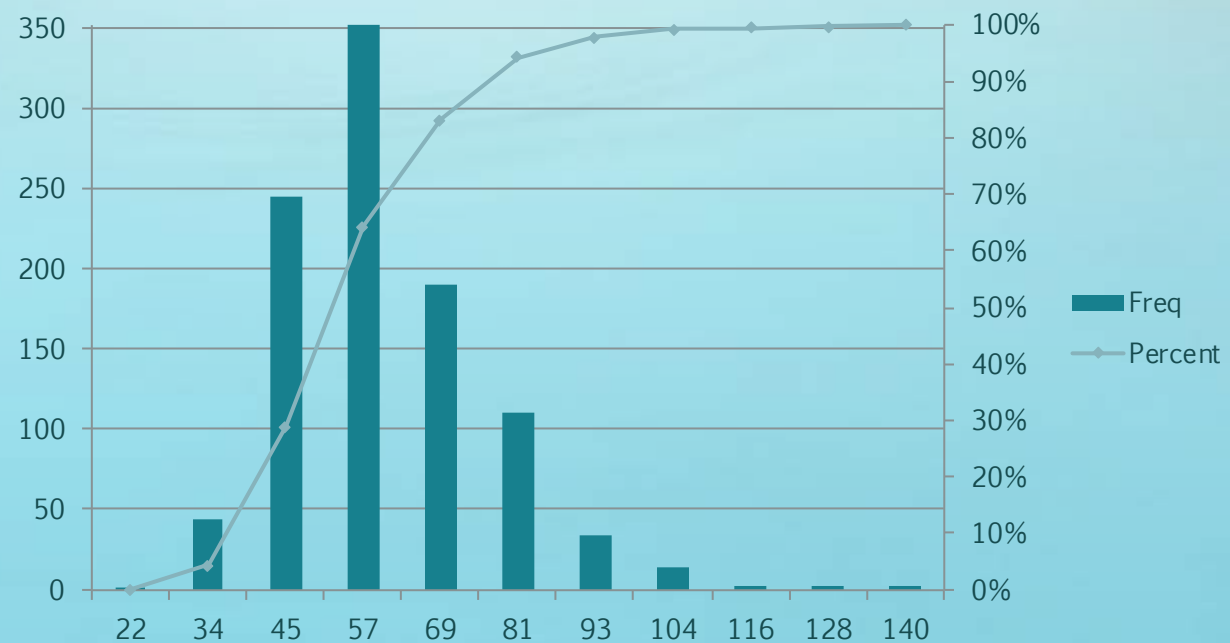
Same Example Monte Carlo Results

- Single Developer
- 5 Tasks
- Historical Performance Data
 - Average 11 days with
 - Standard Deviation 7 days
 - Positive or Right Skewed
 - Lognormal Will Be a Good Approximation ⁽⁶⁾
- 1000 Simulated Experiments
 - Expect Average to Still Be 55 Days, But Also Know
 - 50% of Experiments Were 52 Days or Less
 - 80% of Experiments Were Between 37 and 74 Days

The Results

	Result
Min	22
Max	140
Mean	55
StdDev	16
Percentile 10%	37
Percentile 50%	52
Percentile 90%	74

Days	Freq	Cum Freq	Percent
22	1	1	0%
34	44	45	5%
45	245	290	29%
57	352	642	64%
69	190	832	83%
81	111	943	94%
93	34	977	98%
104	14	991	99%
116	3	994	99%
128	3	997	100%
140	3	1000	100%



Won't Simple Math Work As Well?

- If We Had Used More Historical Data
 - 10% Tasks Finish In 4 Days
 - 50% Tasks Finish In 9 Days
 - 90% Tasks Finish In 21 Days
- Simple Math vs Monte Carlo Simulation
 - 20 ($4*5$) Days or Better Never Occurs or $< 0.1\%$ Chance Not 10%
 - 45 ($9*5$) Days or Better Occurs 298 Times or $<30\%$ Not 50%
 - 105 ($21*5$) Days or Better Occurs 991 Times or $<99.1\%$ Not 90%
- Monte Carlo Simulation Produces Less Exaggerated Results

The Results Allow For A Change In Language

- Towards Reliability Instead of Accuracy
- Towards Confidence, Uncertainty and Risk
- You Are No Longer Limited To Predicting “Average” Results

Supporting Agile Commitments

- Don't Spend More Time Estimating Than You Need To
- More Detailed Estimates Are Not Necessarily More Accurate
- Use The Estimating Techniques You Use Today
- Depend On Your Processes To Provide Consistency
- Collect Information To Inform Your Estimates
- Avoid Wishful Thinking
- Don't Exclude The Bad Data



The Model We Used For Schedule Forecasting

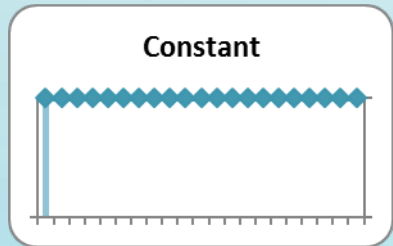
“There’s no point in being exact about something if you don’t even know what you’re talking about.”

—John von Neumann ⁽⁵⁾

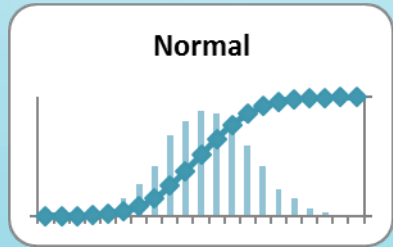
A Simple Model for Scheduling

- *Number of Months* =
$$\frac{\text{Number of Features}}{\text{Number of Features per Month}}$$
- Problem:
 - Number of Features was an Important Planning Number
 - Number of Features Per Month was a Commitment
- Reality:
 - We Don't Know How Many Features a Project Will Be
 - We Don't Know How Long It Will Take To Implement each Feature
 - We Do Have Lots of Data About How Projects Went In The Past
- Approach:
 - Approximate a Cumulative Distribution Function for Feature Growth
 - Approximate a Cumulative Distribution Function for Feature Duration
- *Number of Months* =
$$\frac{\sum_{i=1}^{g(\text{Number of Features})} \times \text{Number of Features } d(i)}{\text{Number of Features per Month}}$$

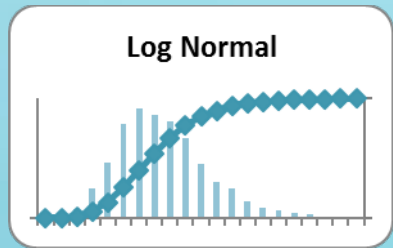
Selecting A Cumulative Distribution Shape



- Constant – Single Point, Only Value Possible
 - Useful for Modeling Single Point Calculations

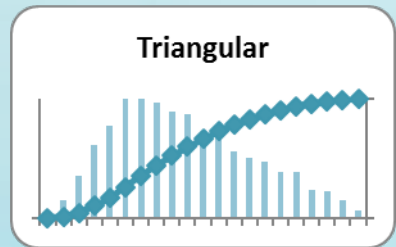


- Normal⁽⁸⁾ – Normal Variation Around a Mean
 - Rolling Multiple Dice

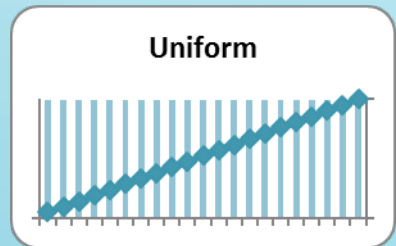


- Log Normal⁽⁶⁾ – Positive Skewed Variation Around a Mean
 - “There is a limit to how well a project can go but no limit to how many problems can occur.”⁽³⁾

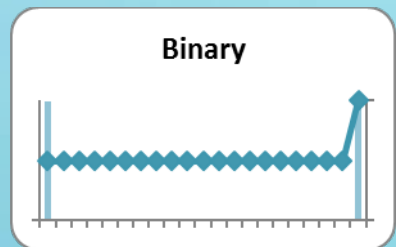
Selecting A Cumulative Distribution Shape



- Triangular⁽⁹⁾ – Known Range, Around a More Likely Value
 - Useful Approximation of Unknown Distribution



- Uniform – A Range High to Low, Equally Likely Values
 - Rolling a Single Dice



- Binary – High Low Value, Only Two Values Possible
 - Useful for Modeling Min, Max Calculations

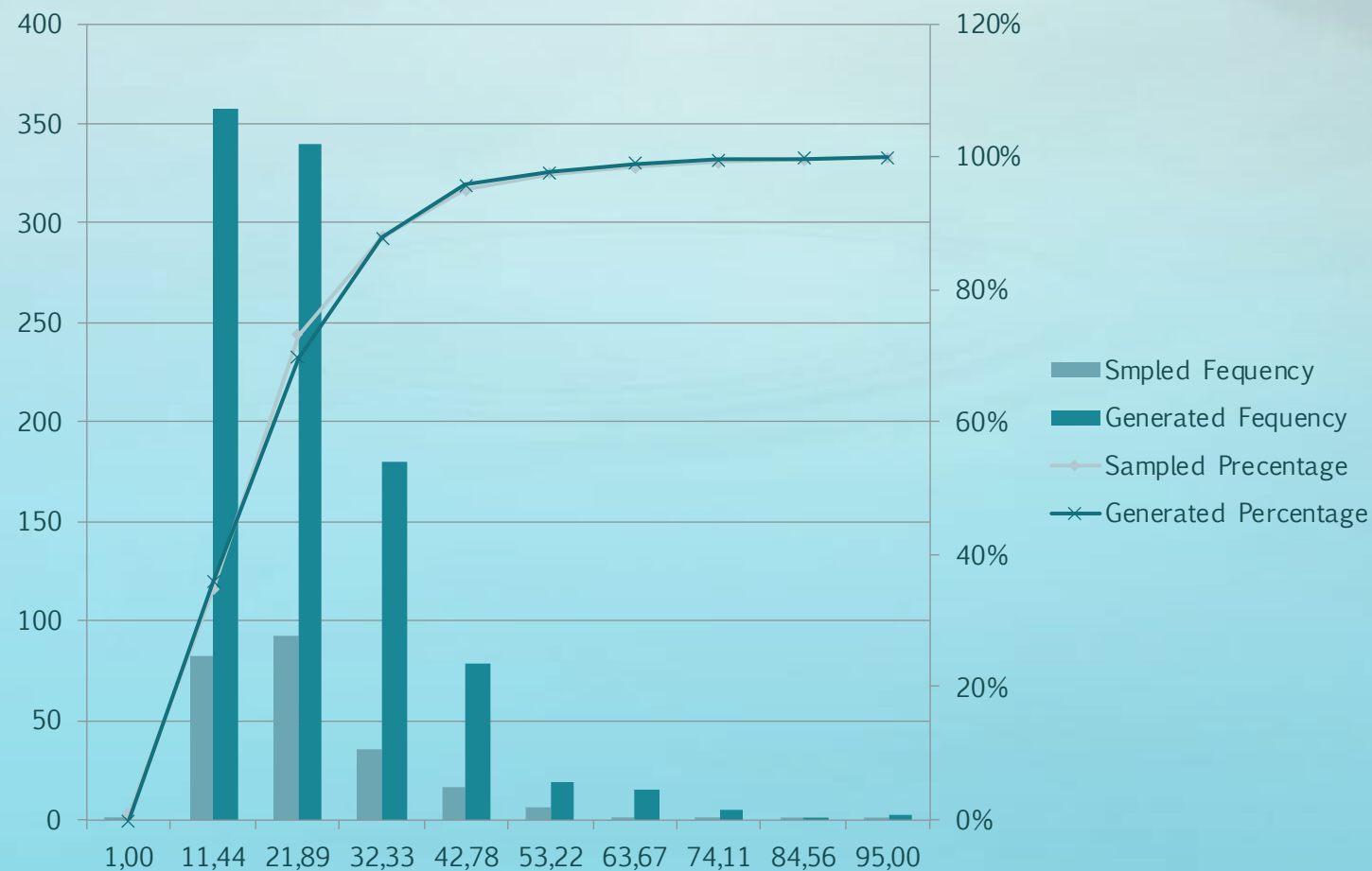
Estimating Feature Duration

- Multiple Years of Historical Data of Actual Cycle Times
- Conscious Efforts Were Made To Keep Size Consistent
- Challenges and Problems Occurred to Extend Cycle Times
- Data Fit Very Closely to a Log Normal Distribution

Feature Duration Cumulative Distribution Function

Distribution Shape:		LogNormal
Min:		1
Max:		95
Mean		18
Standard Deviation:		13
Percentile	10%	6
Percentile	50%	16
Percentile	90%	34

Min:		1
Max:		95
Mean		18
Standard Deviation:		13
Percentile	10%	7
Percentile	50%	15
Percentile	90%	33



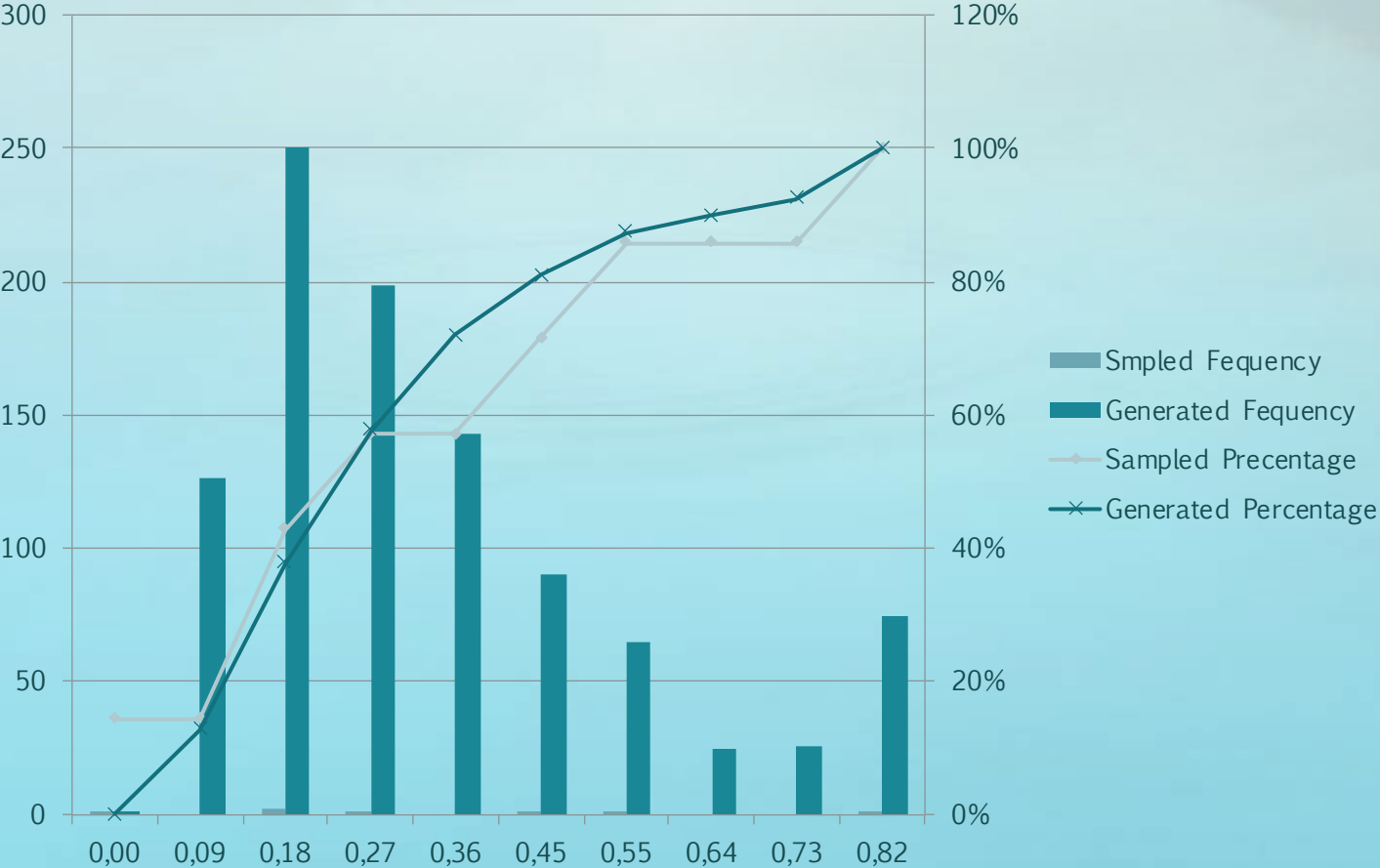
Estimating Feature Growth

- Minimal Historical Data of Initial Estimates to Actuals
- Feature Splitting
- Increased Scope
- Research on Sources of Estimation Error from Steve McConnell's Book *Software Estimation* ⁽¹⁰⁾
 - Error from Initial Concept to Completion from -75% to +300%
 - Fantasy Factor +33%
 - Missed Requirements +20-30%
 - Developer Optimism +20-30%
- Calibrated Estimation from Douglas Hubbard's Book *How To Measure Anything* ⁽¹¹⁾ +20-60%

Feature Growth Cumulative Distribution Function

Distribution Shape:		LogNormal
Min:		0%
Max:		82%
Mean		32%
Standard Deviation:		28%
Percentile	10%	9%
Percentile	50%	22%
Percentile	90%	63%

Min:		1%
Max:		82%
Mean		30%
Standard Deviation:		21%
Percentile	10%	9%
Percentile	50%	24%
Percentile	90%	63%



The background of the slide is a soft, teal-colored image of a misty landscape. It features a calm body of water in the foreground, with rolling hills and mountains in the distance. The scene is shrouded in a light mist or fog, creating a serene and somewhat ethereal atmosphere. The colors are muted, with various shades of blue, green, and grey.

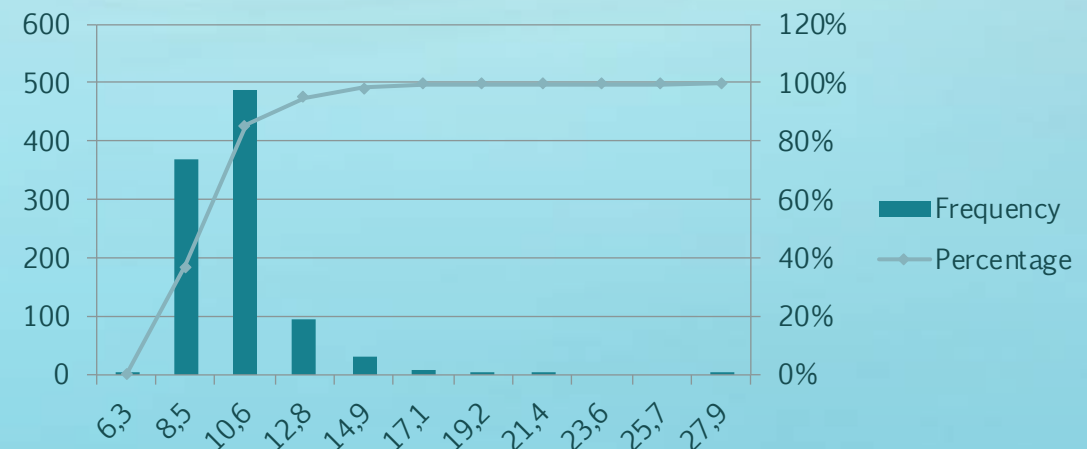
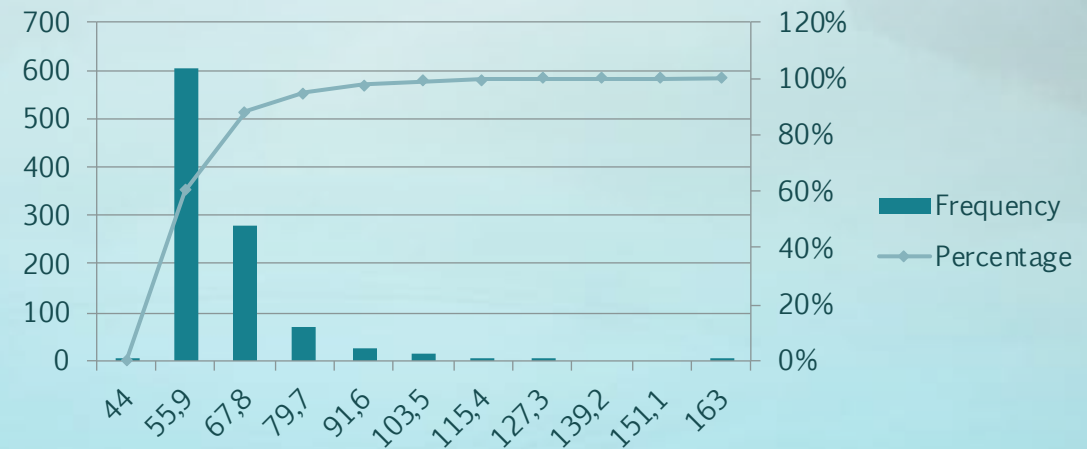
Using the Model Results

“Remember that a model is not the truth. It is a lie to help you get your point across.” ⁽²⁾

Plug In All The Inputs to the Model

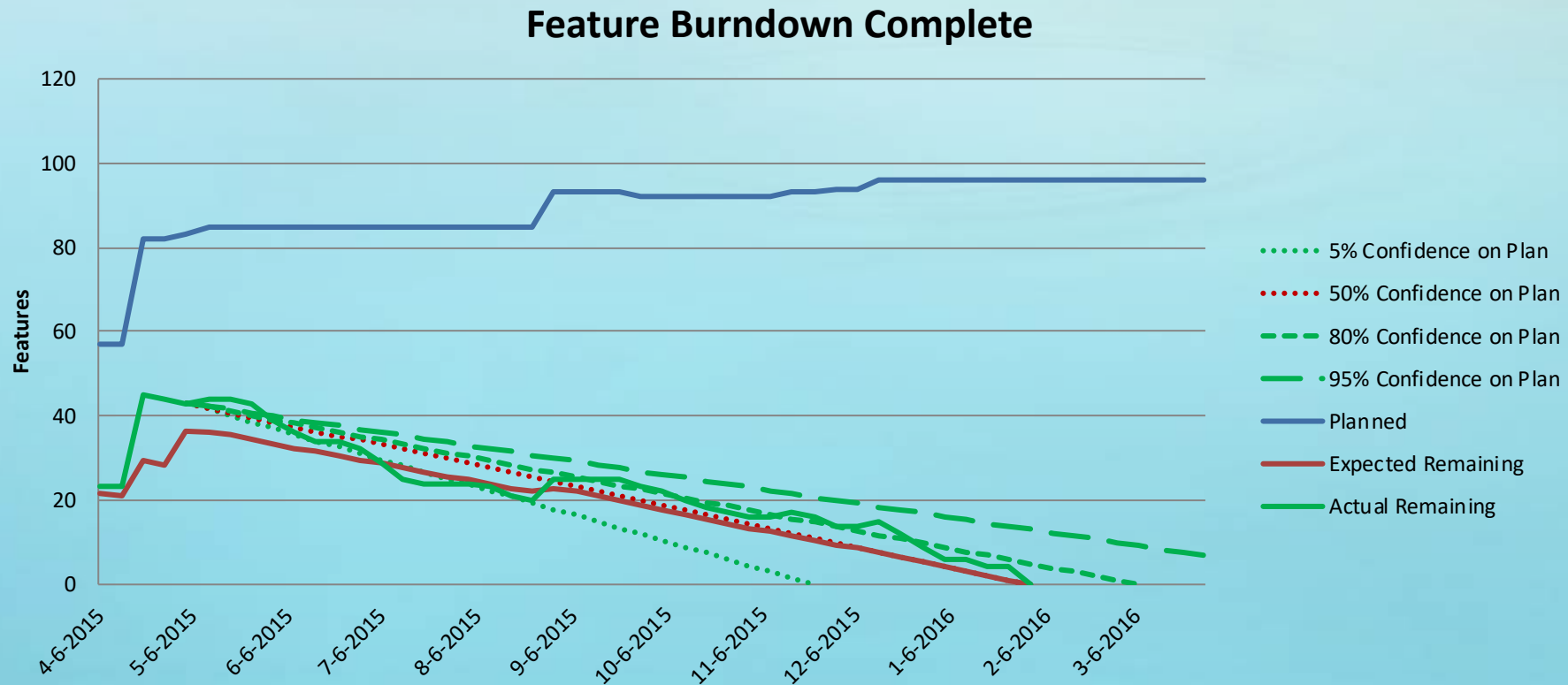
	Min	Mean	Max	StDev	Dist
# of Features		43			Constant
# of Features per Month		6			Constant
Feature Growth		32%		28%	LogNormal
Feature Duration		0.87		0.63	LogNormal

	Features	Months
Min	46.00	6.28
Max	103.00	15.48
Mean	57.66	9.37
Standard Deviation	11.68	1.86
Percentile 10%	47.00	7.55
Percentile 50%	53.00	8.72
Percentile 90%	72.00	11.64



Improve Communication

- Don't Show Single Value Results
- Burn Down Charts Can Show Range of Expectations



Improve System Control

- Several Factors In Our Influence The Calculation
 - Number of Features
 - Number of Features Completed per Month
 - Growth in Features
 - Duration of Features
- All Of These Provide Logical System Controls
- Use The Cumulative Distribution To Select Control Points
- Monitor Results and do Casual Analysis and Resolution on Outliers

The background of the image is a soft, blue-toned landscape. It features a calm body of water in the foreground, reflecting the light. In the distance, there are rolling hills or mountains, some of which are shrouded in a light mist or fog. The overall color palette is a range of blues, from light sky blue to deeper, muted blues and greys, creating a serene and somewhat melancholic atmosphere.

Don't Forget Its Reliable Not
Accurate

A serene landscape featuring a calm body of water, likely a lake or a wide river, which perfectly reflects the surrounding environment. On the left, a steep, forested hillside slopes down towards the water. In the distance, more rolling hills are visible under a vast, clear blue sky. The overall atmosphere is peaceful and contemplative.

Where To Next?

Bibliography

1. Managing by Commitments. (2003). Retrieved March 20, 2016, from <https://hbr.org/2003/06/managing-by-commitments>
2. Savage, Sam L. (2012-03-13). The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty (Kindle Location 1229). Wiley. Kindle Edition.
3. McConnell, Steve (2006-02-22). Software Estimation: Demystifying the Black Art (Developer Best Practices) (Kindle Location 709). Pearson Education. Kindle Edition.
4. Anderson, David J. (2013-11-12). Kanban (Kindle Locations 970-971). Blue Hole Press Inc. Kindle Edition.
5. McConnell, Steve (2006-02-22). Software Estimation: Demystifying the Black Art (Developer Best Practices) (Kindle Locations 1184-1186). Pearson Education. Kindle Edition.
6. How to create a random number following a lognormal distribution in excel? (2014). Retrieved April 10, 2016, from <http://stackoverflow.com/questions/23699738/how-to-create-a-random-number-following-a-lognormal-distribution-in-excel>
7. Joel on Software. (2007). Retrieved April 10, 2016, from <http://www.joelonsoftware.com/items/2007/10/26.html>

Bibliography

8. Produce random numbers with specific distribution with Excel. (2011). Retrieved April 11, 2016, from <http://stackoverflow.com/questions/6241784/produce-random-numbers-with-specific-distribution-with-excel>
9. VBA Create a Random Number from a Triangular Distribution. (2013). Retrieved April 10, 2016, from http://answers.microsoft.com/en-us/office/forum/office_2003-excel/vba-create-a-random-number-from-a-triangular/cc8b77fc-6cf6-4c79-a2c3-6bd8a61db7a0
10. McConnell, Steve (2006-02-22). Software Estimation: Demystifying the Black Art (Developer Best Practices) (Kindle Location 1183). Pearson Education. Kindle Edition.
11. Hubbard, Douglas W. (2014-02-24). How to Measure Anything: Finding the Value of Intangibles in Business (Kindle Location 2500). Wiley. Kindle Edition.



Our Model for Schedule Estimation

“Remember that a model is not the truth. It is a lie to help you get your point across.” ⁽²⁾