
atlanta
GEORGIA
July 25-29, 2016
Agile Alliance

Forecasting using Data

Introduction to probabilistic forecasting
Using data rather than estimates

Every spreadsheet and exercise worksheet is here:

Bit.ly/SimResources (gitHub)

or **FocusedObjective.com** (see “free stuff”)

or **@t_magennis** (I’ve post links here in my twitter feed)

Or email me: **troy.magennis@focusedobjective.com**

Every spreadsheet and exercise worksheet is here:

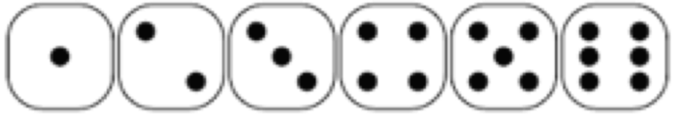
Bit.ly/SimResources (gitHub)

or **FocusedObjective.com** (free stuff)

or **@t_magennis** (I’ve post links here)

Understanding probability - Exercises

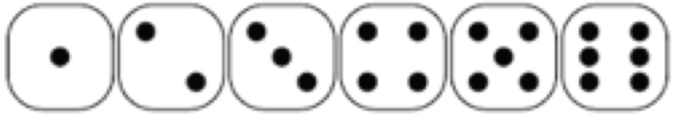
Q1. How many different possible values are there for a standard six-sided dice?



A:

Q2. How many values of a six sided dice are less than 4?

Tip: Circle the values that are less than 4.



A:

Q3. What is the probability of rolling a value less than 4 on a standard six side dice?

Tip: Count the number of "right" values and divide by the total number.

$p = \frac{Number\ of\ "right"\ values}{Total\ possible\ values}$ A:

$$p = \frac{Number\ of\ "right"\ values}{Total\ possible\ values}$$

$$p = \frac{Number\ of\ "right"\ values}{6}$$

$$p = \frac{3}{6} \qquad p = \frac{1}{2} \qquad p = 0.5$$

Sampling

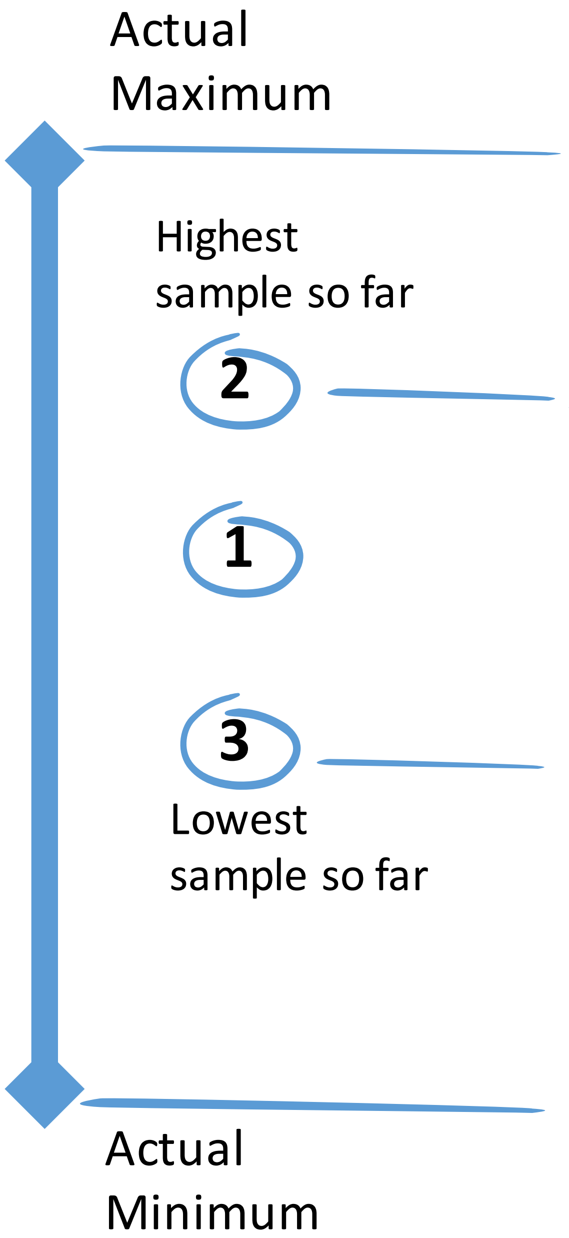
A way to use the data we do have to make predictions & forecasts

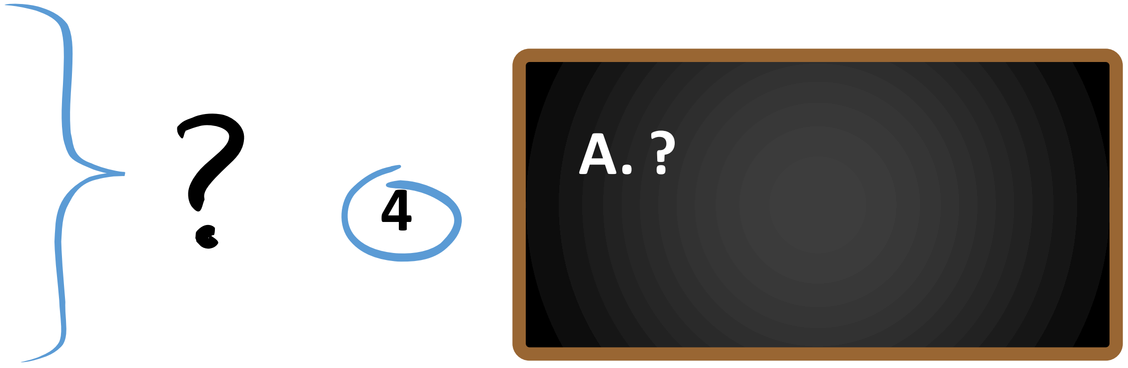
Q. How quickly do we discover a range of values by sampling?

Why? Because as we get story count, story size, velocity, Throughput, cycle-time. How confident should we be of having found the full range values.

Q. On average, what is the chance of the 4th sample being between the range seen after 3 random samples?

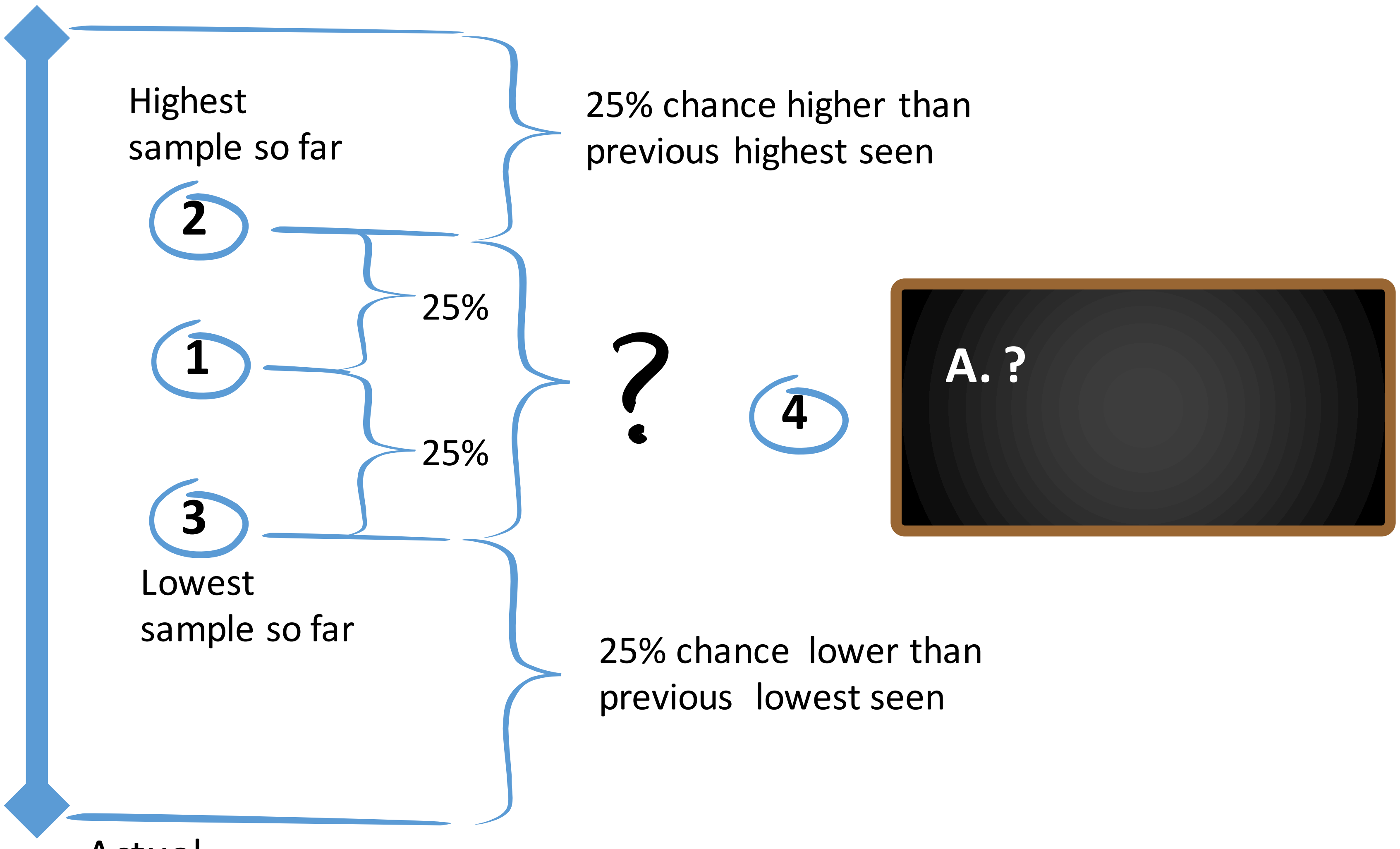
(no duplicates, uniform distribution)





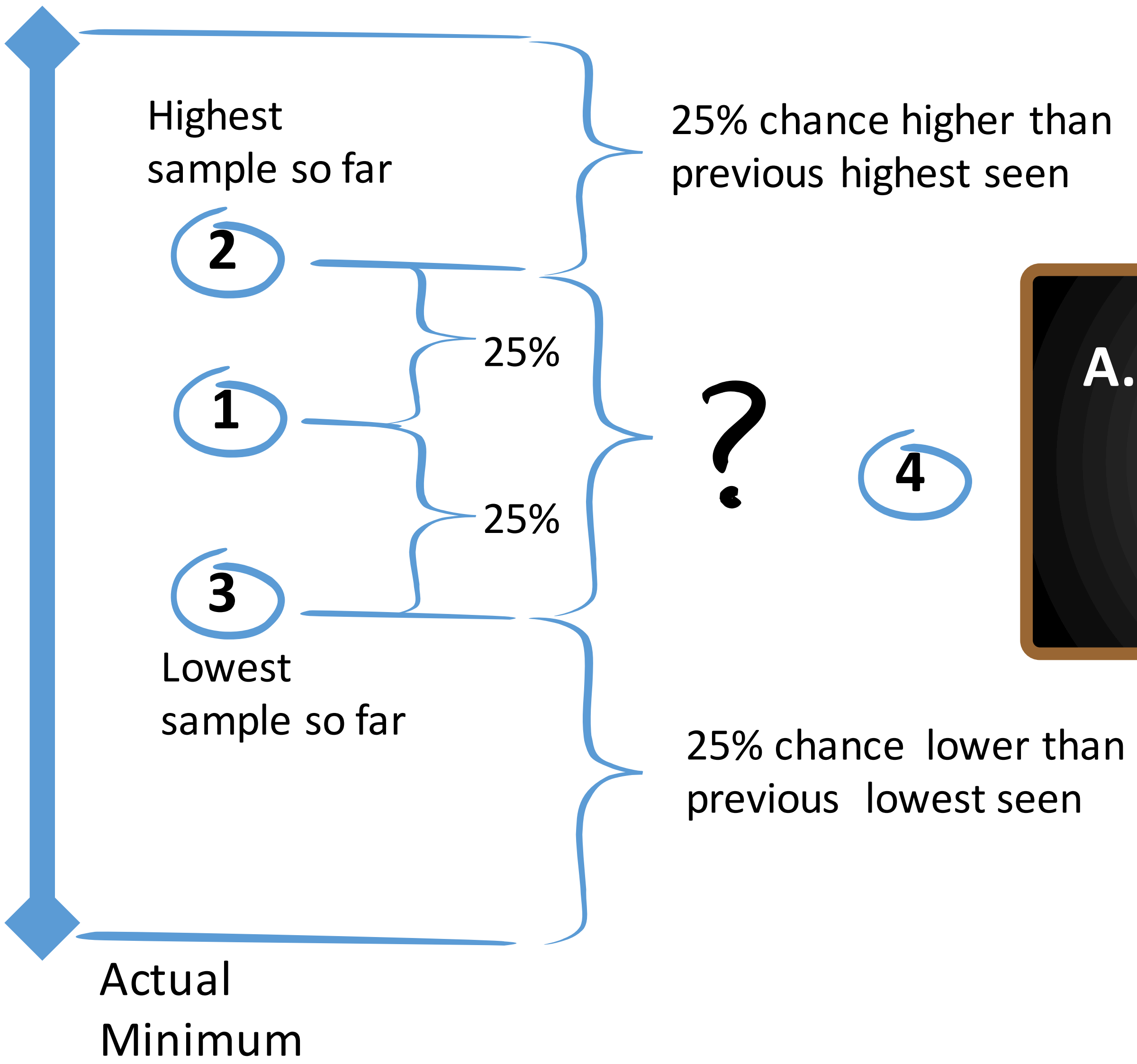
Q. On average, what is the chance of the 4th sample being between the range seen after 3 random samples?

(no duplicates, uniform distribution)



Q. On average, what is the chance of the 4th sample being between the range seen after 3 random samples?

(no duplicates, uniform distribution)



A. 50%

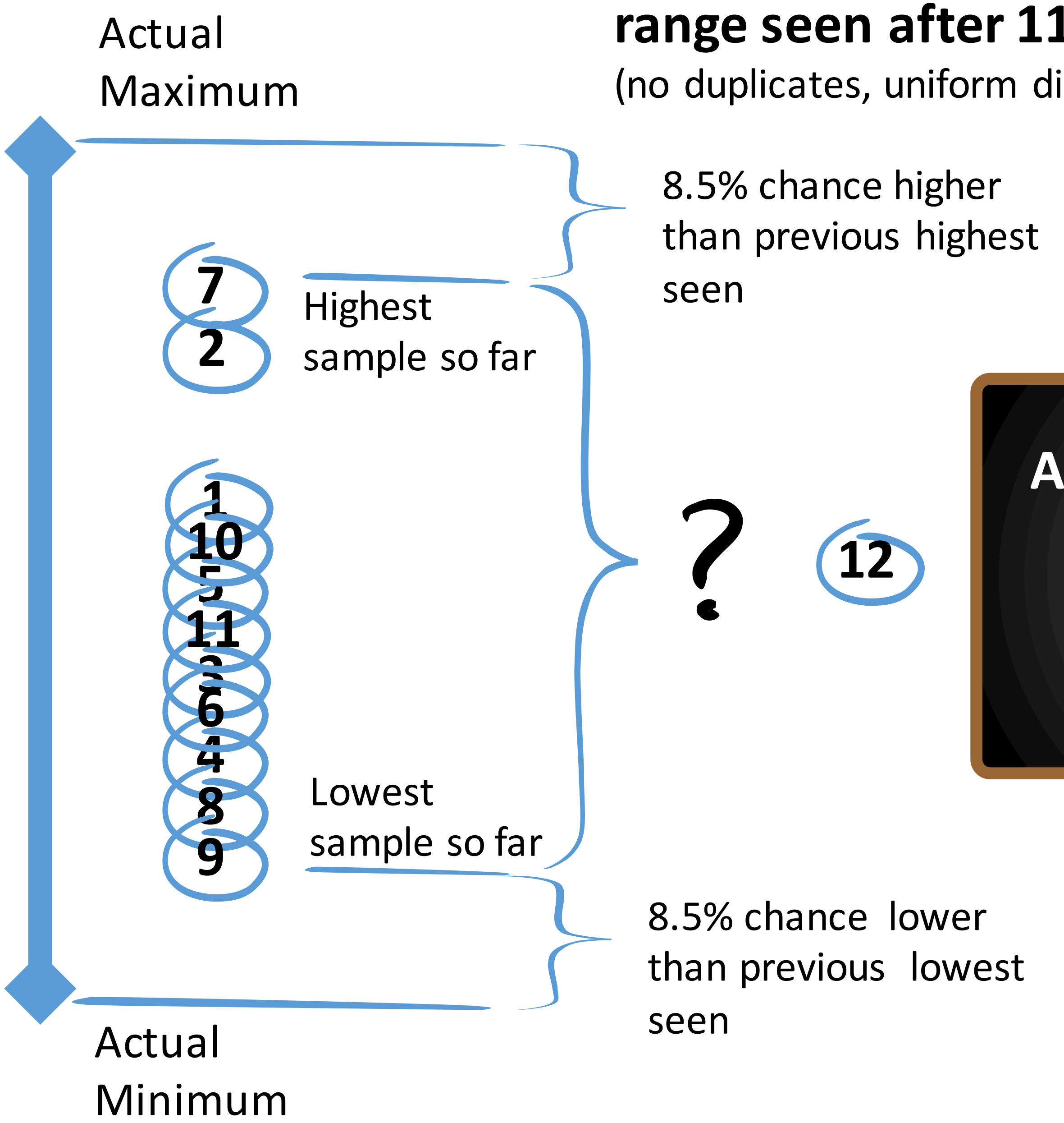
% = (n – 1)/(n+1)

% = (3-1)/(3+1)

% = 0.5

Q. On average, what is the chance of the 12th sample being between the range seen after 11 random samples?

(no duplicates, uniform distribution)



A. 83%

$\% = (n-1)/(n+1)$

$\% = (11-1)/(11+1)$

$\% = 0.833$

Predicted Expected

- “n” = number of prior samples
- A is the % chance next sample in previous range

n	(n-1)/(n+1)	n	(n-1)/(n+1)
2	33%	16	88%
3	50%	17	89%
4	60%	18	89%
5	67%	19	90%
6	71%	20	90%
7	75%	21	91%
8	78%	22	91%
9	80%	23	92%
10	82%	24	92%
11	83%	25	92%
12	85%	26	93%
13	86%	27	93%
14	87%	28	93%
15	88%	29	93%
		30	94%

Experiment

From a *known* range of values, take samples at random and see how fast we can determine what the full range *might* be.

Compare two ways –

1. From the computed probability formula
2. By doubling the average (double what you are told)

Prediction Intervals Exercise

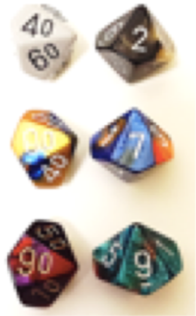
To find how many samples it takes to find the lower and upper bounds of a sample set on average? This exercise simulates finding the upper and lower boundary of a sequential range by sampling the result of dice rolls.

The process

1. **Roll Dice**: Create a random number with a range of 1 to 100. Options:
   a. A random number generator app on your phone (Randomizers)
   b. Use three rolls of a six-sided dice (see next page for chart)
   c. Sum two 10 sided dice (00 – 90 by 10's) and a traditional (0-9)
2. **Repeat**: Repeat 20 times and record the results in the table below.
3. **Examine Results**: Look at the range between the lowest rolled and highest rolled. Compare against expected.

3 x 6 Sided Dice

2 x 10 Sided Dice



42

7

99

Note: Rolling a 00 and 0 = 100

Questions and discussion topics

1. **What probability distribution is a single roll?**
2. **What guarantee do I have that I have found the range expected?**
3. **What happens if the data is a Normal (bell curve) distribution?**
4. **What happens if the data is left or right skewed?**

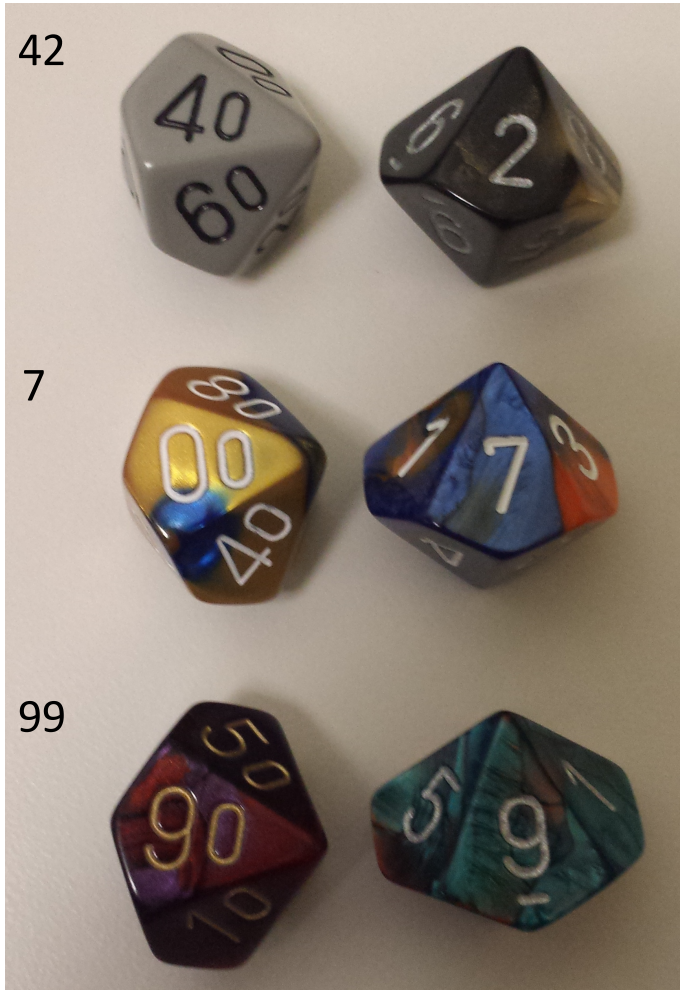
Results table

Record each roll and calculate the ranges seen so far after each roll. Are you ahead or behind expected?

n	This Roll	Lowest So Far	Highest So Far	Range So Far = Highest-Lowest	Expected Range $\frac{(n-1)}{(n+1)} \times 100$	Average So Far (expected 50)
1					0	
2					33.3	

Exercises

- Dice rolling exercise
  - Roll samples from Dice
  - Values from 0 to 99
  - How many rolls before you see: < 10 AND > 90 values



Percentage Dice

D10 (10 sided)

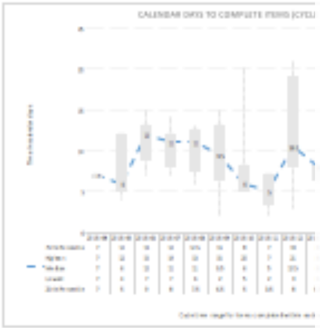
Come to the front when completed. Compare with expected.
How close to 9 samples is range of 80 found? (80% range, 10% above?)

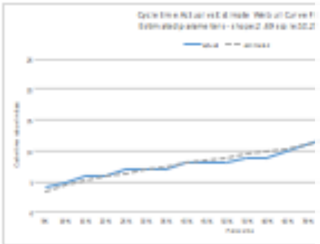
Group	# samples > range > 80	# samples until 2 x avg > 80
1		
2		
3		
4		
5		
6		
7		

Group	# samples > range > 80	# samples until 2 x avg > 80
8		
9		
10		
11		
12		
13		
14		

http://bit.ly/Throughput

Completed Date	Start Date (optional)	Type (optional)
1/21/15	1/14/15	
1/26/15	1/14/15	Story
1/26/15	1/14/15	Defect
1/26/15	1/21/15	Story
1/26/15	1/22/15	Story
1/29/15	1/23/15	Story
2/2/15	1/23/15	Story
2/2/15	1/20/15	Defect
2/2/15	1/20/15	Defect
2/4/15	1/20/15	
2/4/15	1/26/15	
2/4/15	1/23/15	
2/4/15	1/22/15	

17 charts so far…

Throughput (planned & un-planned)

Throughput Histogram(s)

Cycle Time (planned & un-planed)

Cycle Time Histogram(s)

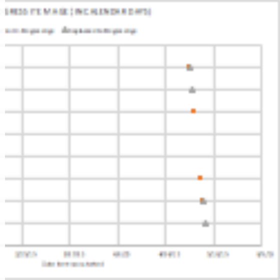
Work In Process

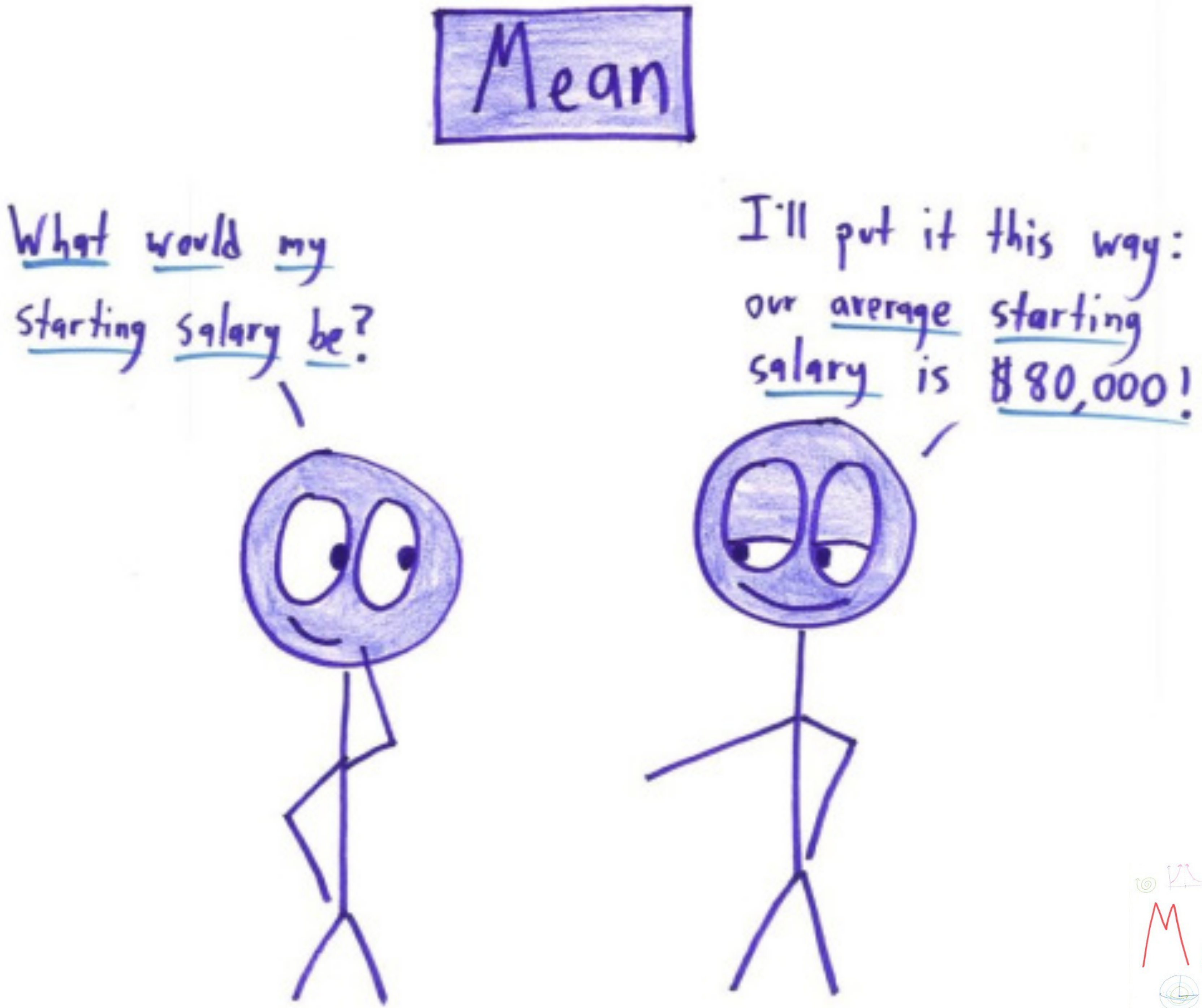
Cumulative Flow

Arrival vs Departure Rate

Un-planned work Percentage

Cycle Time Distribution Fitting





Mean
What would my starting salary be?
I'll put it this way: our average starting salary is $80,000!

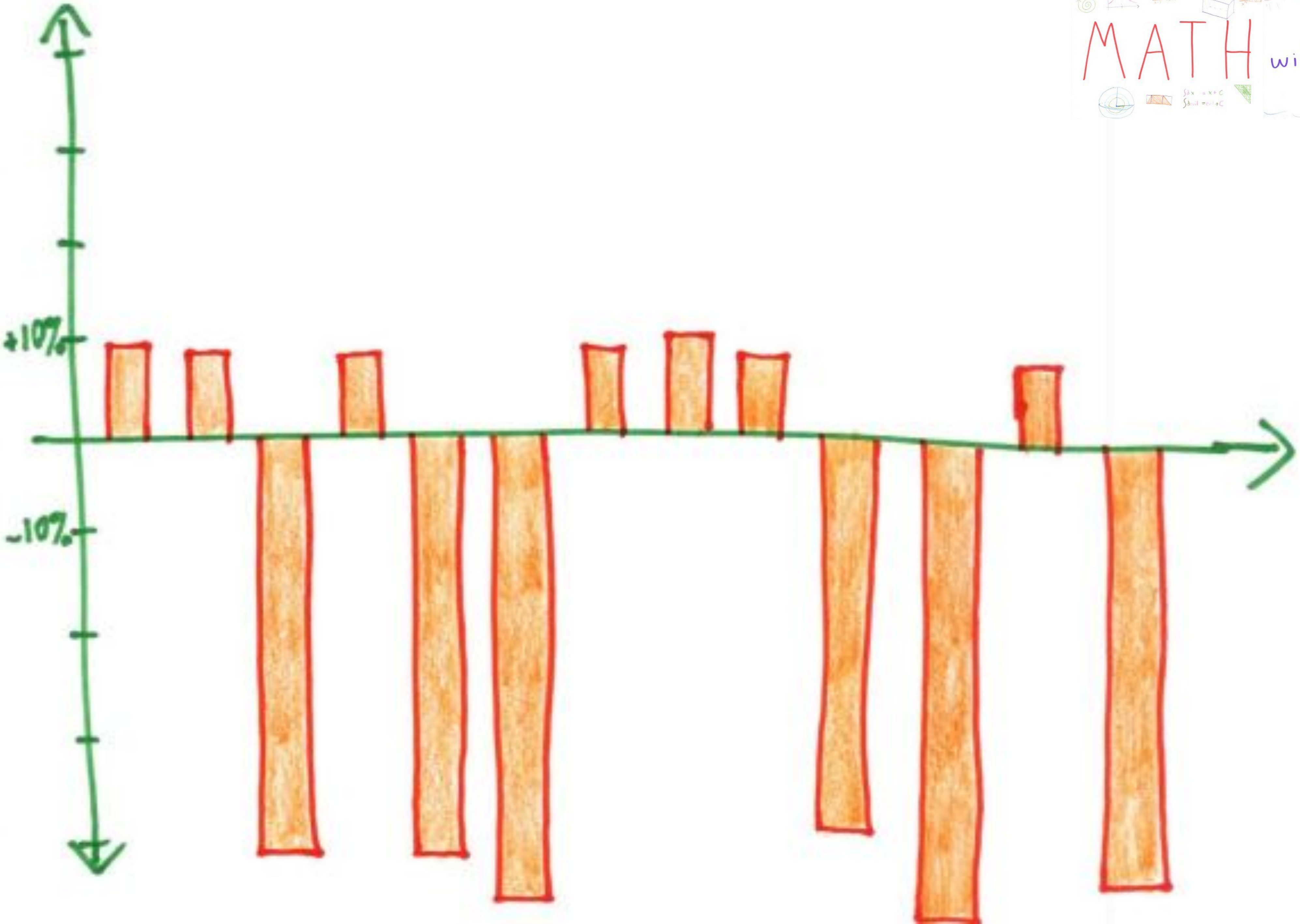
you → $30,000

all your coworkers
- $30,000
- $30,000
- $30,000
- $30,000
- $30,000
- $30,000

CEO's son → $430,000

Average: $80,000.


Median
So, why should I invest with you?
Well, not to brag, but my fund has a median gain of 8% per year!



MATH with BAD DRAWINGS

On average (or median), Arithmetic fails….

1 to 6 days + 1 to 6 + 1 to 6 + 1 to 6 + 1 to 6
= 5 to 30 days

3.5 days + 3.5 + 3.5 + 3.5 + 3.5 = 17.5 days

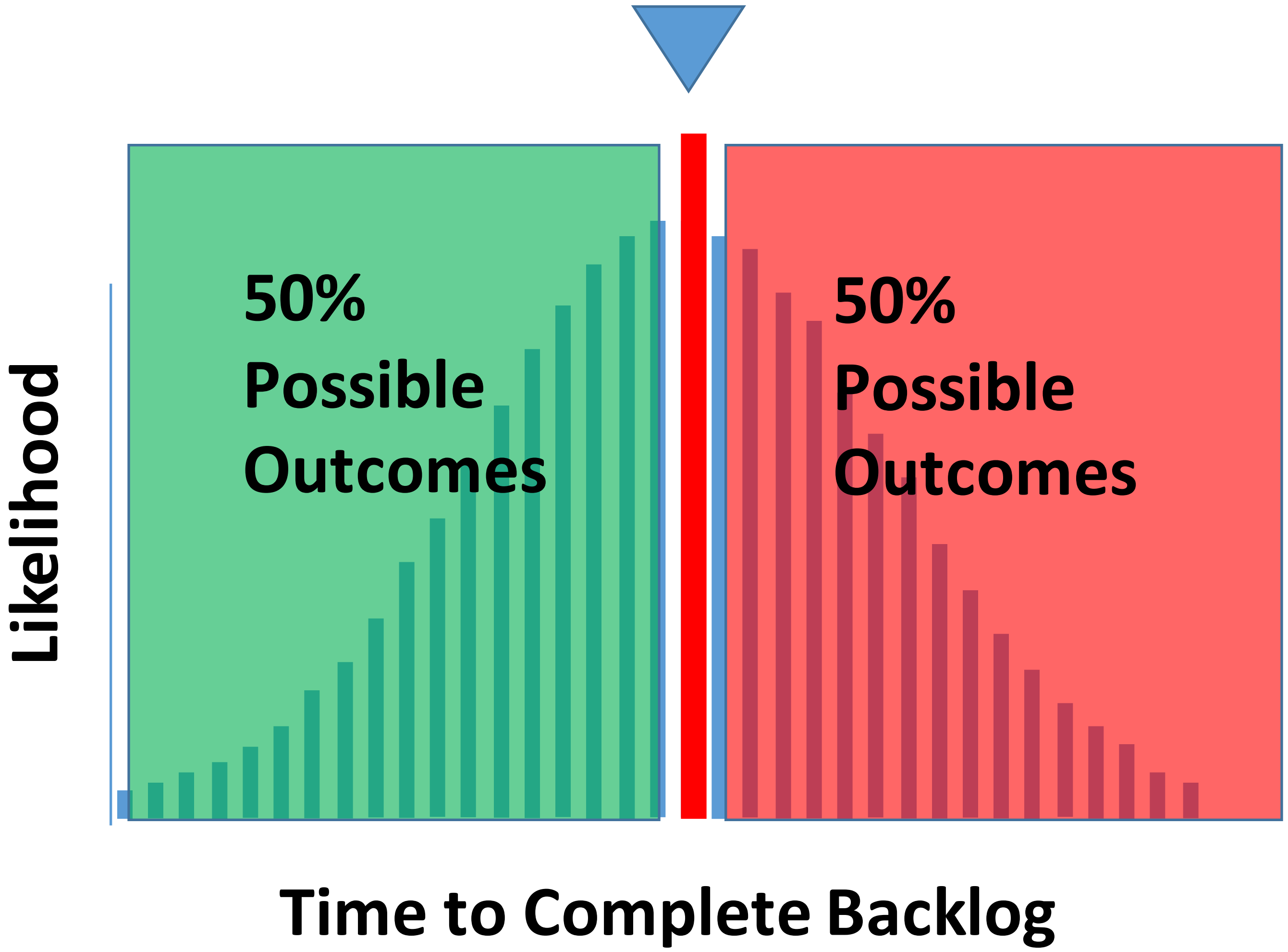
Siri, Add 1 to 6 five times.

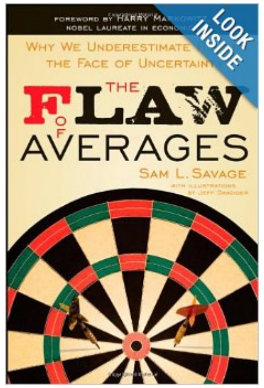
Cortana, Add 1 to 6 five times.

(sometime later)

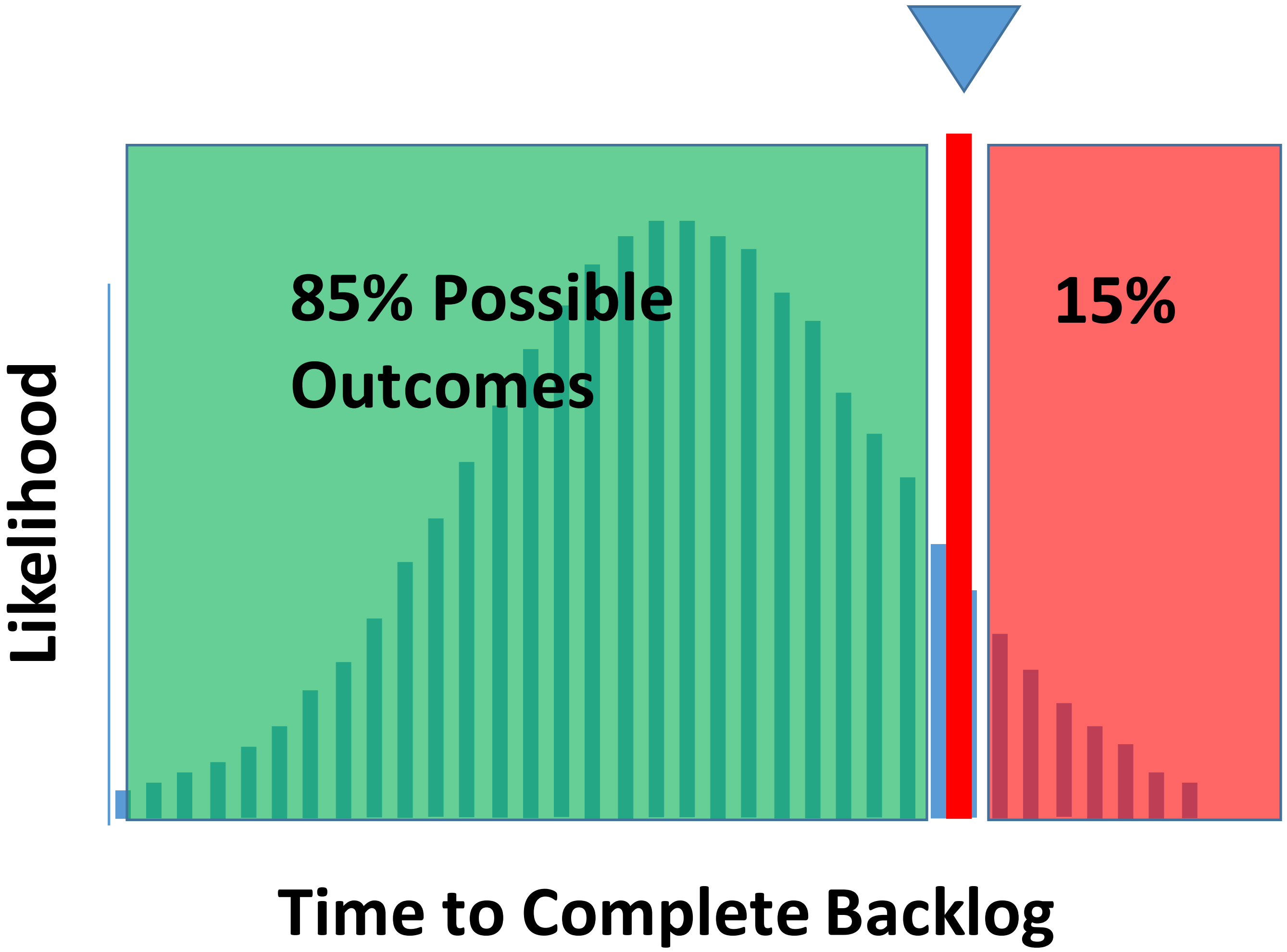
Alexa, Buy me some Vodka….

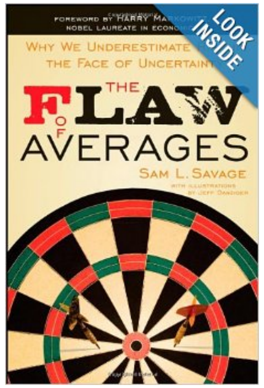
Probabilistic Forecasting combines many uncertain inputs to find many possible outcomes, and what outcomes are more likely than others



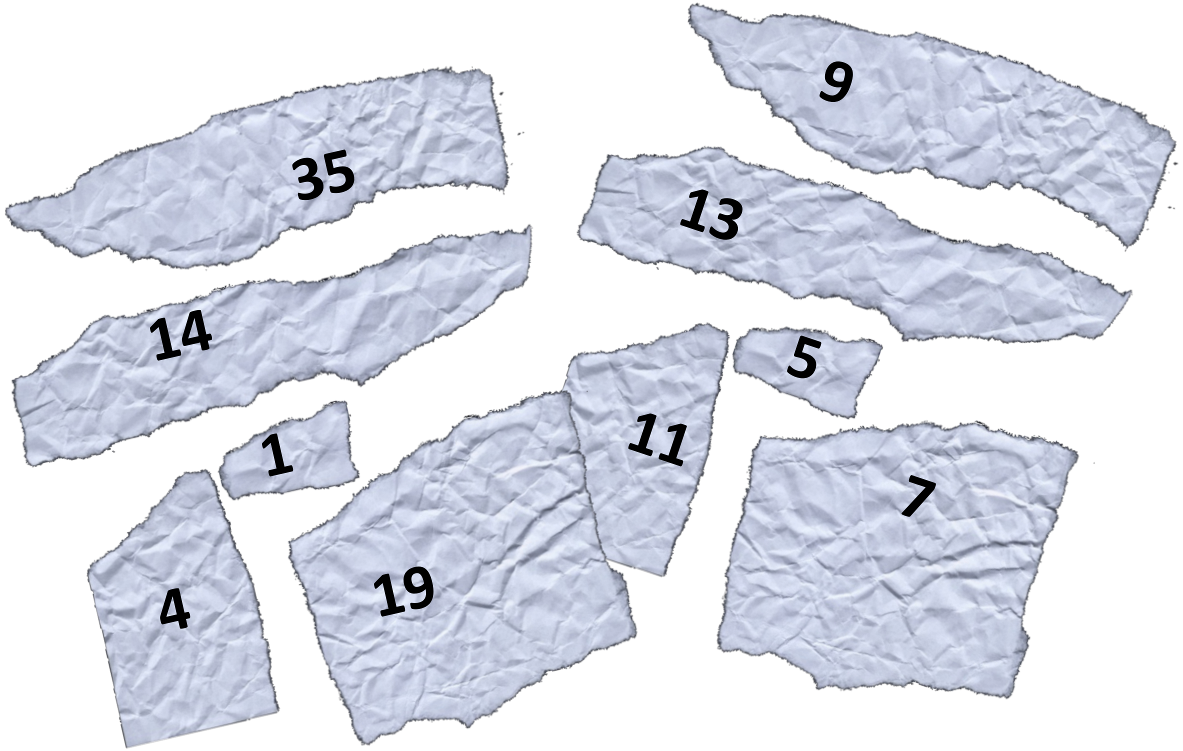


Seeing “How Likely”





Sampling with replacement



	Trial 1	Trial 2		Trial 100
		1		35
		4		19
		7		5
		5		13
		11		11
Sum:	**51**	**28**	…	**83**

Q. Could I make a simple forecast tool that worked?

Without macros or add-ins!

http://bit.ly/ThroughputForecast

http://bit.ly/ThroughputForecast

Forecast Completion Date

1. Start Date 4/1/15

2. How many stories are remaining to be completed?

(enter the range estimate of stories. Tip: start wide and narrow as certainty increases)

Low guess	20	Highest guess	30

3. Stories are often split before and whilst being worked on. Estimate the split rate low and high bounds.

(often the throughput in the backlog is pre-split, but captured throughput post-split. Adjust for this here)

Low guess	1.00	Highest guess	1.00

4. Throughput. How many completed stories per week or sprint do you estimate low and high bounds?

Throughput estimate/samples are per Week 7 days

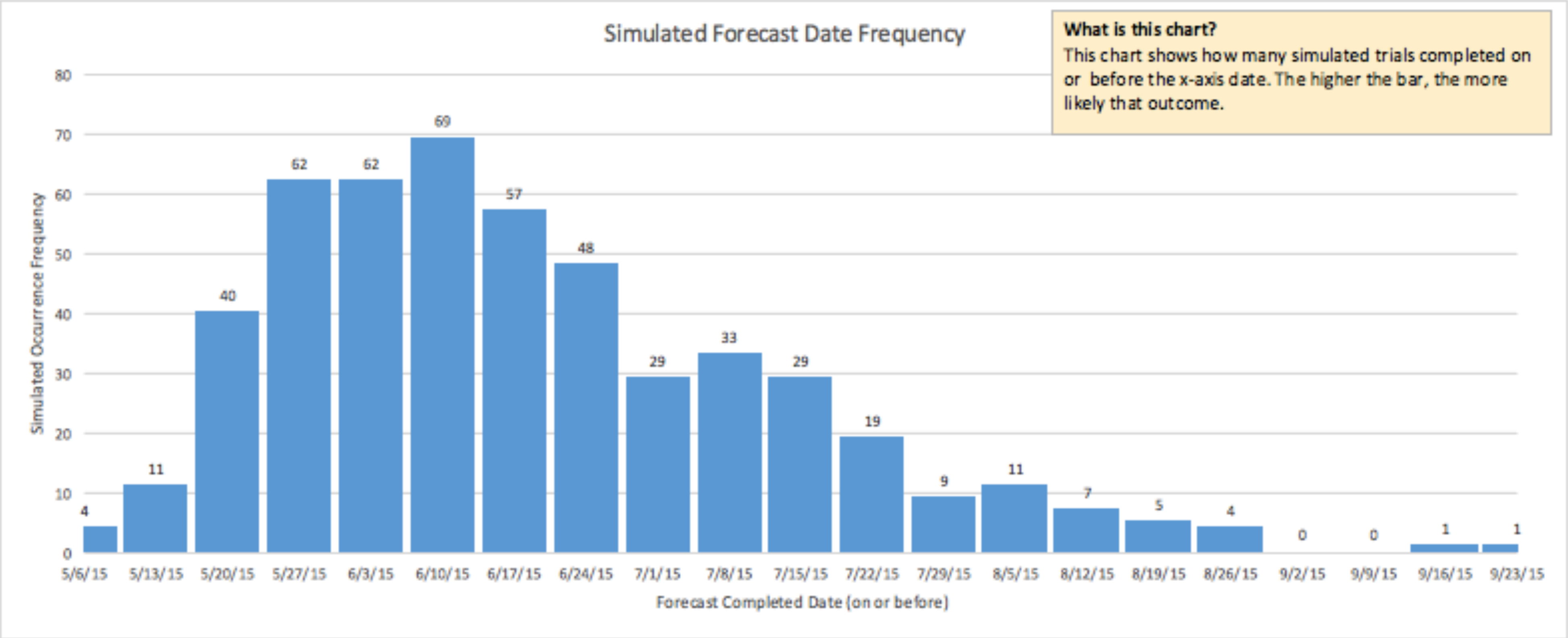
Use historical throughput data **OR** enter a low and high estimate below. Use: Estimate

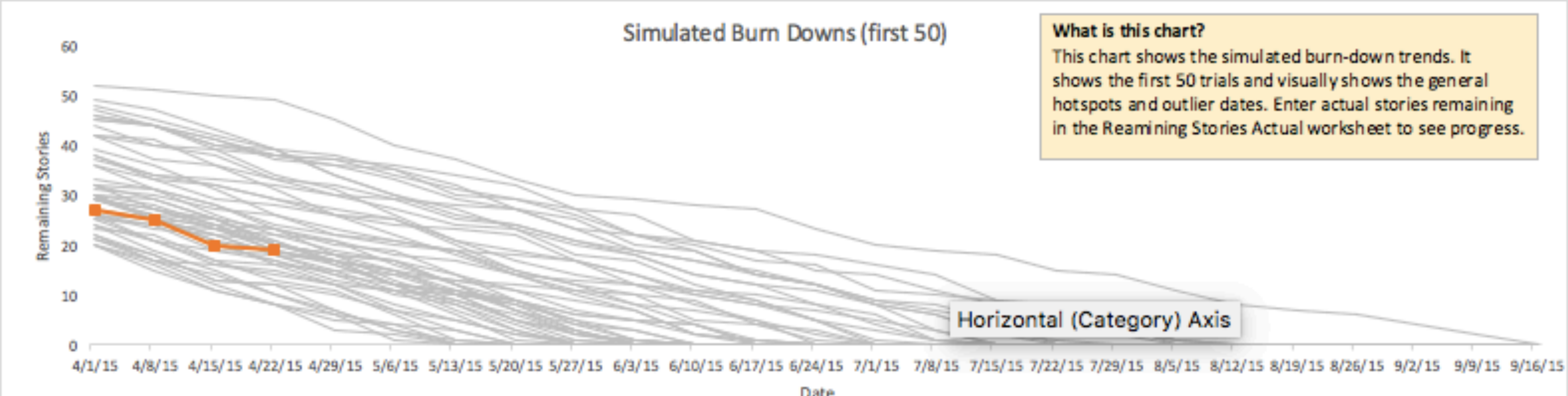
Low guess	1	Highest guess	5

Can I use velocity rather than throughput?

Yes. If you do have estimates in story points, then you can sume all of the estimates and use that for input 2 and estimate or use historical team velocity for input 4. The benefit of using throughput (count of

http://bit.ly/ThroughputForecast





http://bit.ly/ThroughputForecast

Results

Likelihood	Duration in Week's	Date
100%	25	9/23/15
95%	18	8/5/15
90%	16	7/22/15
85%	15	7/15/15
80%	14	7/8/15
75%	13	7/1/15
70%	12	6/24/15
65%	12	6/24/15
60%	11	6/17/15
55%	11	6/17/15
50%	11	6/17/15
45%	10	6/10/15
40%	10	6/10/15
35%	9	6/3/15
30%	9	6/3/15
25%	9	6/3/15
20%	8	5/27/15
15%	8	5/27/15
10%	7	5/20/15
5%	7	5/20/15

Almost certain

Somewhat certain

Less than coin-toss odds. But if you are game?

Experiment

From a set of *prior* throughput samples, compute the completion rate(s) for the next 6 (six) weeks.

Process –

1. Repetitively sample prior throughput in sets of 6
2. Compute how many trials complete at least 10, 20, 30, 40, 50, 60 items in 6 weeks

24 Throughput (or velocity) Samples Randomly picked by throwing a dice

First dice throw

	⚀	⚁	⚂	⚃	⚄	⚅
⚀	16	3	10	6	19	11
⚁	17	17	15	9	11	8
⚂	5	13	5	7	8	6
⚃	10	10	8	5	5	7
⚄	Roll again	Roll again	Roll again	Roll again	Roll again	Roll again
⚅	Roll again	Roll again	Roll again	Roll again	Roll again	Roll again

Second dice throw

1. Throw a 6-sided dice. Pick the column.
2. Throw a six-sided dice and pick the row
3. If it doesn't say "Roll again" this is your throughput sample.

Fill in the numbers for Trials 1, 2 and 3. I've done Trials 4 to 11 so you don't want to kill me!

Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6
			7	11	7
			19	7	10
			6	5	5
			6	19	5
			5	7	10
			5	7	19

Exercise – Throughput Forecast Monte Carlo Worksheet

Aim: To estimate the number of stories that will be completed by a team for a six (6) week timespan using historical weekly throughput samples for that team. To understand the probability of achieving those estimates.

Process:

1. Shuffle the 24 throughput cards or dice (whichever method you choose)
2. Pick a card at random or throw dice and record sample in the table below
3. Return the card to the deck and reshuffle ("sample with replacement")
4. Repeat until all squares are filled

We randomly sampled trials 4 to 11 for you to save time.

Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Trial 9	Trial 10	Trial 11
			7	11	7	5	17	5	10	16
			19	7	10	5	13	13	5	7
			6	5	5	3	5	16	6	5
			6	19	5	3	5	3	6	3
			5	7	10	5	6	8	8	6
			5	7	19	10	16	8	10	16

5. **Sum of all samples for each trial by column (upper) / Nearest "tens" grouping rounded down (lower)**

			48	**56**	**56**	**31**	**62**	**53**	**45**	**53**
			40+	50+	50+	30+	60+	50+	40+	50+

6. **Sum all trials (a):** **Average all trials (a/11):**

   Actual data average 6 week throughput = **57.75**. How close was your average?

7. **Probabilities of achieving at least n stories for a six-week timespan**

Six Week Throughput	Count trial sum groups at least 30,40, 50, etc. stories	(Count / 11) Likelihood
At least **30** stories		
At least **40** stories		

This value is 0 to 1

Multiply it by 100 to get a percentage.
0% = no chance,

Come to the front and give me your Likelihood of 60, 70 and 80 stories

Group	% >= 40 stories	% >= 50 stories	% > 60 stories
1			
2			
3			
4			
5			
6			
7			

Group	% >= 40 stories	% >= 50 stories	% > 60 stories
8			
9			
10			
11			
12			
13			
14			

Every spreadsheet and exercise worksheet is here:

Bit.ly/SimResources (gitHub)

or **FocusedObjective.com** (free stuff)

or **@t_magennis** (I’ve post links here)

Every choice we make changes the outcome