Raytheon
Customer Success Is Our Mission

Schedule Risk Analysis

Building Models and Validating Estimates with Historical Data

Andrew Uhlig

Raytheon Missile Systems

January 27, 2015

Schedule Risk Analysis

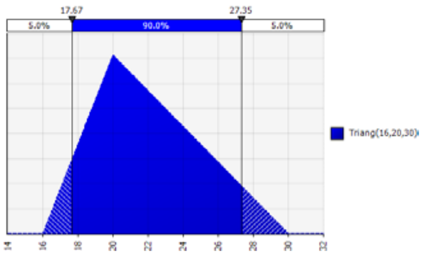
- Statistical analysis performed on a regular basis enhances the use of the IMS as a **management tool**
- Fitting distributions to historical data **more accurately** models project performance and contributes to program management efficiencies
- Focusing on baseline duration variance **is measurable** and demonstrates continuous improvement

Raytheon

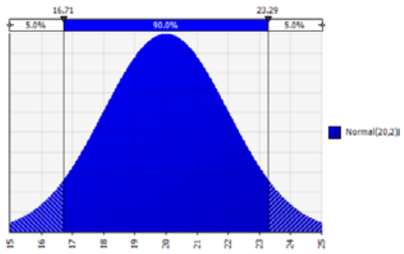
Question for the Audience

- **Do common distribution curves accurately model your data?**

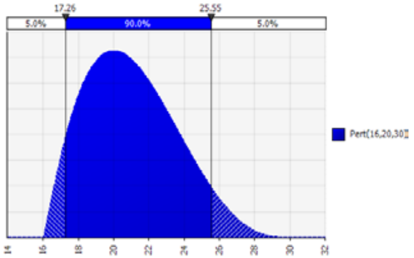
- Some examples of common curves used when conducting Schedule Risk Assessments



Triangle

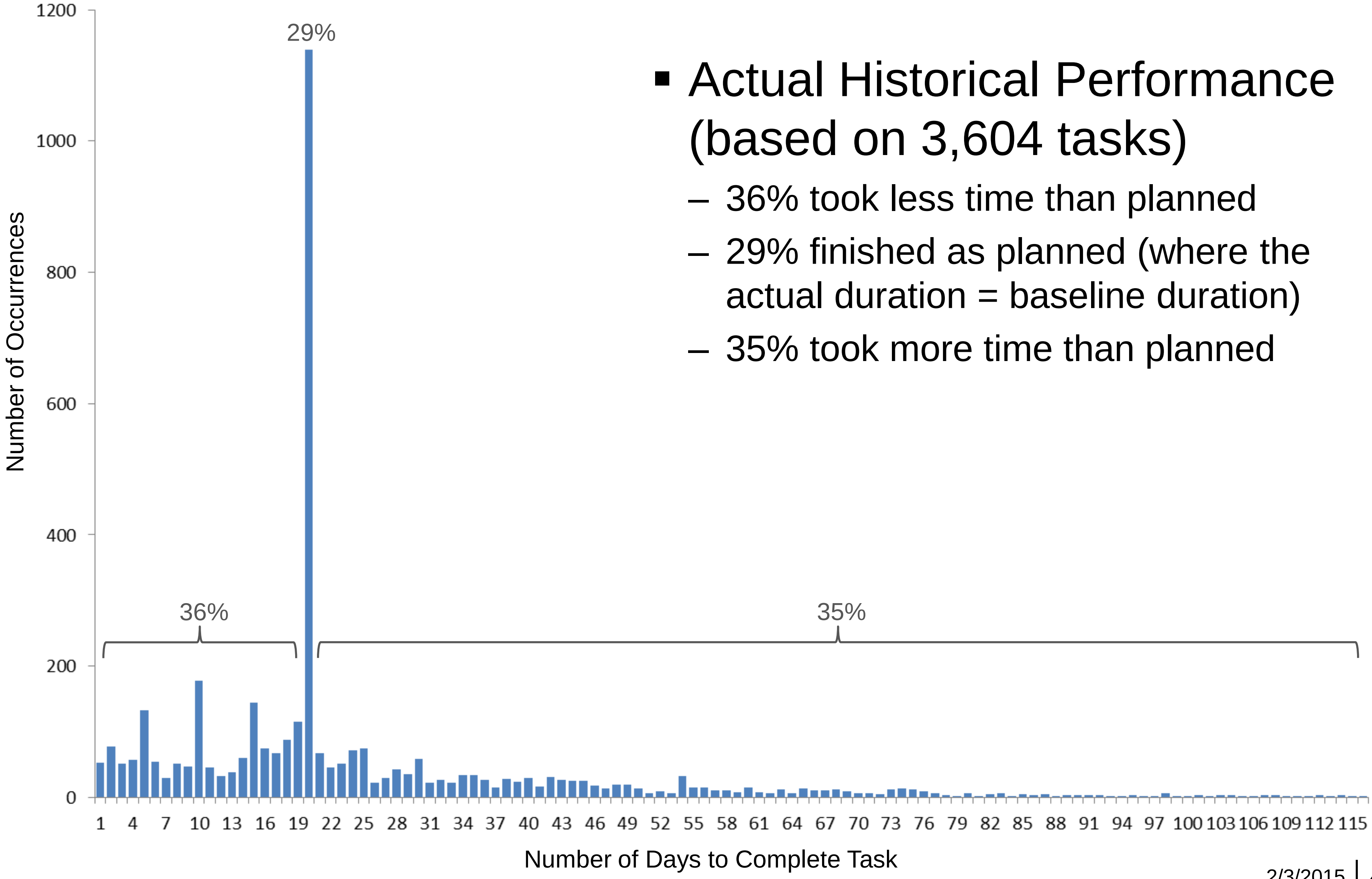


Normal



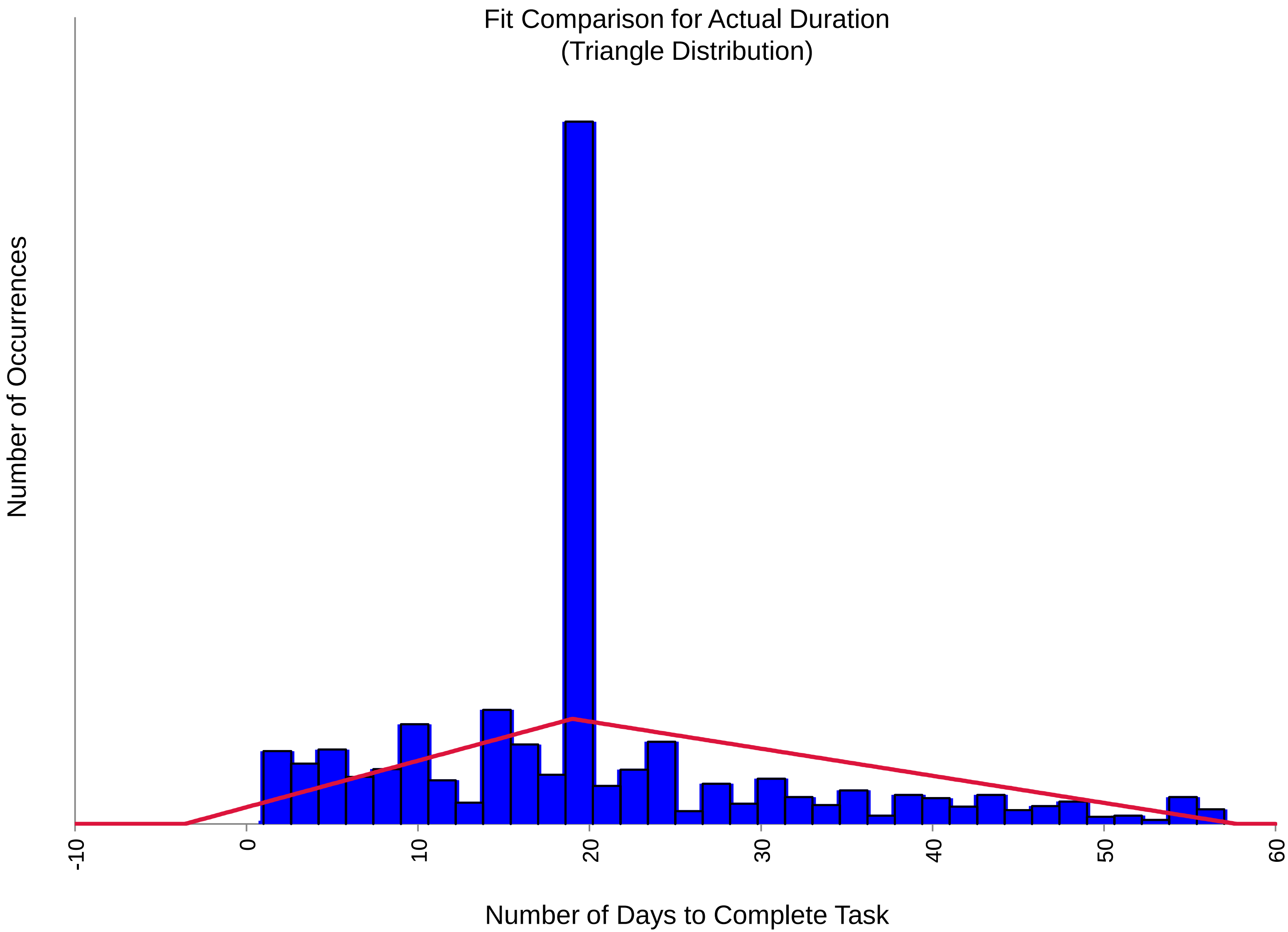
Pert

Sample of Tasks with 20 Day Baseline Durations

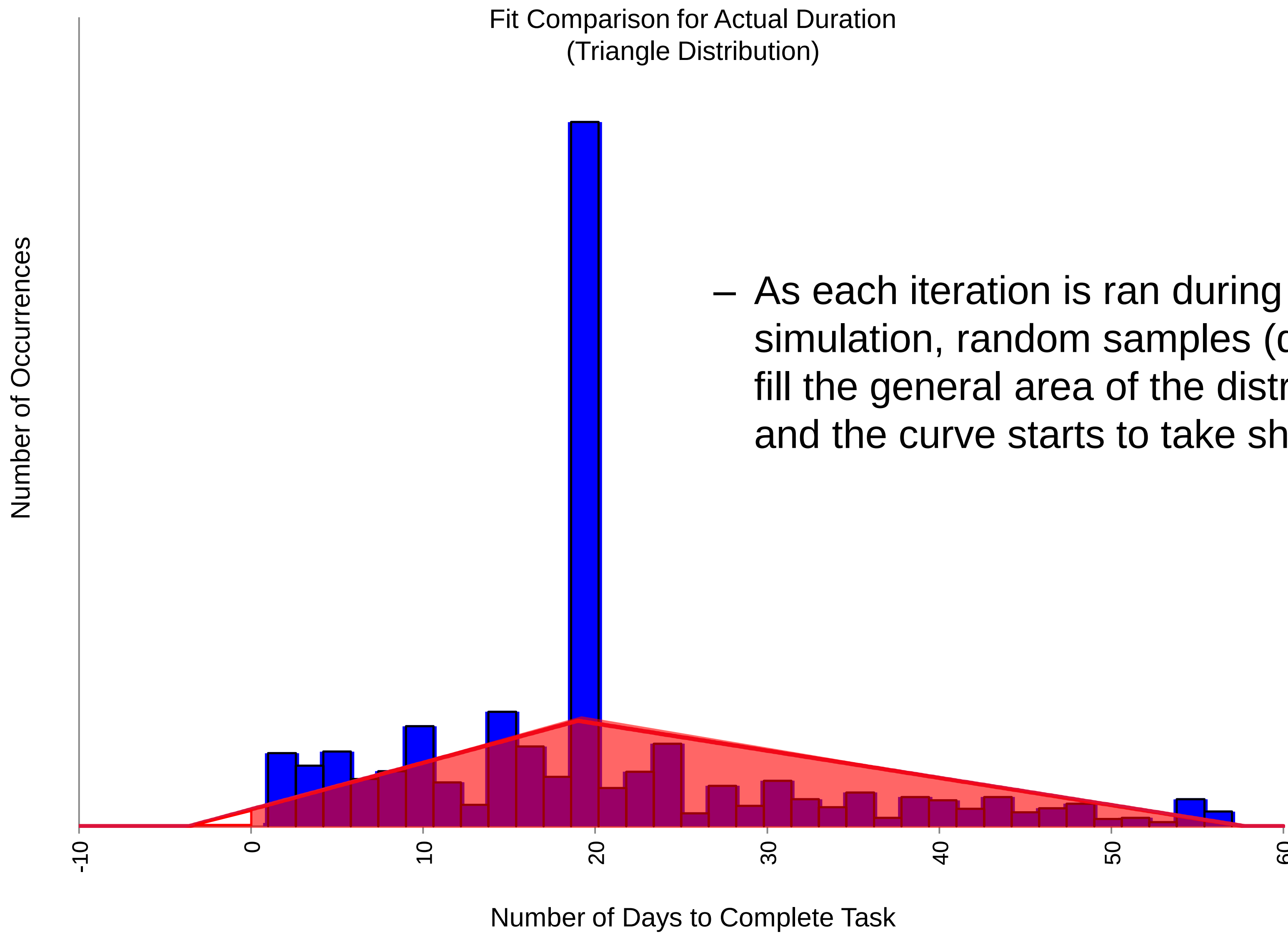


- Actual Historical Performance (based on 3,604 tasks)
  - 36% took less time than planned
  - 29% finished as planned (where the actual duration = baseline duration)
  - 35% took more time than planned

Distribution Fitting Function

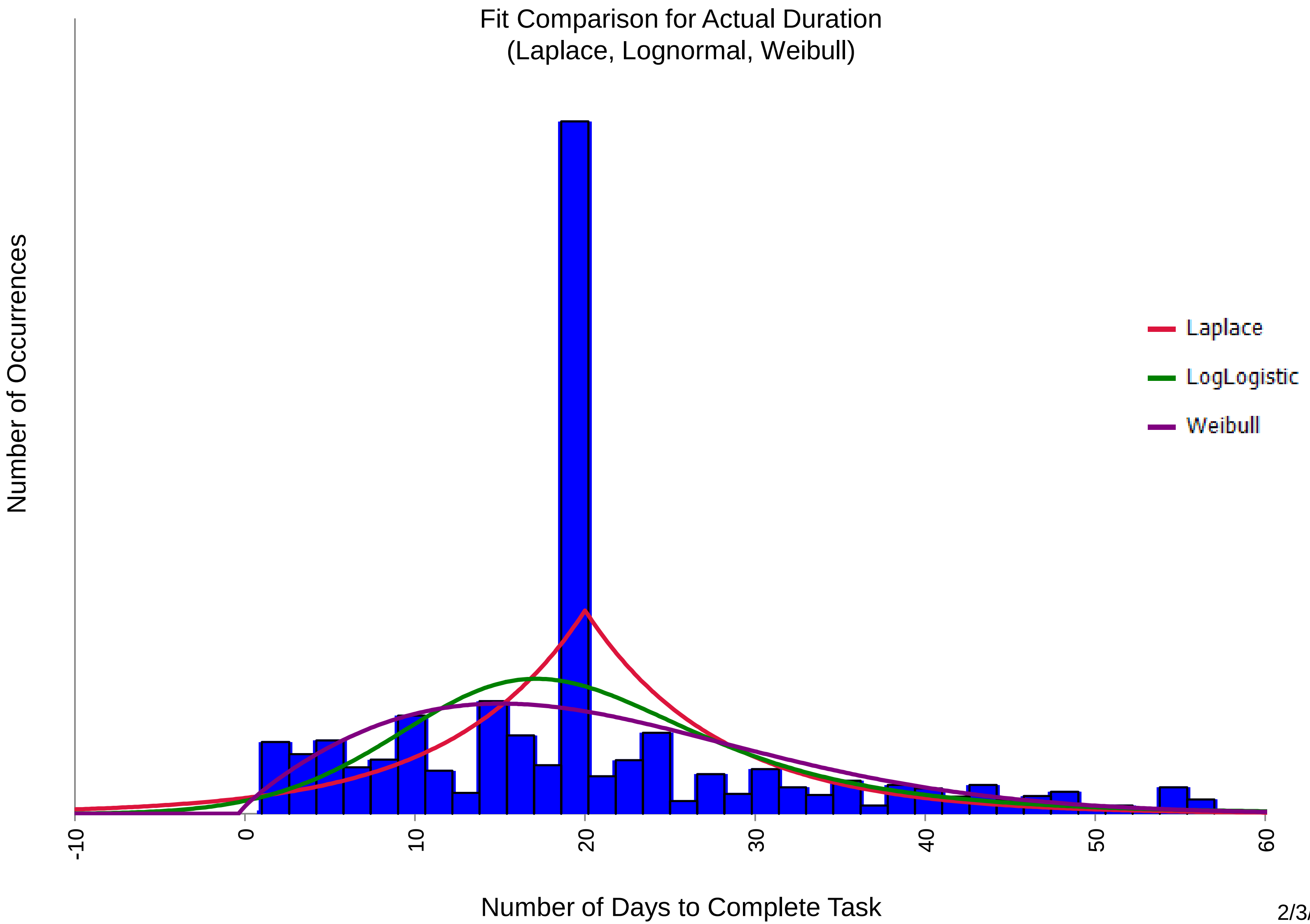


Distribution Fitting Function Cont.



– As each iteration is ran during the simulation, random samples (durations) fill the general area of the distribution and the curve starts to take shape

Fit Ranking – Top Three Distributions



Engineering and Manufacturing Development

- Common distribution curves **do not** model data well in **our** development environment

Custom Distribution Curves

- Build a custom distribution curve to more accurately model the data
  - Assumes there is some historical data available based on actual performance

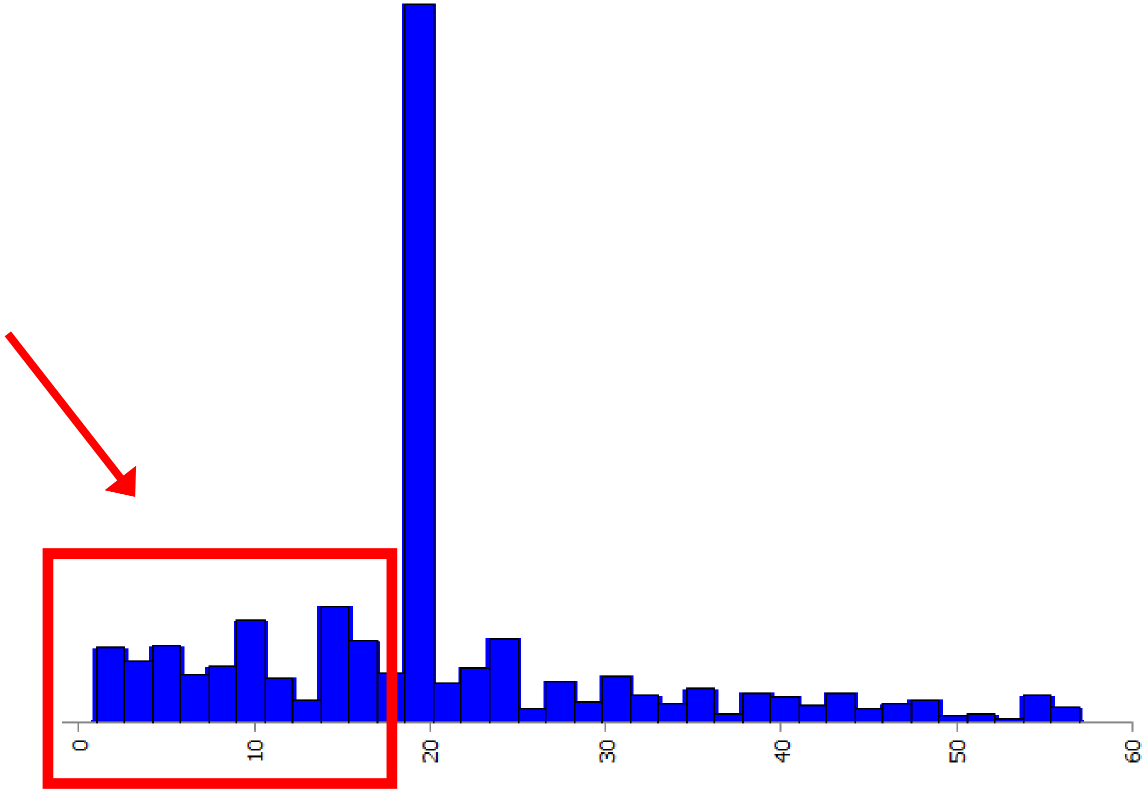
- Some tasks will take less time than planned and some will take more

- We also know that a large number of tasks will finish as planned (where the actual duration = baseline duration)

Think about these as three separate measurable events

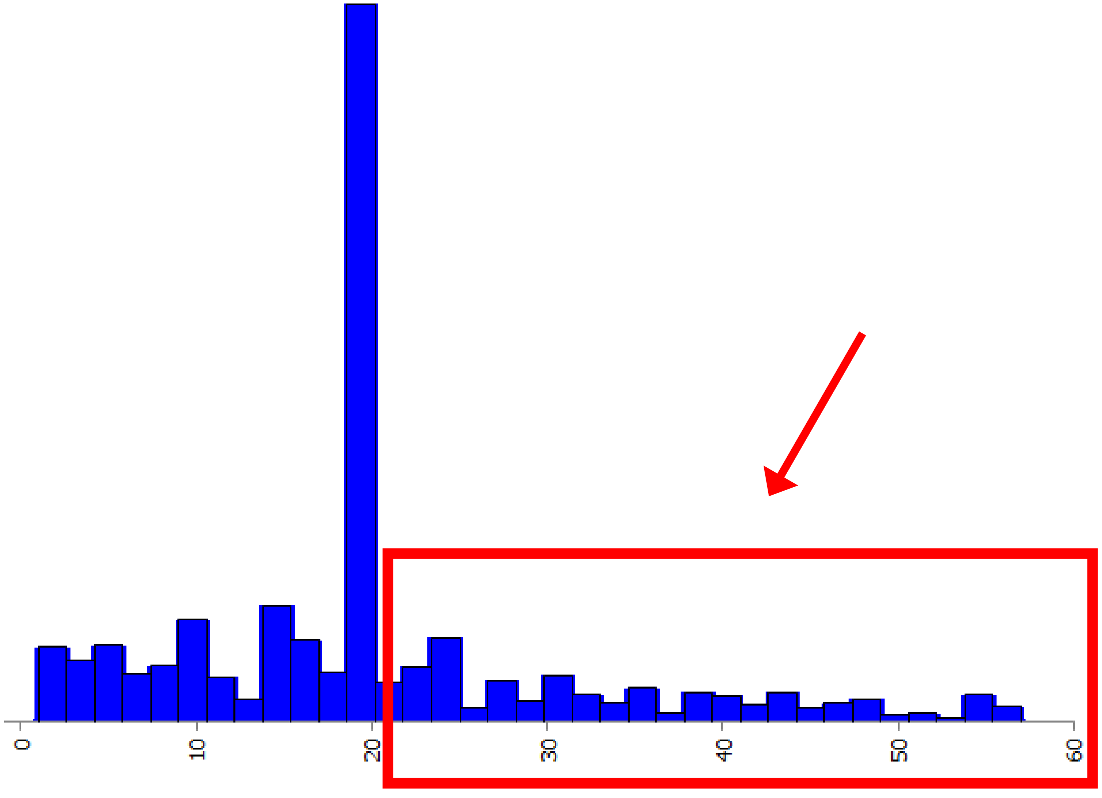
Breaking out the Distribution

- For example, consider the first event all by itself. Run a distribution fit on data points where you perform better than planned
  - This will be “Event *x*”



Breaking out the Distribution Cont.

- Now consider the second event all by itself and fit a distribution on the data points where performance took longer than planned
  - Consider this “Event *y*”



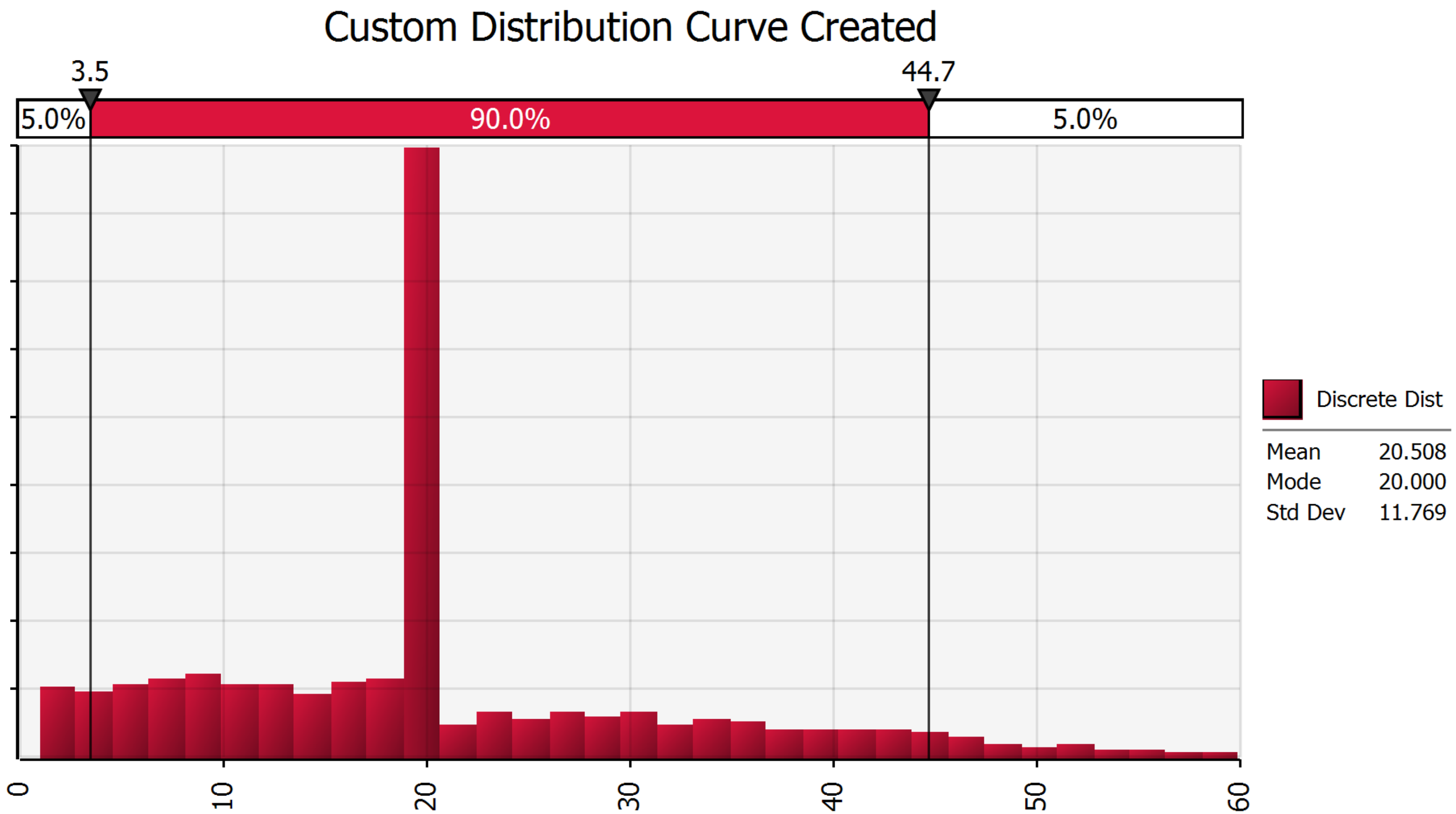
Discrete Distribution

- A discrete distribution can build a more accurate curve
  - Specifies a number of outcomes *n*. In this case there are three
  - Each outcome (event) has a value (it’s own distribution) and a weight which specifies the outcome's probability of occurrence

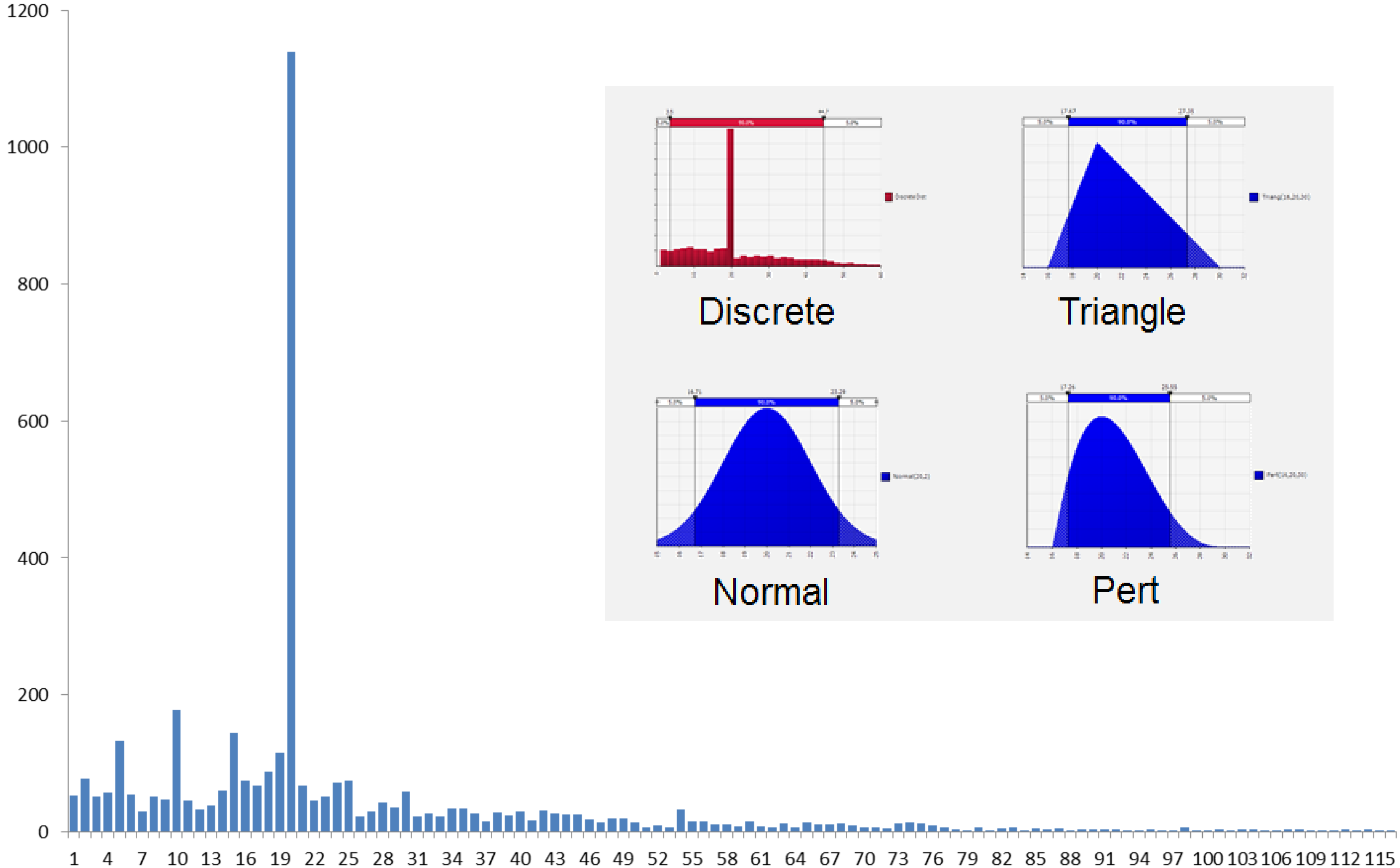
- Event *x* fitted distribution
  - Actual duration < baseline duration
  - Has a 36% Probability of occurrence
- Event *y* fitted distribution
  - Actual duration > baseline duration
  - Has a 35% Probability of occurrence
- Event *z* will have no variation
  - Actual duration = baseline duration
  - Has a 29% Probability of occurrence

Simulation technique used to combine separate events into one overall distribution

Simulation Result



Which Curve Most Accurately Models the Data?



Two Approaches to Collecting Data

- The “Art” aspect
  - Tribal knowledge (Very specific to the project being analyzed)
  - Human input based on an Engineer’s judgment
  - Collaboration among the team executing the project

- The “Mathematical” approach
  - Actual historical performance data used to build the model
  - Built in data validation and justification
  - No emotion or individual bias involved

Three Point Duration Estimates

- Historical performance data can be used to validate minimum and maximum duration estimates provided by Engineers

- There are also cases where historical data can be used to generate three point estimates without human input and provide **more accurate simulation results**

Models Based on Historical Performance

- There are almost endless possibilities when it comes to analyzing data and developing statistical models
- At a higher level, let’s discuss a model that works well in my environment (multi-year development programs)
- Basic requirements to apply this type of model
  - Need at least three months of historical performance data (Could be less in a weekly status environment)
  - Need enough data points (completed tasks) to run a distribution fitting function with statistical software
  - Need a process to identify and remove outliers from historical data
  - Simulation results are more accurate when performing the analysis anywhere from eighteen to three months before the milestone event is scheduled to complete

Baseline Duration Variance

- First you have to determine how well the project has actually performed to date
  - In order to determine this you need to compare the actual duration of a completed task to the task's original planned baseline duration
  - This can be done by calculating the percentage of baseline duration variance
    - **(Act Dur – BL Dur) / BL Dur**

- This analysis is independent of whether or not we finished a task on time (e.g. a task meeting the baseline finish date)

Baseline Duration Grouping

- Next you need to determine how well the program is performing in a specific range of planned durations
  - For example, there could be a lot more variance in tasks with lower baseline durations
- You can group the data by 10 day increments and calculate the baseline duration variance for each completed task within that range
  - For example, how well you performed on tasks with planned durations between 1 and 10 days, 11 and 20 days, 21 and 30 days, etc.
- Build custom distribution curves specific to each grouping

Custom Distributions Applied by Group

- Filter the IMS for all of the remaining effort leading up to the milestone event being measured
  - Apply the appropriate distribution to the remaining duration of each task (based on the original baseline duration grouping)
  - Don’t apply distributions where it obviously does not make sense

- Run the initial simulation with your statistical software

- Afterwards, every task with a distribution curve applied should have its own unique distribution based on the simulation results
  - You can then determine what values to use for three point estimates
  - For example, P=.05 of the distribution can be used for the minimum value and P=.95 can be used to determine the maximum value

Raytheon

Simulated Three Point Estimates

- Now let’s take a timeout for a minute!
- My original intention was to run an initial simulation on the file without any input from the program and therefore placing no additional burden on the team
- Then I would simply use the “auto-generated” three point estimates to compare against the Engineer’s estimates
  - Gives the Planner additional information to facilitate discussions and question inputs
  - Data could also be used as additional backup justification

And, oh by the way, this is undeniably how you have actually performed against your plan to date!

Validating the Model

- Then something very interesting happened! 
- I decided to back test this model against some SRAs conducted in the past using Engineer's three point estimates
- This testing requires a project where a major milestone event has completed
  - All you need to know is the actual finish date
- Then you need an archived copy of the original IMS (from a point back in time) where an SRA was ran to that milestone before it was completed

Comparison Results

- Using the archive copy of the file, follow the steps previously outlined (same assumptions apply)
  - Pull the historical data out of the archive IMS
  - Build custom distribution curves for each baseline duration group
  - Apply the appropriate distribution to the remaining duration of each task
  - Run a simulation on the archive file to determine the expected milestone date
- Now that you have a date based on this model you can compare it to the date produced by the original SRA
  - Just be sure to compare dates that have the same level of confidence in both assessments
- Simulation results using the auto-generated three point estimates were **more accurate**, and produced expected dates closer to the dates that the milestones had actually occurred

Raytheon

Combine the Best of Both Techniques

- Processes for obtaining three-point estimates directly from Engineer's can vary widely, lack definition, and sometimes even reasoning
  - However, I would never discount the Engineer's true judgment

- Take advantage of the Engineer's knowledge and experience (the Art) combined with historical performance (the Math)

- Run a "pre-simulation" to provide auto-generated three point estimates to the Engineers as a starting point
  - They can address high risk activities or tasks on the critical path providing justification where necessary
  - They can also identify tasks that should be excluded from the simulation
  - Provides them with an opportunity to update minimum and maximum durations where it makes more sense before running another simulation

Wide Range of Application

- There are so many possibilities to create models when it comes to grouping data
  - You can model human behavior by focusing on an individual Engineer’s historical planning performance
  - Model specific efforts or types of work (e.g. qualification testing)
  - Modeling performing organizations or Integrated Product Teams
  - Modeling the build of similar products

- Always be cautious
  - Are you really comparing like sets of data?
  - Can you identify any correlation or the lack thereof?
  - Will the analysis help the program team make real decisions?

If you go overboard you can really get lost in the data

Raytheon

Conclusion

- Statistical analysis can be applied on a regular basis and does not have to break the bank

- The application of actual performance trends leads to **program management efficiencies**
  - Models easily adjusted at major planning intervals (e.g. rolling wave)
  - Continuously improve duration estimates and planning with passage of time
  - Helps to optimize the Integrated Master Schedule as a predictive tool