University of Maryland

Project Management Symposium

NEXT SESSION

Turbocharging Your Estimation Process Using Data You (Probably) Already Have

Nana Kwame

Writer & speaker on PMOs
Senior Manager, Enterprise PMO

Project Management Center for Excellence
A.J. Clark School of Engineering
Civil & Environmental Engineering Department

This session will be recorded.

Project Management Symposium

Turbocharging Your Estimation Process Using Data You (Probably) Already Have

Nana Kwame

University of Maryland 1856

Project Management Center for Excellence

A.J. Clark School of Engineering

Civil & Environmental Engineering Department

About DC Water

McMillan Water Treatment Plant (photo credit: DC Water)

Blue Plains Advanced Wastewater Treatment Plant (photo credit: DC Water)

DC Water Headquarters (photo credit: SK&A Engineers)

Some Awards….

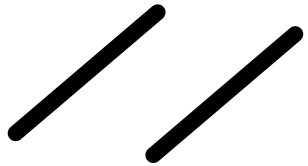
PROJECT MANAGEMENT CENTER FOR EXCELLENCE
A.J. CLARK SCHOOL OF ENGINEERING
Civil & Environmental Engineering Department

GLOBAL WATER AWARDS

CSWEEK
Expanding Excellence in **Customer Service**

Project Management Institute

What is an Estimate?



“a quantitative assessment of the likely amount or outcome of a variable”

- PMBOK Guide 7th edition

Some key terms….

Range

A list of possible outcomes within a set lower & upper limit

01

Accuracy

The correctness of an estimate

02

Precision

The degree of exactness ascribed to the estimate

03

Confidence

The level of certainty (%) that the estimate will fall within a given range

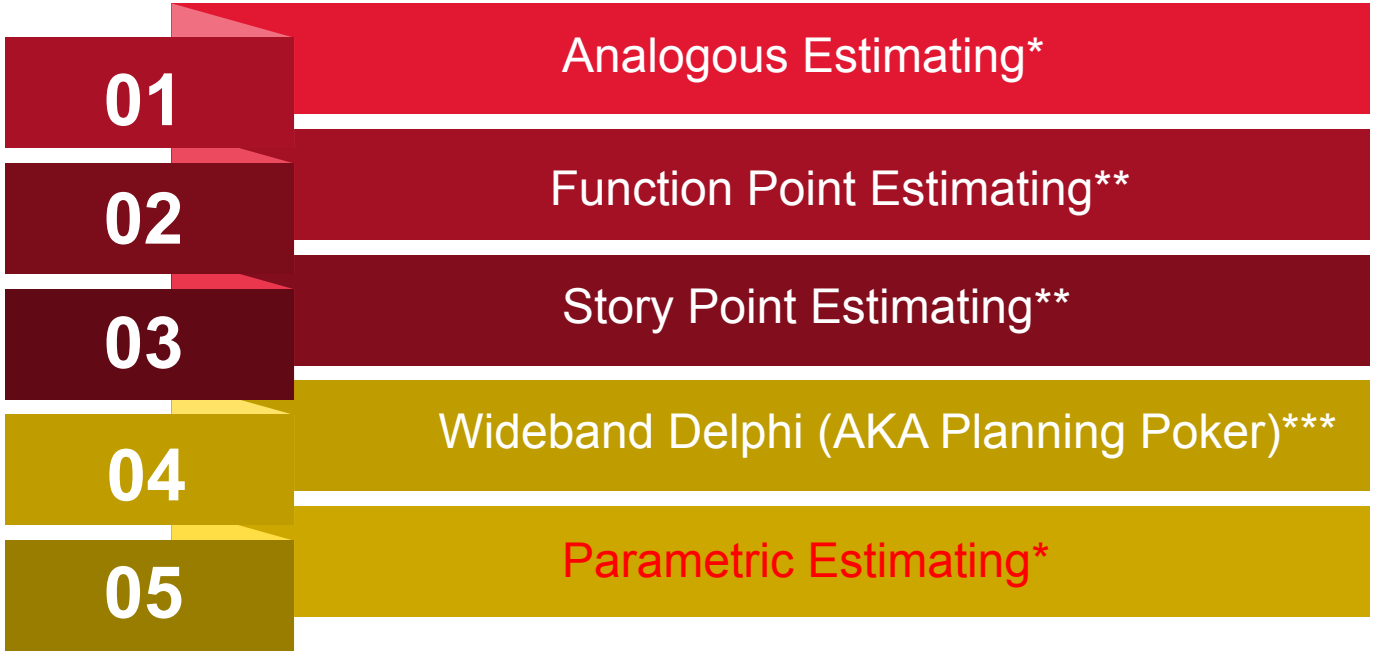
04

Deterministic vs Probabilistic

Point (or point in time) estimate vs range of estimates (with associated probabilities)

05

Common Estimation Methods



* Top-down estimation method
** Bottom-up estimation method
*** Top-down or bottom-up estimation method

Data Collection: Building your Estimates Database

PROJECT MANAGEMENT CENTER FOR EXCELLENCE
A.J. CLARK SCHOOL OF ENGINEERING
Civil & Environmental Engineering Department

Actuals

- Timesheet Hours
- Costs
- Durations
- Story Points
- FTE

Estimates

- Effort Hours
- Costs
- Durations
- Story Points
- FTE

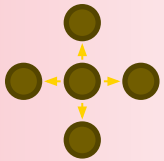
Automated (Weekly) Inputs

Manual Inputs

ESTIMATES DATA WAREHOUSE

Dimensions include:

- Task
- Assignment
- Sprint
- Release
- Epic
- Phase
- Project/Program
- Resource
- Portfolio
- Project Size
- Complexity
- Project Methodology
- Estimation Method



Reporting

Data Analytics

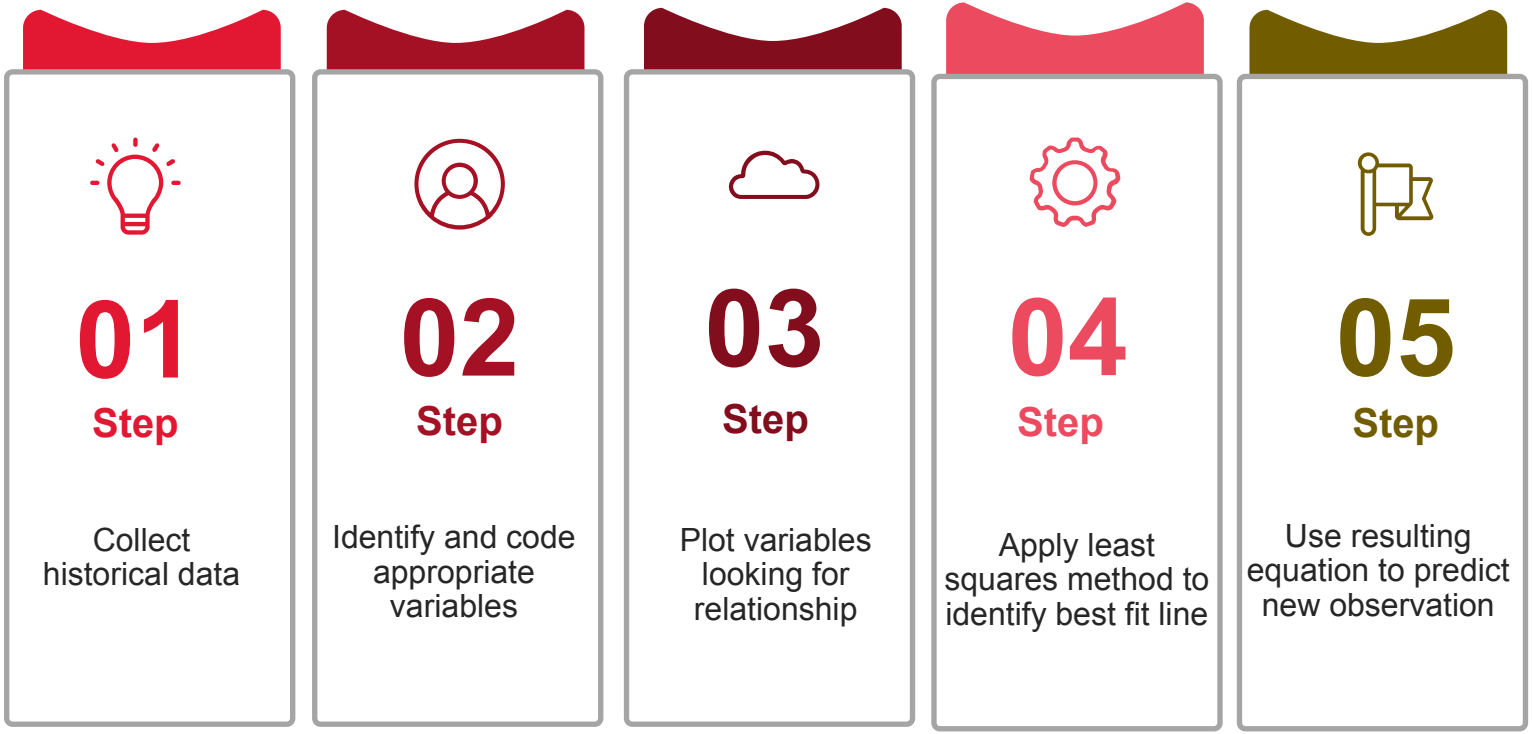
Predictive Modeling

Error Reduction

Estimates Generation: Parametric Models

- Compares each variable against Y to determine the best fit
- Historical data must be collected
- Model must be scalable
- Use excel's data analysis add-in, BI tools or build into your PMO tool!

Parametric Models - Process



Parametric Models - Example

What independent variable that best predicts individual sprint costs on an agile project?

Effort Hours?

Median Resource Rate?

Story Points + Effort Hours?

Story Points?

Story Points + Median Rate?

Effort Hours + Median Rate?.

All 3 Variables?

Is story points the "best fit" variable?

Regression Statistics	
Multiple R	0.928153612
R Square	0.861469127
Adjusted R Square	0.860486638
Standard Error	18415.5081
Observations	143

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	170.3697701	3405.237144	0.050032	0.960168	-6561.550767	6902.290307	-6561.550767	6902.290307
Story Points	1991.696973	67.26159075	29.61121	2.18E-62	1858.725417	2124.66853	1858.725417	2124.66853

…maybe effort hours is the “best fit”

Regression Statistics	
Multiple R	0.986691192
R Square	0.973559508
Adjusted R Square	0.973371986
Standard Error	8045.354577
Observations	143

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	394.7730812	1415.175161	0.278957	0.780687	-2402.931257	3192.47742	-2402.931257	3192.47742
Effort Hours	94.64305085	1.313507913	72.05366	4E-113	92.0463357	97.239766	92.0463357	97.239766

…perhaps median rate is the “best fit”?

Regression Statistics

Multiple R	0.055019181
R Square	0.00302711
Adjusted R Square	-0.004043619
Standard Error	49402.8603
Observations	143

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	147345.0176	87582.12938	1.682363955	0.094711542	-25798.85186	320488.887	-25798.85186	320488.887
Median Rate	-604.2361611	923.4740748	-0.65430766	0.513979431	-2429.881138	1221.408816	-2429.881138	1221.408816

… will effort hours + story points do the trick?

Regression Statistics	
Multiple R	0.986926911
R Square	0.974024728
Adjusted R Square	0.973653653
Standard Error	8002.690308
Observations	143

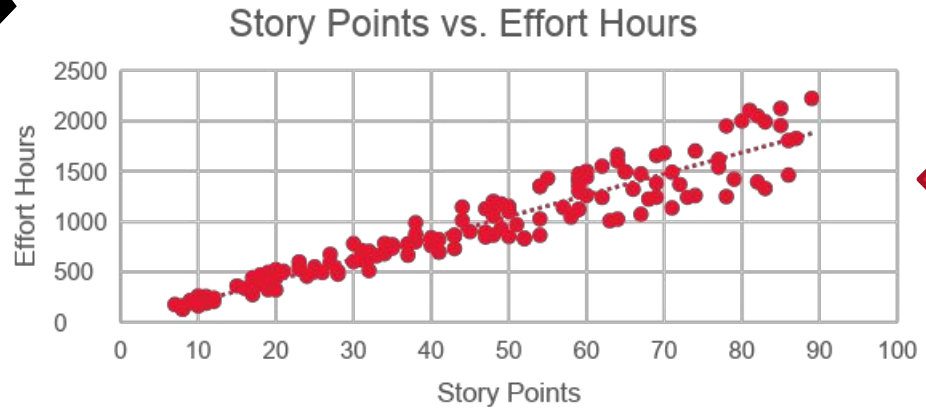
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1119.939367	1480.290693	0.756567188	0.45058	-1806.674854	4046.553589	-1806.674854	4046.553589
Story Points	-144.9539788	91.54124662	-1.583482683	0.115568	-325.9359409	36.02798336	-325.9359409	36.02798336
Effort Hours	100.7832641	4.091861243	24.63017637	9.63E-53	92.69343459	108.8730936	92.69343459	108.8730936

…. what about adding all 3 variables?

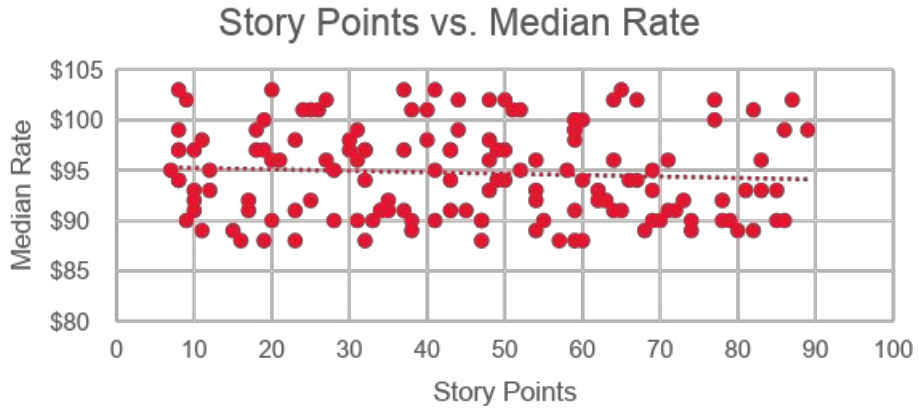
Regression Statistics	
Multiple R	0.986930044
R Square	0.974030911
Adjusted R Square	0.973470427
Standard Error	8030.469419
Observations	143

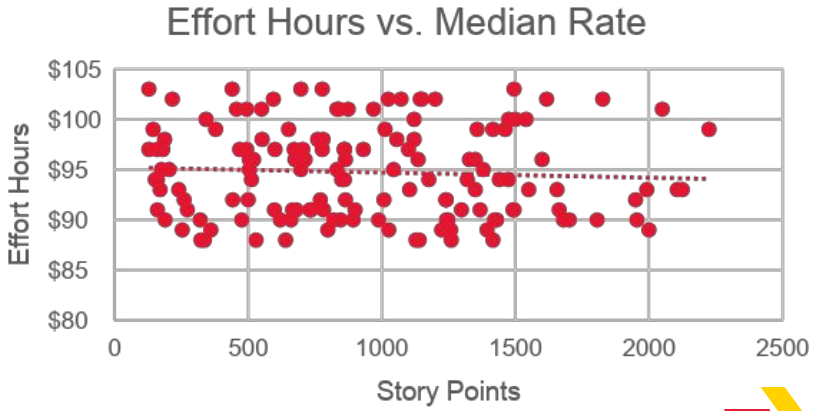
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-1493.244092	14440.95581	-0.103403411	0.917791844	-30045.58023	27059.09205	-30045.58023	27059.09205
Story Points	-144.0676245	91.98812561	-1.566154583	0.119586212	-325.9444974	37.8092484	-325.9444974	37.8092484
Effort Hours	100.759938	4.108066487	24.52733868	2.38746E-52	92.63756049	108.8823156	92.63756049	108.8823156
Median Rate	27.39527198	150.588517	0.181921388	0.855909452	-270.3449894	325.1355334	-270.3449894	325.1355334

Checking for collinearity….



Collinearity problem!





And the winner is….

ANSWER: Effort Hours, represented by:

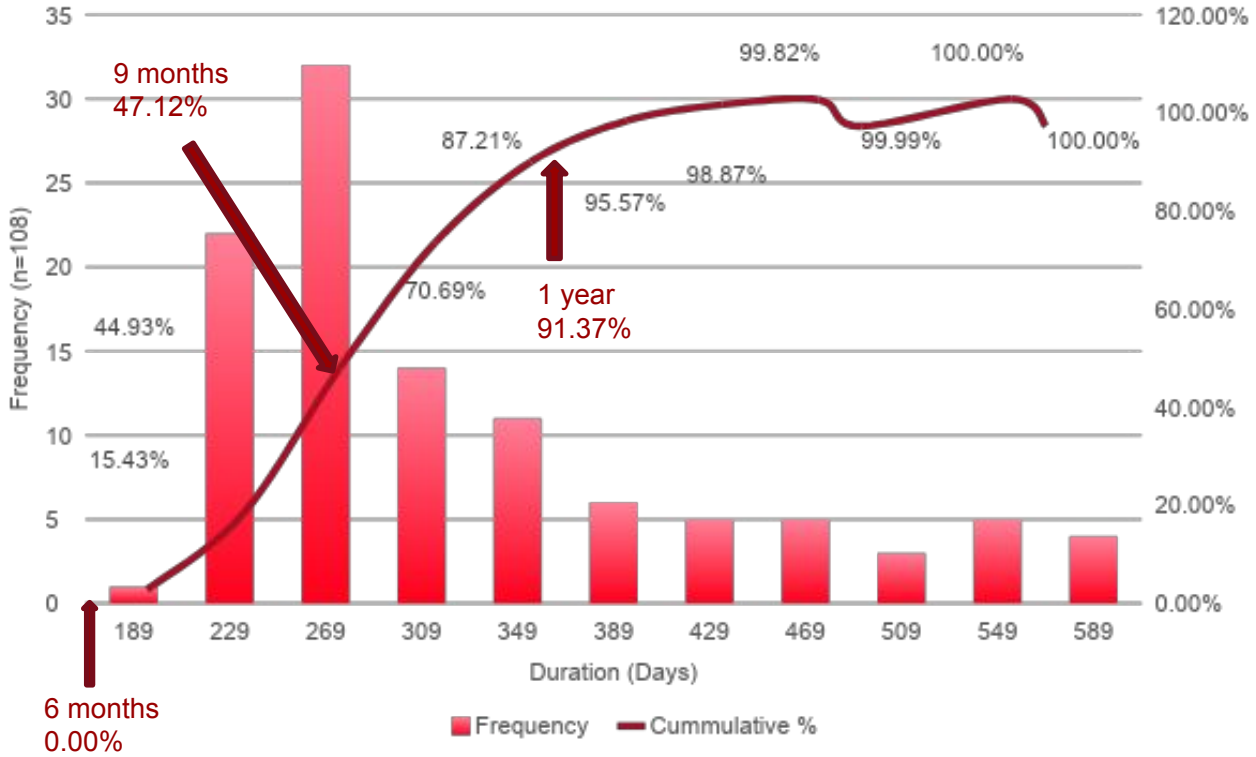
$$y = 94.059x + 394.77$$

(i.e. 1 hour of effort is worth 94.059 + 394.77 = $488.83 @95% confidence level)

Estimates Generation: Beta Distribution

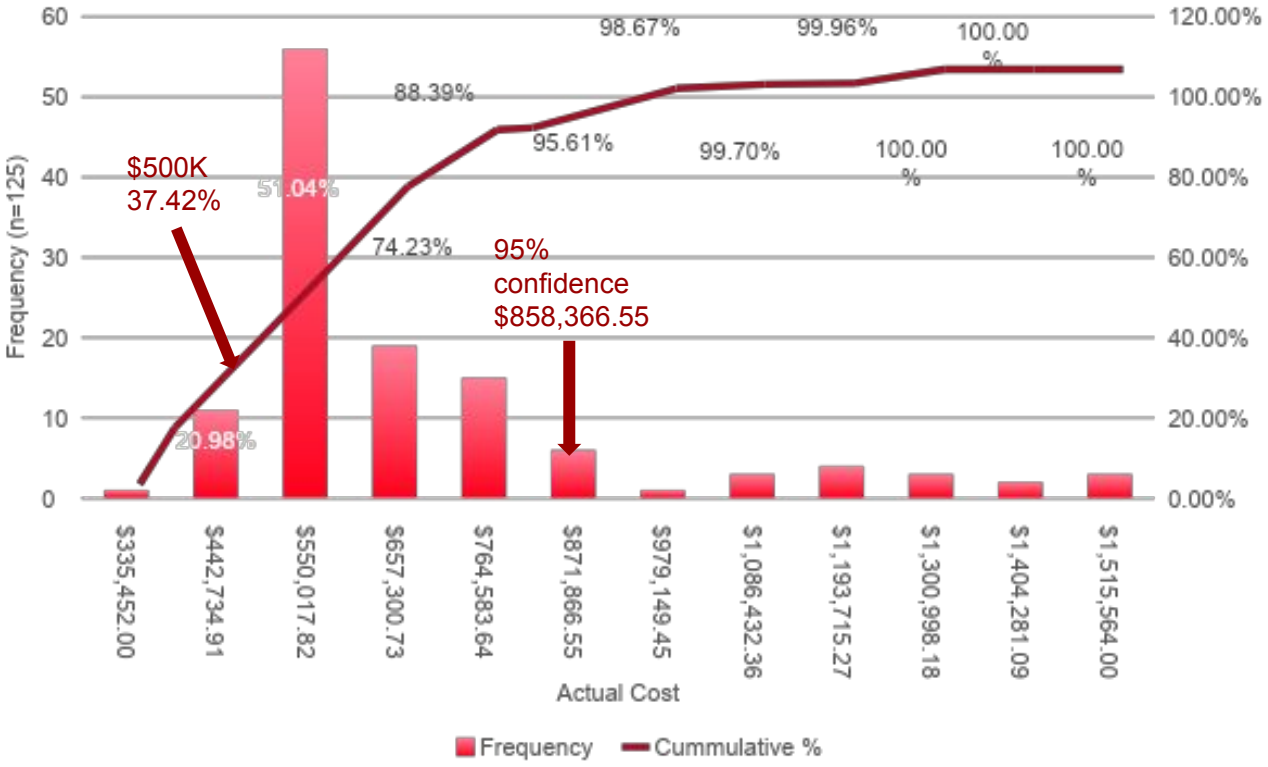
- Our cost and duration data has “best fit” into a beta
- Historical data must be collected
- Use excel’s beta dist. function or build into your PMO tool!

Beta Distribution– Duration Prediction Example



Sample data from 108 App Dev projects
Mean = 308.42 days
Standard deviation = 102.63
Median = 267 days

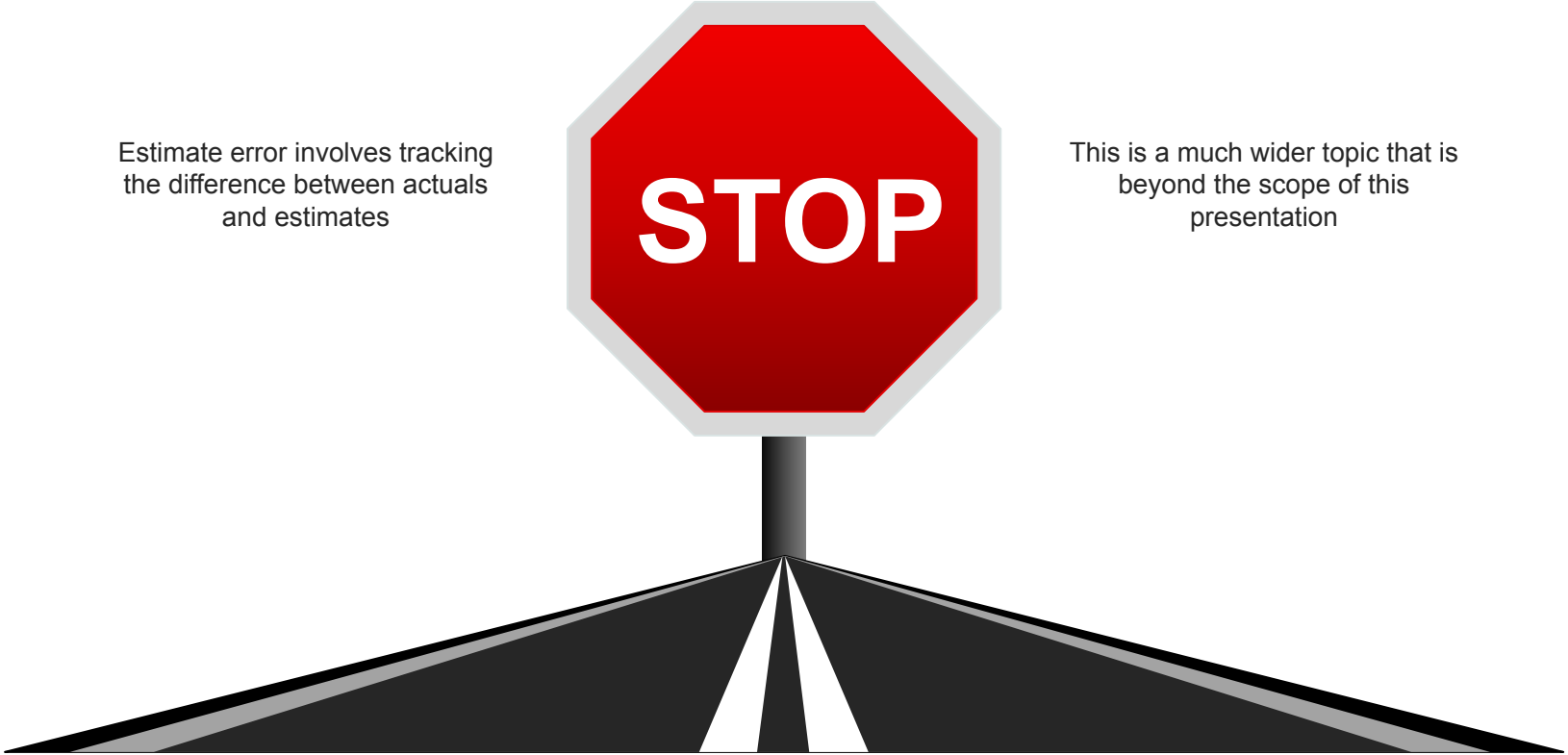
Beta Distribution– Cost Prediction Example



Sample data from 125 COTS projects
Mean = $633,762.57
Standard deviation = 259,427.28
Median =$540,431.00

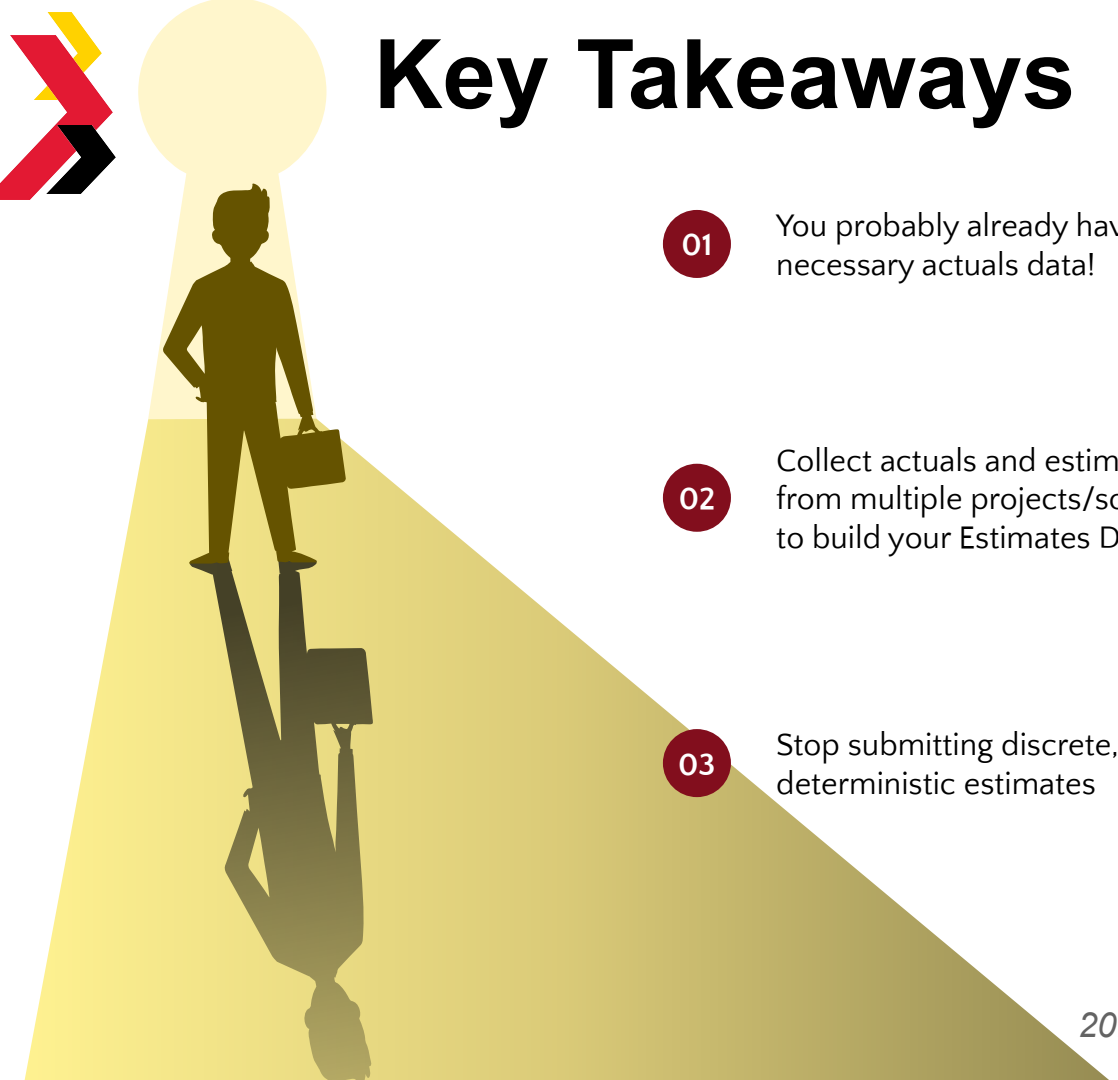
Estimates Optimization: Tracking Estimate Error

Estimate error involves tracking the difference between actuals and estimates



This is a much wider topic that is beyond the scope of this presentation

Key Takeaways



PROJECT MANAGEMENT CENTER FOR EXCELLENCE
A.J. CLARK SCHOOL OF ENGINEERING
Civil & Environmental Engineering Department

01 You probably already have the necessary actuals data!

02 Collect actuals and estimates from multiple projects/sources to build your Estimates DB.

03 Stop submitting discrete, deterministic estimates

04 Both models presented today are “top-down”

05 Use parametric models to test which variables best influence an estimate for better prediction

06 Use β distributions for prediction if parametric models are not an option

Let's continue the conversation!

Nana Kwame

Building metrics-based PMOs | 2020 PMI PMO of the Year A...

Evaluate Session

PROJECT MANAGEMENT
CENTER FOR EXCELLENCE
A.J. CLARK SCHOOL OF ENGINEERING
Civil & Environmental Engineering Department