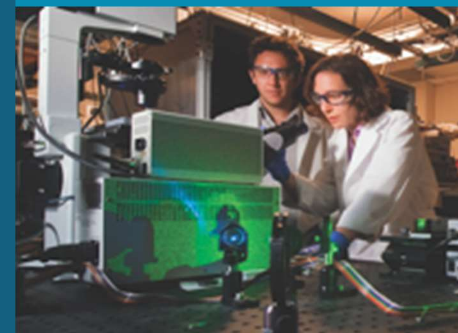
Sandia National Laboratories

Employment Application Arrival Model for Talent Acquisition Simulation and Management



PRESENTED BY

Thor D. Osborn, PhD, MBA, CAP

ENERGY NNSA

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.
SAND2021-2973 C

Presentation Objectives and Outline

Make the business case for modeling employment application arrival patterns and rates

Describe a straightforward process for creating concise, broadly applicable models based on historical application data

Demonstrate the process on data from a large research organization and analyze the results

- Business Case
- Technical Framework
  - Understanding the Data
  - Source Models
  - Analysis
  - Model Development
- Discussion
- Concluding Remarks

Motivations for Modeling Employment Application Arrival Patterns and Rates (1 of 2)

- Continuous Improvement of Talent Acquisition
  - Hire rate and lag depends on rate of flow through vetting stages in the Talent Acquisition Pipeline (TAP)
  - Performance of TAP processes relies on sufficiency of employment applications
  - Application rates and patterns vary widely – field, specificity, competition, and advertisement are frequently cited explanatory variables
  - A common mathematical framework for application arrival may enable better understanding of trade space for improving application capture rates
  - Key relationships
    - Application capture rate and variance vs. employment context – job site, career level, field of practice
    - Capture rate impacts of adjustable variables – advertisement, job posting specificity, job posting language, targeted recruiting efforts
    - Capture rate impacts of external factors – economic conditions, competitors for field of practice, professional population within rational recruiting area

How can the Talent Acquisition function best address the *triple constraint* - time to collect sufficient applications, quality of applicants, and cost per application?

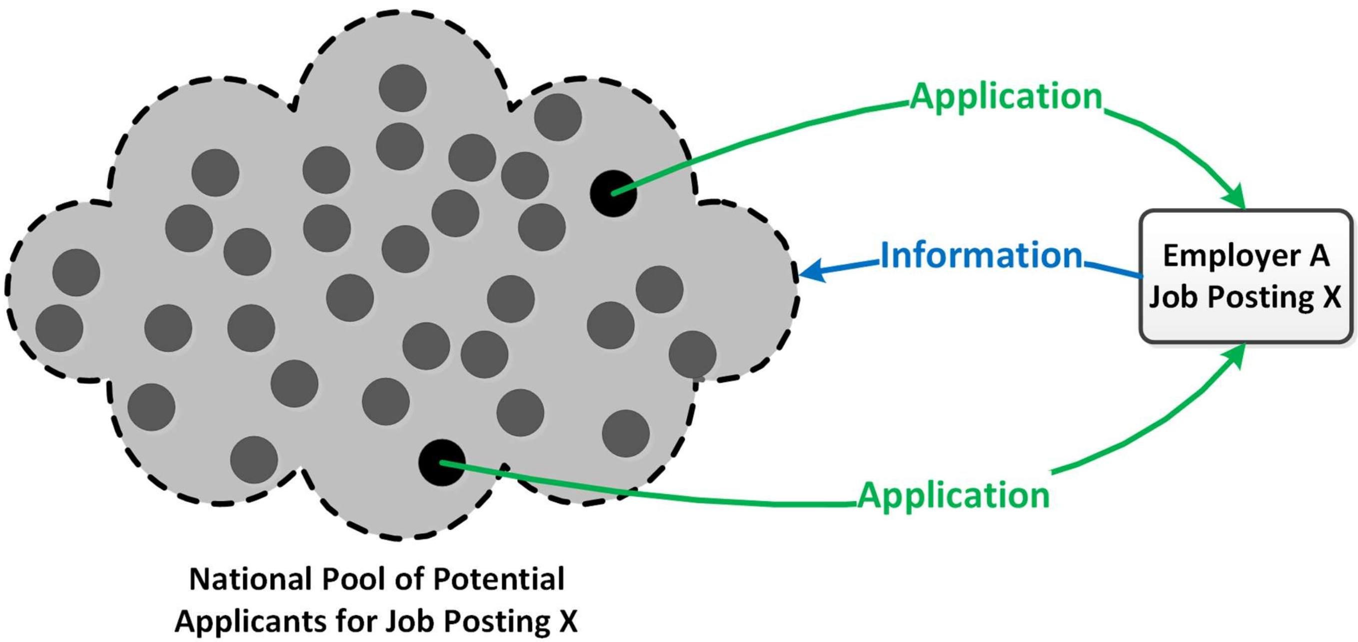
Motivation for Modeling Employment Application Arrival Patterns and Rates (2 of 2)

- Managing Executive Expectations
  - Executive leadership often sets headcount goals through Work-Force Planning (WFP)
  - Absent relevant models, consideration of triple constraint in allocation of Staffing budget may be subjective or absent
- Managing Hiring Manager Expectations
  - Arrival patterns of small numbers of applications may activate pattern biases
    - Any of several cognitive biases characterized by a tendency to imbue meaning to patterns within data that could readily be explained by random action
    - Examples include identification of trends based on a few successive outcomes or assignment of complex rationales to explain short bursts
  - Unchecked, intuitive response to biases may lead to detrimental decisions
    - Appearance of declining application rate may encourage premature closure of posting window based on perception of increasing scarcity
    - Comparison of immediate response vs. expectations based on prior experiences (anchoring bias) may lead to dissatisfaction with Talent Acquisition function
    - Models based on more comprehensive data may help to reset expectations

Employment Application Arrival Data

- Arrival data are tied to specific job requisitions
- Job requisitions are characterized in several ways
  - Job site (location)
  - Career phase (early career vs. experienced professional)
  - Visibility (broadly accessible vs. internal only)
  - Field of practice (e.g., mechanical engineer, chemist, electronics technician)
  - Specific requirements
- Applications may be submitted during the window of time when the job posting for the requisition is accessible
- Submission of completed applications is tracked by date
- Date of last submitted application is treated as posting closure date
- Submissions are counted by date: days within the active posting window without a submission are treated as counts of zero

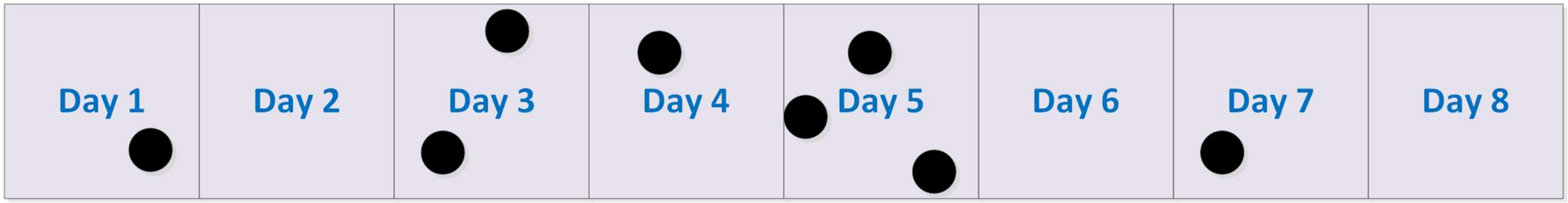
Application Arrivals from the Applicant Source Pool



- Potential applicants learn about an opportunity via the internet, recruiters, or their personal networks
- Few of the potential applicants apply – pool is assumed to be large relative to total applicants
- Applicant pool may be viewed as a source of applications
- Applications arrive at the employer at different times

Application Arrivals Example with Terms and Definitions

Application Arrivals



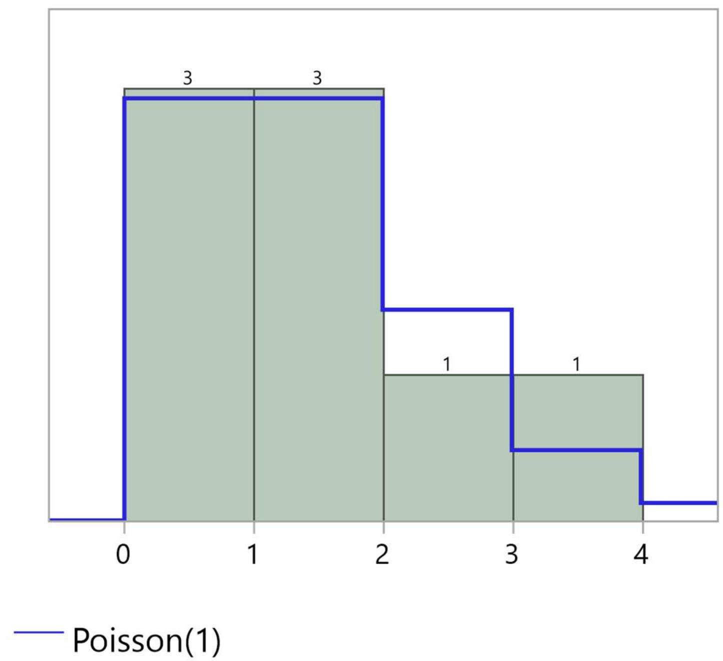
Daily Sums

1 Instance Count = 1	1 Instance Count = 0	1 Instance Count = 2	1 Instance Count = 1	1 Instance Count = 3	1 Instance Count = 0	1 Instance Count = 1	1 Instance Count = 0

Total Instances by Count

Total Instances by Count
3 Instances Count = 0
3 Instances Count = 1
1 Instance Count = 2
1 Instance Count = 3

- Applications are tallied by day of arrival into the organization's HRIS
- The daily tally provides an Instance of a Count
- Total Instances by Count constitute Count Frequency
- Distribution is approximated by Poisson model (at right, blue) with an average rate of one Application Arrival per day



The Poisson Source Model

- The Poisson distribution may be used to describe the probability of count (k) produced in one unit of time by a randomly emitting source of discrete items with a constant mean rate of emission (λ) per unit time, provided that the items behave independently

$$P(\lambda) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}$$

- As a first hypothesis, members of the nationally distributed pool of potential applicants for a specific, broadly advertised job are assumed to act in an uncoordinated manner (i.e., independently) regarding employment opportunities
- Only a small portion of the potential applicant pool is expected to be interested in applying for a specific job at a specific career level and work site location at any given time
- The Poisson distribution offers a reasonable initial hypothesis for the arrival behavior of employment applications

The Gamma-Poisson Source Model

- The variance of the Poisson distribution is equal to the mean rate
- This also applies to an aggregate of Poisson sources
- When the application source is more readily conceived as a collection of non-independent but otherwise 'Poisson-like' sources the variance will exceed the aggregate rate
- The gamma-Poisson distribution is often used to represent such phenomena, and comprises a mixture of Poisson components using the gamma distribution as the mixing distribution

$$Var(P(\lambda)) = \lambda$$

$$Var(\sum_i P(\lambda_i)) = \sum_i \lambda_i$$

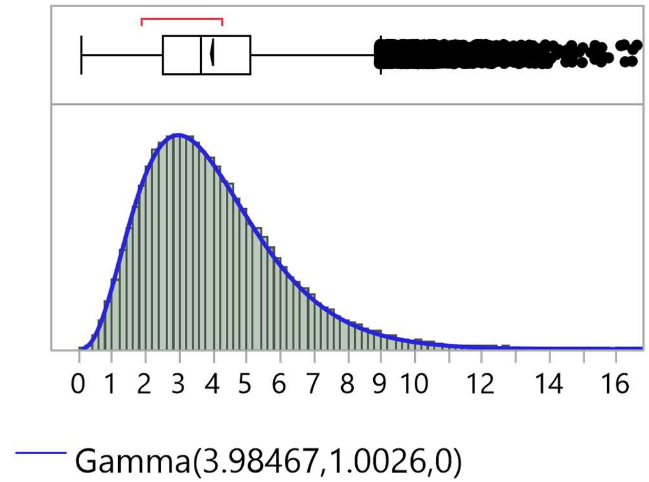
$$Var(\sum_i Q(\lambda_i)) > \sum_i \lambda_i$$

GammaPoisson is expected to fit better under the alternative hypothesis that applicants for employment behave in a substantially coordinated manner

$$GP(\lambda,\sigma) = \frac{\Gamma\left(k + \frac{\lambda}{\sigma - 1}\right)}{\Gamma(k+1)\cdot\Gamma\left(\frac{\lambda}{\sigma-1}\right)} \cdot \left(\frac{\sigma - 1}{\sigma}\right)^k \cdot \sigma^{\left(\frac{\lambda}{\sigma-1}\right)}$$

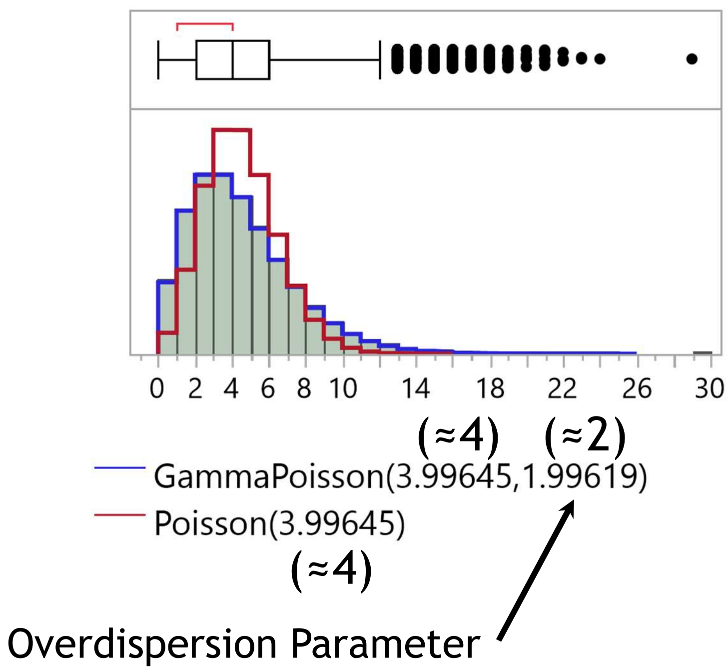
Generation of the Gamma-Poisson Distribution

- Generate random gamma distributed data (mean = 4, N = 100K)



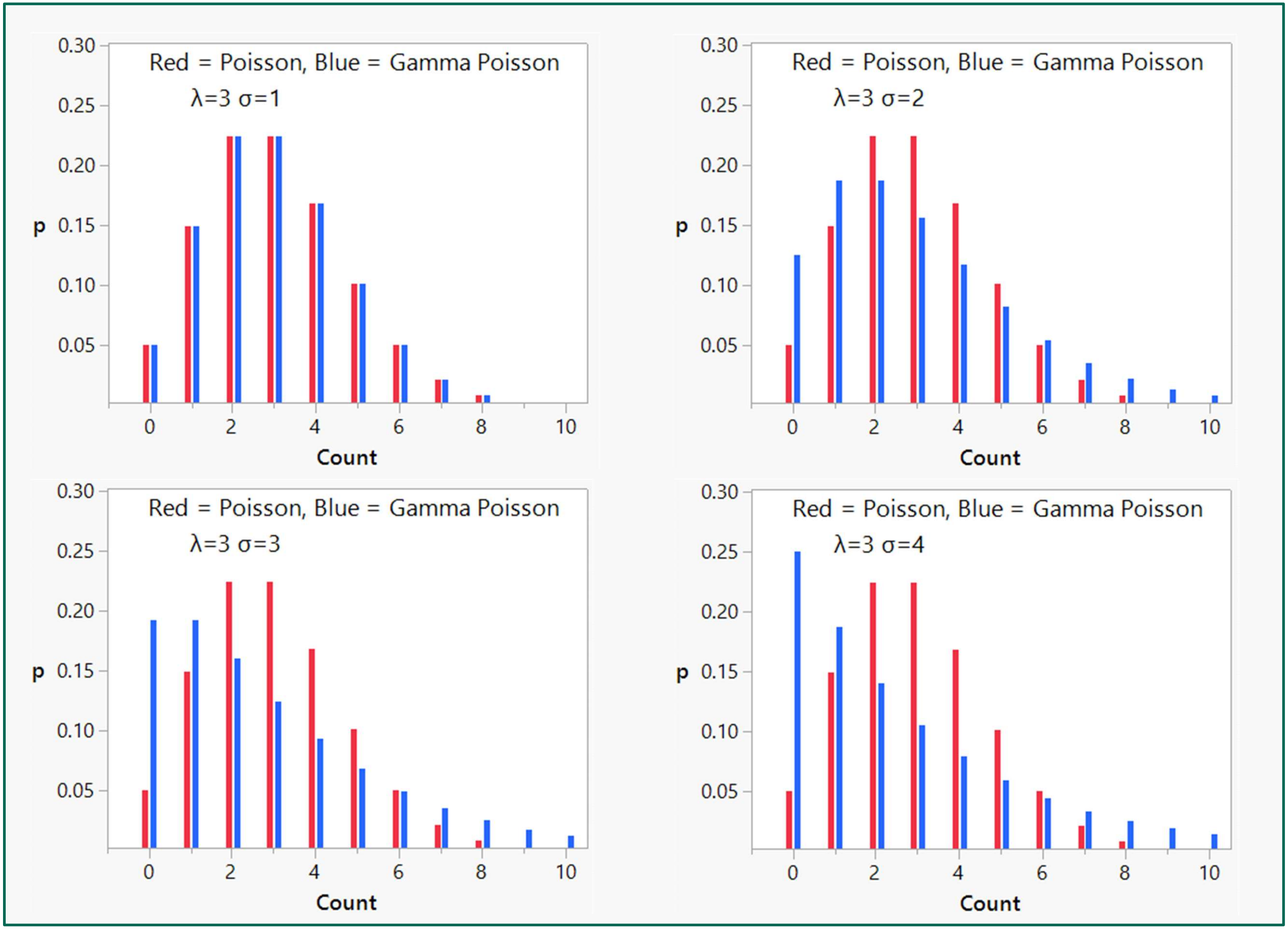
- Generate mixed random Poisson distributed outcomes using random gamma as the Poisson parameter (λ)

- Result is a discrete distribution with greater variance than the corresponding Poisson

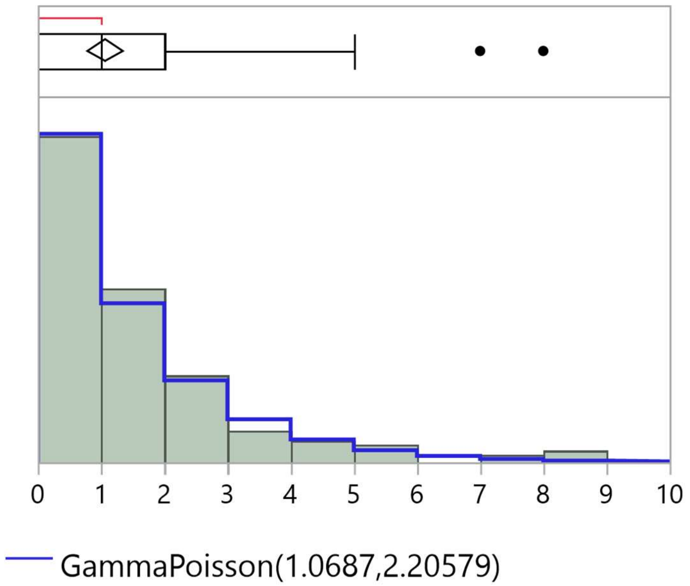


Impact of Gamma-Poisson Overdispersion Parameter

The overdispersion parameter, σ , reflects increased variance

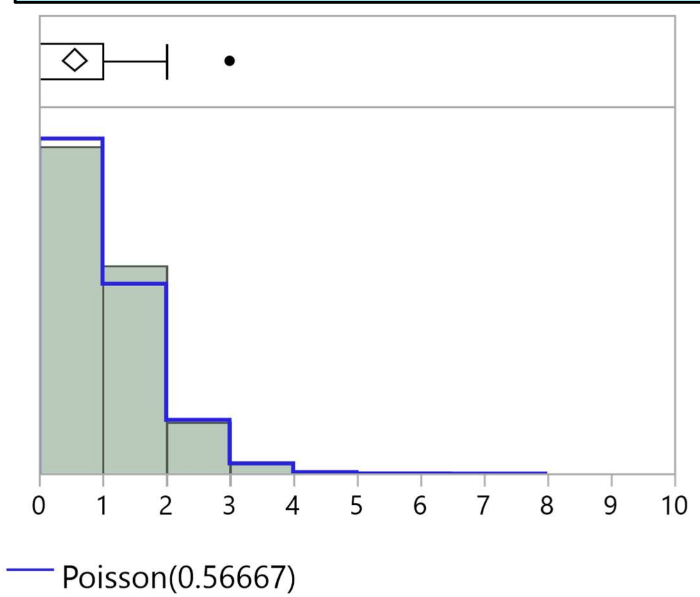


Example Count Distributions for Application Arrivals



Count distribution for a broadly accessible early-career mechanical engineering discipline job posting (140 applications / 131 counts)

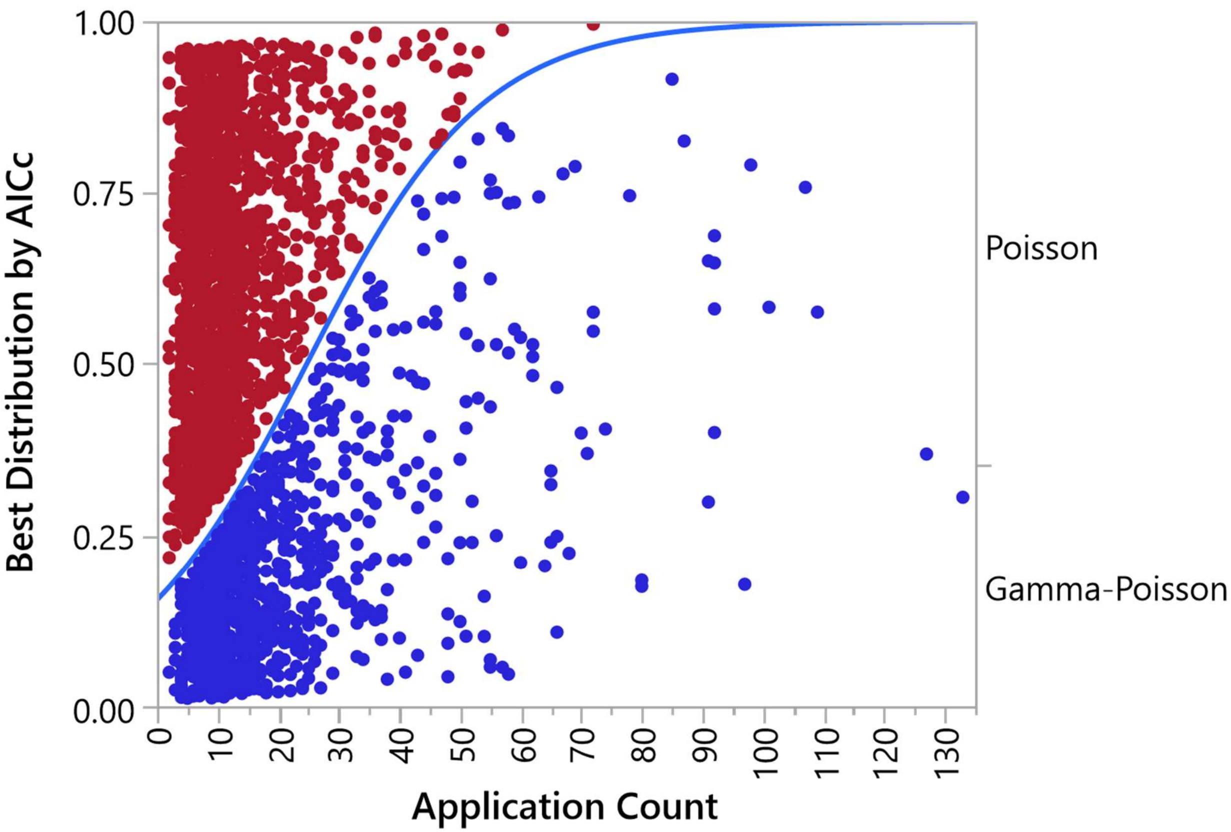
Distribution shown is best fit among Poisson and GammaPoisson by Akaike's criterion



Count distribution for a broadly accessible experienced professional mechanical engineering job posting (34 applications / 60 counts)

Best Distribution and Total Applications

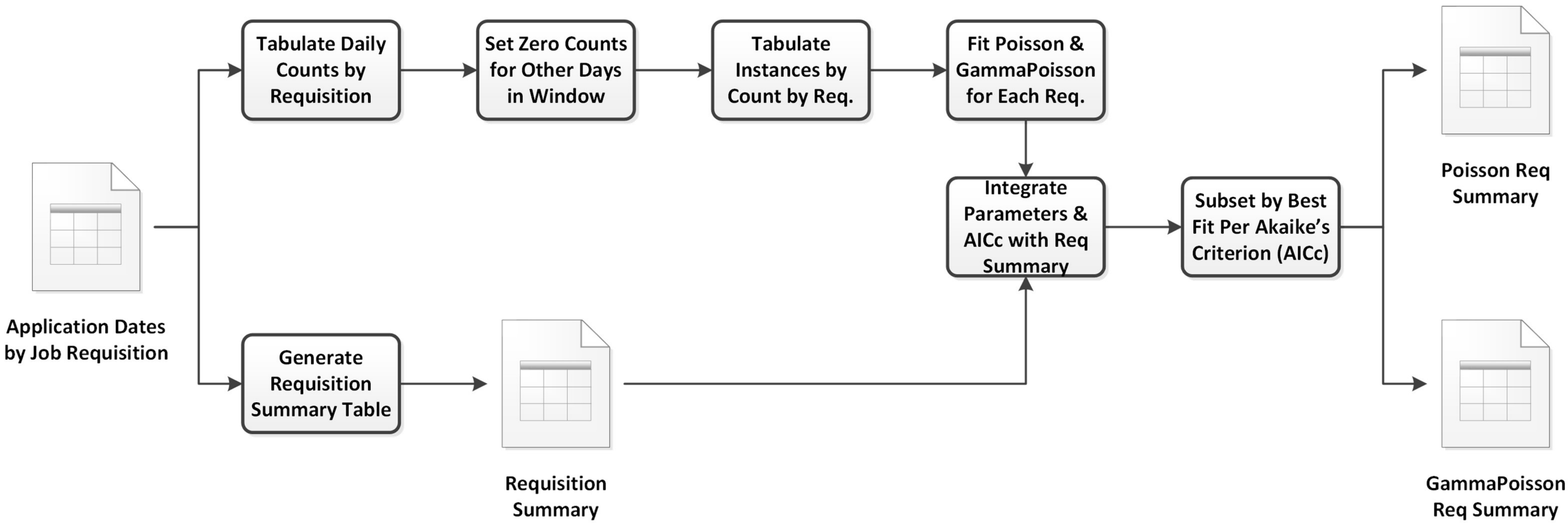
The best-fitting model is strongly related to the total application count – the requisitions that garner the most attention are most likely to be Gamma-Poisson distributed



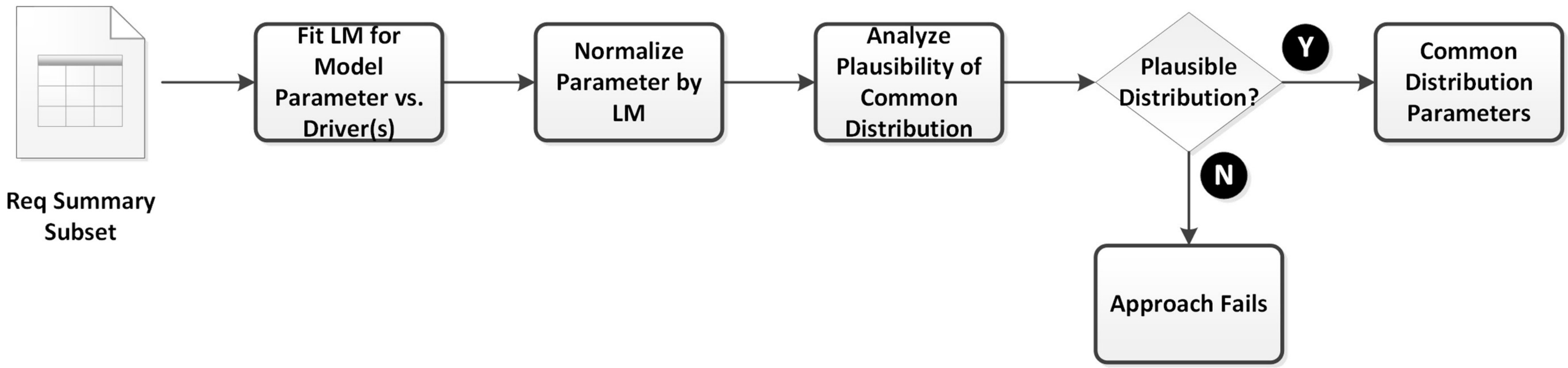
Career Stage (Early) and FLSA Status (Non-Exempt) were also significant factors favoring the GammaPoisson distribution

Causation has not been attributed; however, circumstances encouraging greater sharing of information or synchronization of information could lead to larger and more coordinated applicant response

Data Preparation



Normalization and Distribution Fitting

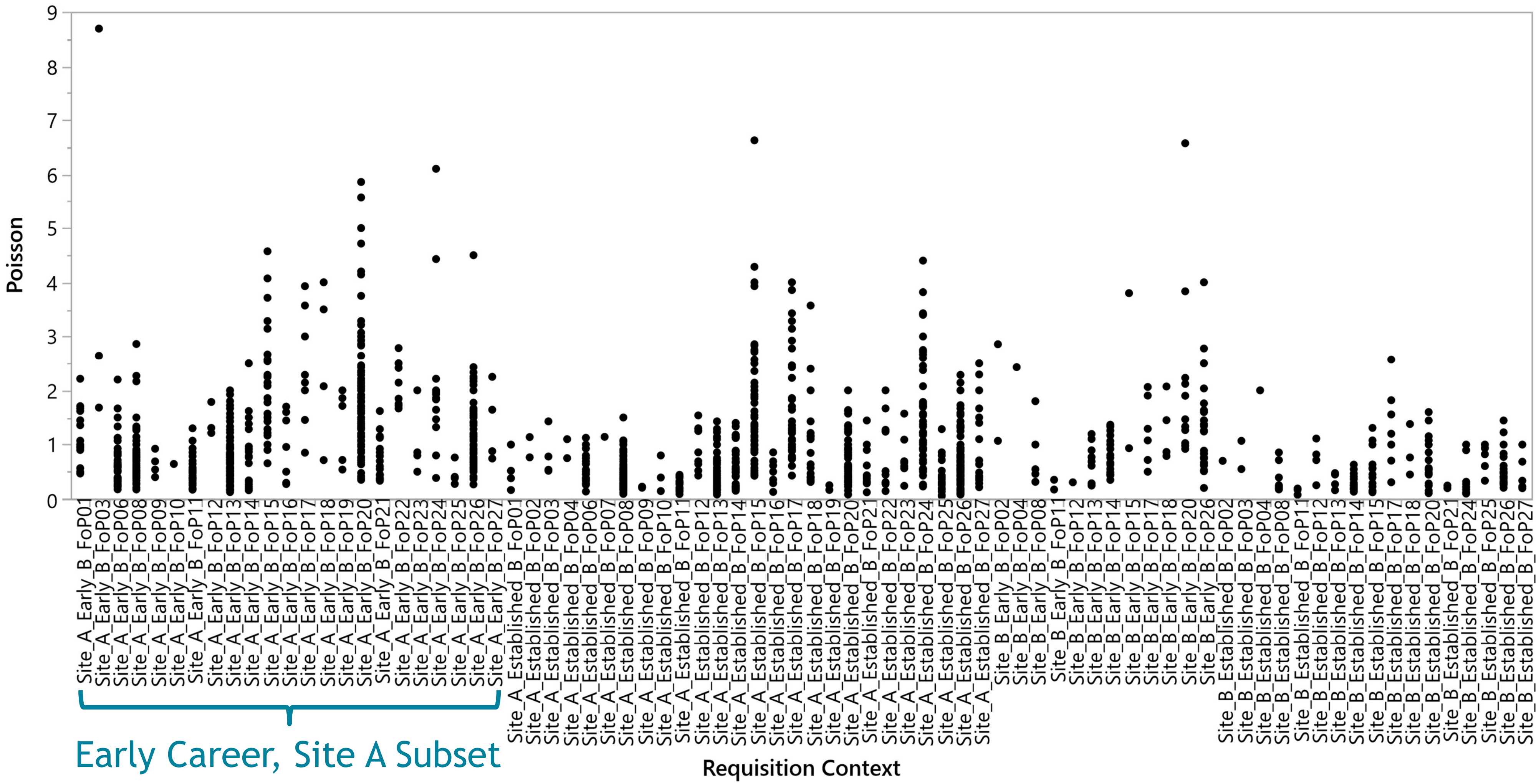


Raw Poisson Parameter Distribution by Context

Data Source:

Poisson Req Summary

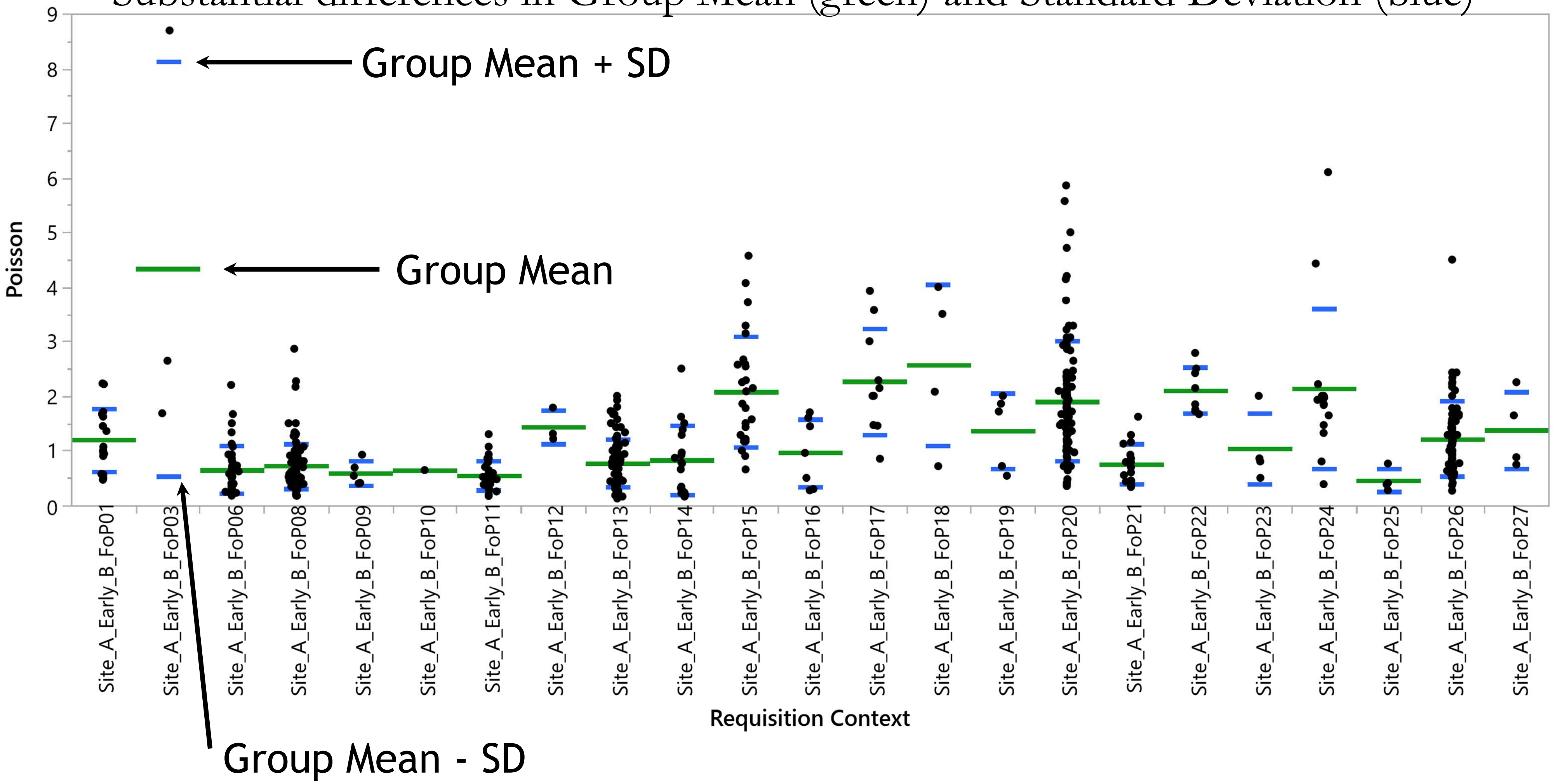
- Mean and variance clearly differ by Requisition Context



Analysis

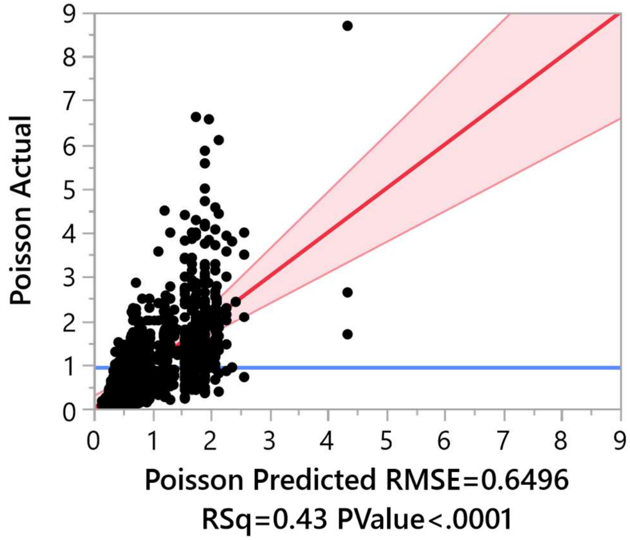
Raw Poisson Parameter Distribution Subset by Context

- Subset of Early Career requisitions for Site A
- Substantial differences in Group Mean (green) and Standard Deviation (blue)

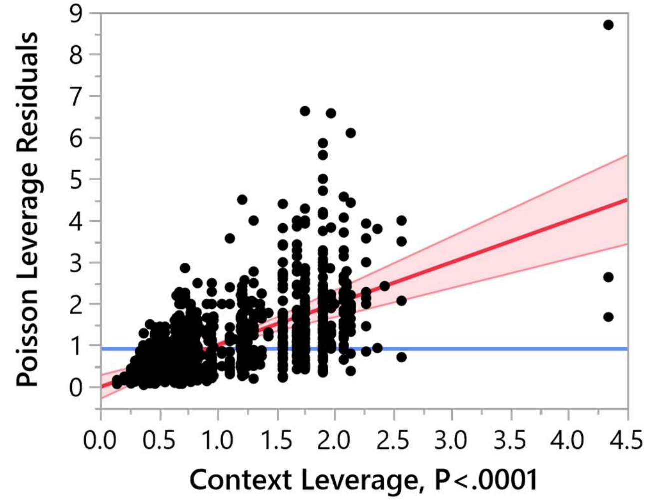


Context Based Normalization Function for Poisson Parameter Distribution

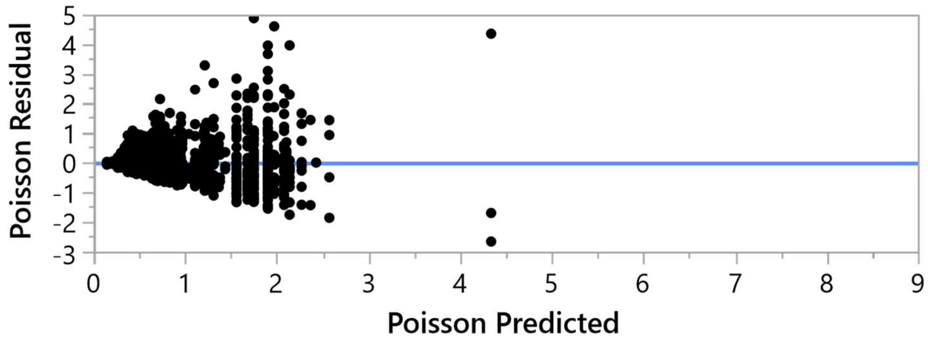
Actual by Predicted Plot



Leverage Plot



Residual by Predicted Plot



- Variance grows with mean prediction
- Normalization is expected to decrease dispersion of variance across Context

Summary of Fit

RSquare	0.43167
RSquare Adj	0.401572
Root Mean Square Error	0.649615
Mean of Response	0.940054
Observations (or Sum Wgts)	1532

Analysis of Variance

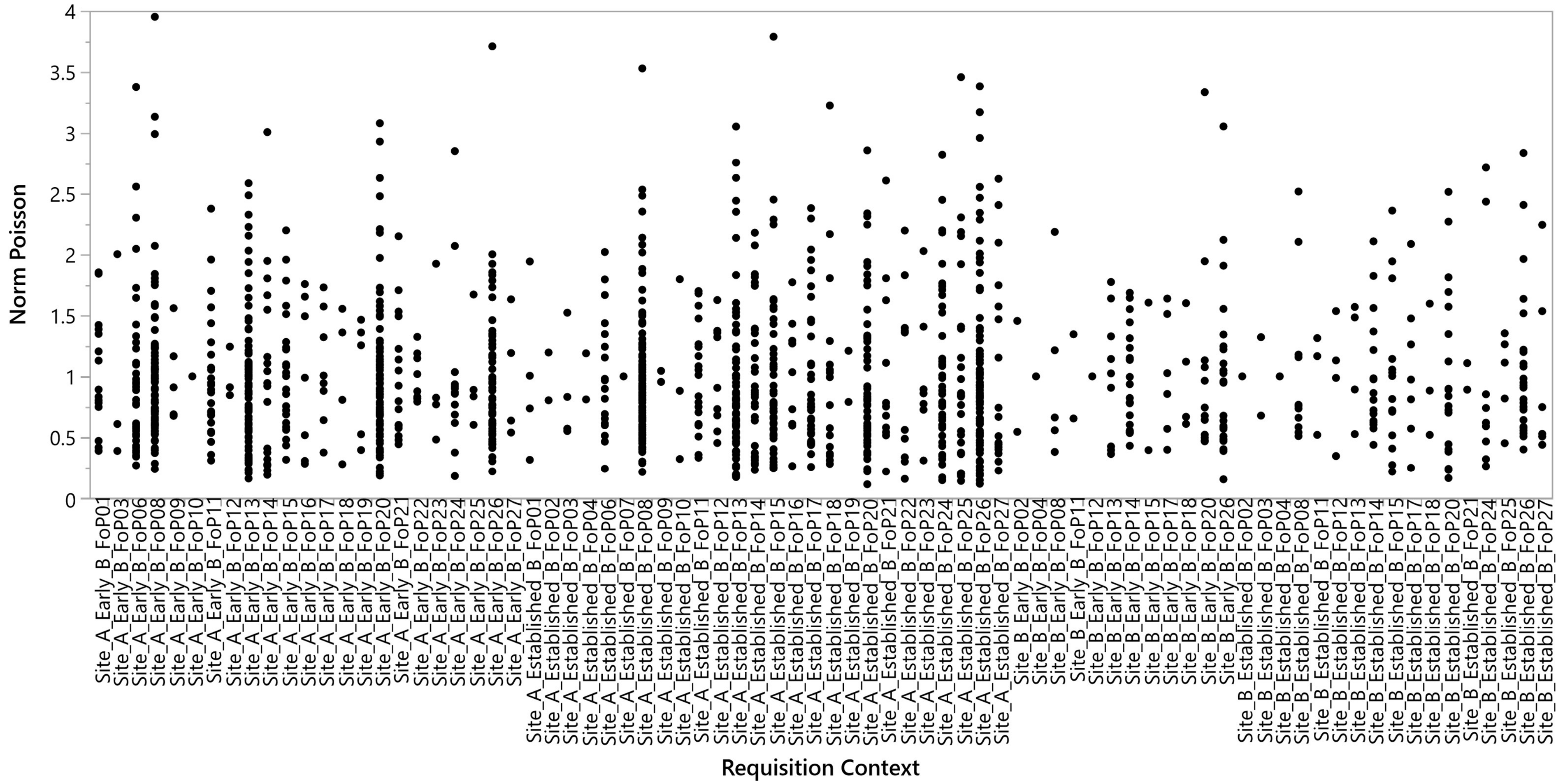
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	77	466.0450	6.05253	14.3425
Error	1454	613.5882	0.42200	Prob > F
C. Total	1531	1079.6332		<.0001*

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Context	77	77	466.04501	14.3425	<.0001*

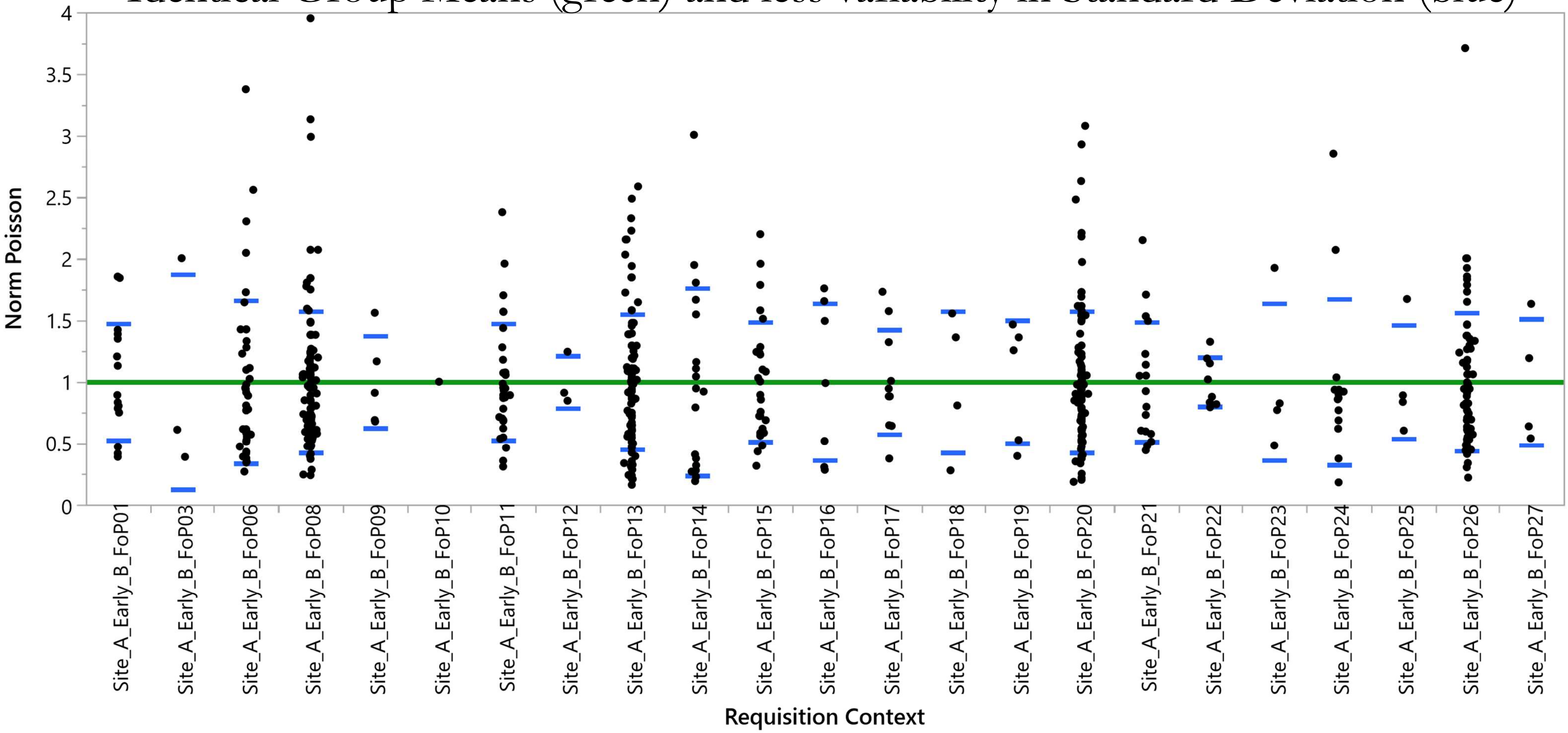
Normalized Poisson Parameter Distribution by Context

- Variance by Context is not dissimilar per O’Brien’s test
- KS test of each Context vs. Remainder (Bulk) showed PValue < 0.05 for only one case out of 78: assumption of a common distribution is reasonable



Normalized Poisson Parameter Distribution Subset by Context

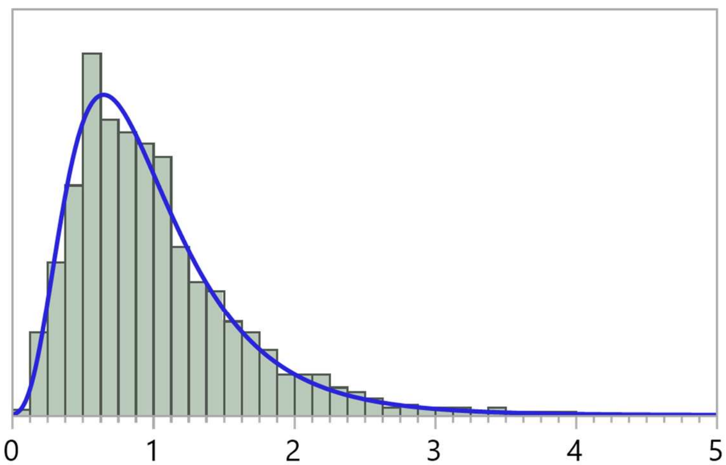
- Subset of Early Career requisitions for Site A
- Identical Group Means (green) and less variability in Standard Deviation (blue)



Normalized Poisson Parameter Distribution and Model Fit

- The complete normalized Poisson parameter data closely resemble a Johnson Sl distribution
- Shapiro-Wilk goodness-of-fit test indicates the Johnson Sl is plausible

Norm Poisson



Summary Statistics

Mean	1
Std Dev	0.58022
Std Err Mean	0.0148239
Upper 95% Mean	1.0290774
Lower 95% Mean	0.9709226
N	1532

Fitted Johnson Sl

Parameter Estimates

Type	Parameter	Estimate
Shape	γ	0.0038163
Shape	δ	2.0173339
Location	θ	-0.127288
Scale	σ	1

Measure	
-2*LogLikelihood	2191.5872
AICc	2197.6029
BIC	2213.5902

Goodness-of-Fit Test

Shapiro-Wilk W Test

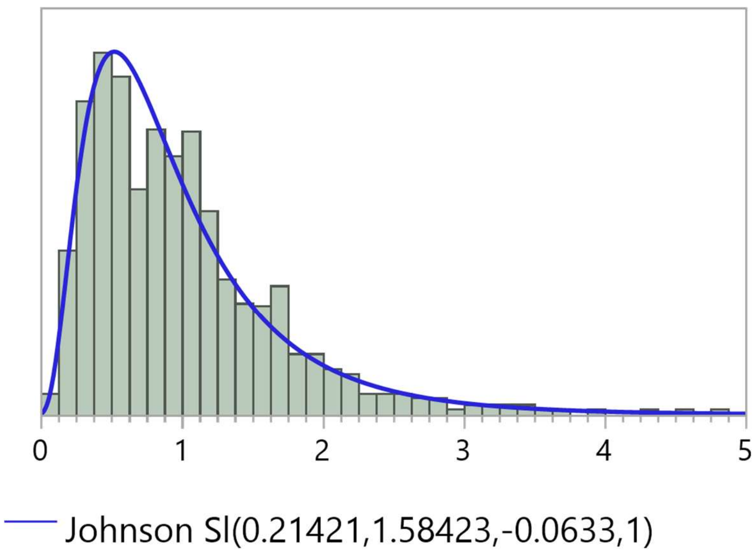
W	Prob<W
0.998494	0.1978

Note: Ho = The data is from the Johnson Sl distribution. Small p-values reject Ho.

Normalized GammaPoisson Lambda Parameter Distribution and Model Fit

- Normalization process for GammaPoisson lambda parameter was nearly identical to process for Poisson parameter

Norm Lambda



Summary Statistics

Mean	1
Std Dev	0.7266599
Std Err Mean	0.0252076
Upper 95% Mean	1.049478
Lower 95% Mean	0.950522
N	831

Fitted Johnson Sl

Parameter Estimates

Type	Parameter	Estimate
Shape	γ	0.2142067
Shape	δ	1.5842313
Location	θ	-0.06333
Scale	σ	1

Measure	
-2*LogLikelihood	1368.8689
AICc	1374.8979
BIC	1389.0368

Goodness-of-Fit Test

Shapiro-Wilk W Test

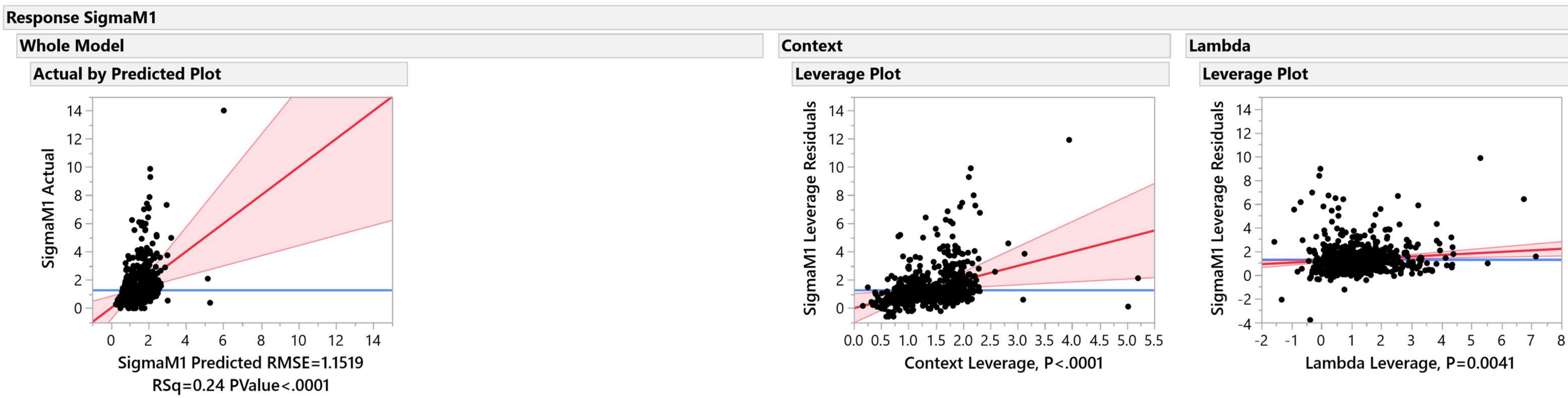
W	Prob<W
0.997156	0.1533

Note: Ho = The data is from the Johnson Sl distribution. Small p-values reject Ho.

- The complete normalized GammaPoisson Lambda parameter data closely resemble a Johnson Sl distribution
- Shapiro-Wilk goodness-of-fit test indicates the Johnson Sl is plausible

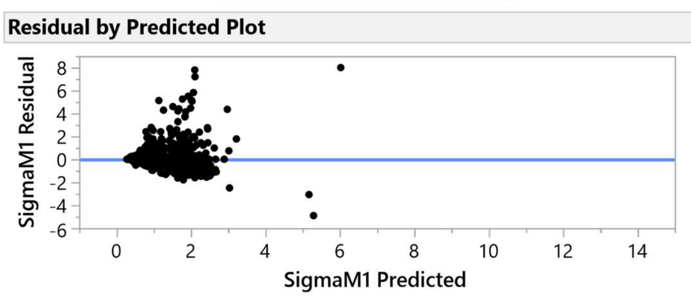
Context and Lambda Based Normalization Function for GammaPoisson SigmaM1 Parameter Distribution

- Normalization of GammaPoisson Sigma parameter was on σ-1 (SigmaM1)
- Lambda used in addition to Context as an input to represent substantive correlation between Lambda and Sigma



Effect Summary

Source	LogWorth	PValue
Context	5.876	0.00000
Lambda	2.385	0.00412

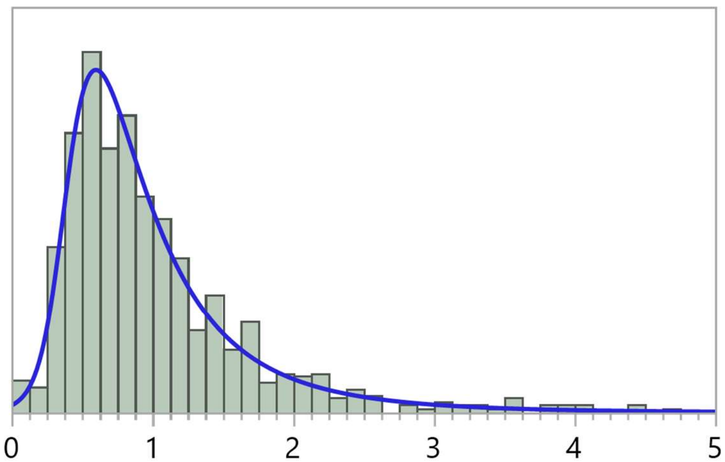


- Variance grows with mean prediction
- Normalization is expected to decrease dispersion of variance across inputs
- Impact of Lambda is less than Context but significant

Normalized GammaPoisson SigmaM1 Parameter Distribution and Model Fit

- Representing parameter as SigmaM1 = σ-1
- The complete normalized GammaPoisson SigmaM1 parameter data closely resemble a Johnson Su distribution
- Shapiro-Wilk goodness-of-fit test indicates the Johnson Su is plausible

Norm SigmaM1



Summary Statistics

Mean	1.0006934
Std Dev	0.6897576
Std Err Mean	0.0239274
Upper 95% Mean	1.0476588
Lower 95% Mean	0.953728
N	831

Fitted Johnson Su

Parameter Estimates

Type	Parameter	Estimate
Shape	γ	-1.603808
Shape	δ	1.2328824
Location	θ	0.3250709
Scale	σ	0.2879808

Measure	
-2*LogLikelihood	1248.2406
AICc	1256.289
BIC	1275.1311

Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.996841	0.1003

Note: Ho = The data is from the Johnson Su distribution. Small p-values reject Ho.

Generation of Synthetic Random Distribution Parameters

- For the data subset best fitting either the Poisson or GammaPoisson distribution
- Generate a linear model for the parameter based on Context and/or other variables
- Normalize the parameter distribution by dividing by the linear model outcome for each datum
- Fit the normalized parameter distribution to a common parametric continuous distribution model
- To generate a synthetic parameter
  - Obtain a random number from the normalized parameter distribution
  - Multiply by the appropriate linear model outcome ("de-normalize")

Evaluation of Synthetic Random Model Parameters

- For the data that best fit the Poisson distribution
- Product of random number from the best fit to the normalized data and the normalization factor
- Synthetic Poisson parameter distribution is indistinguishable from the Poisson parameter distribution for the original data, per KS test

Kolmogorov Smirnov Two-Sample Test

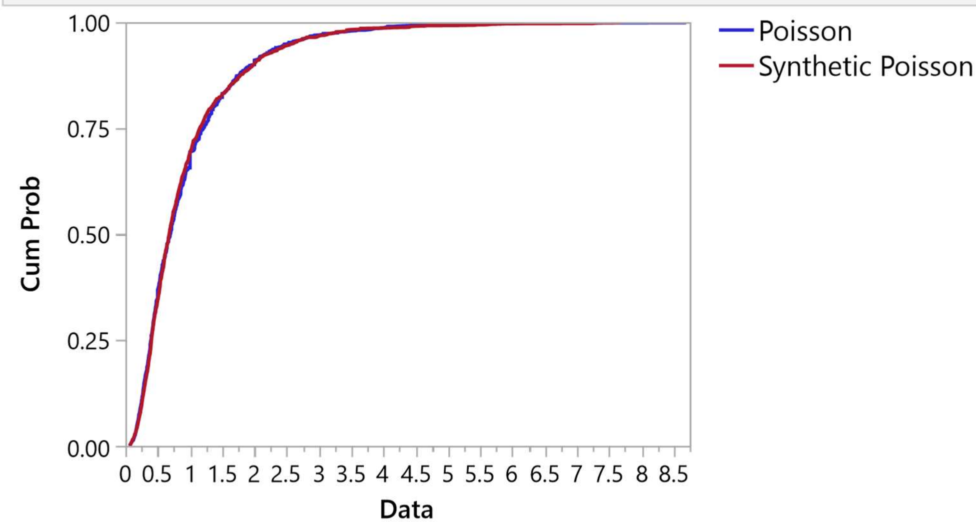
Level	Count	EDF at Maximum	Deviation from Mean at Maximum
Poisson	1532	0.657	-0.677
Synthetic Poisson	1532	0.691	0.677
Total	3064	0.674	

Maximum deviation occurred at observation 346,
value of Data at maximum = 0.997831343640188.

Kolmogorov-Smirnov Asymptotic Test

KS	KSa	D=max|F1-F2|	Prob > D	D+=max(F1-F2)	Prob > D+	D-=max(F2-F1)	Prob > D-
0.0172977	0.9574839	0.0345953	0.3184	0.0261097	0.3519	0.0345953	0.1598

CDF Plot



The GammaPoisson synthetic Lambda and SigmaM1 parameter distributions were also found to be indistinguishable from those for the original data by KS test

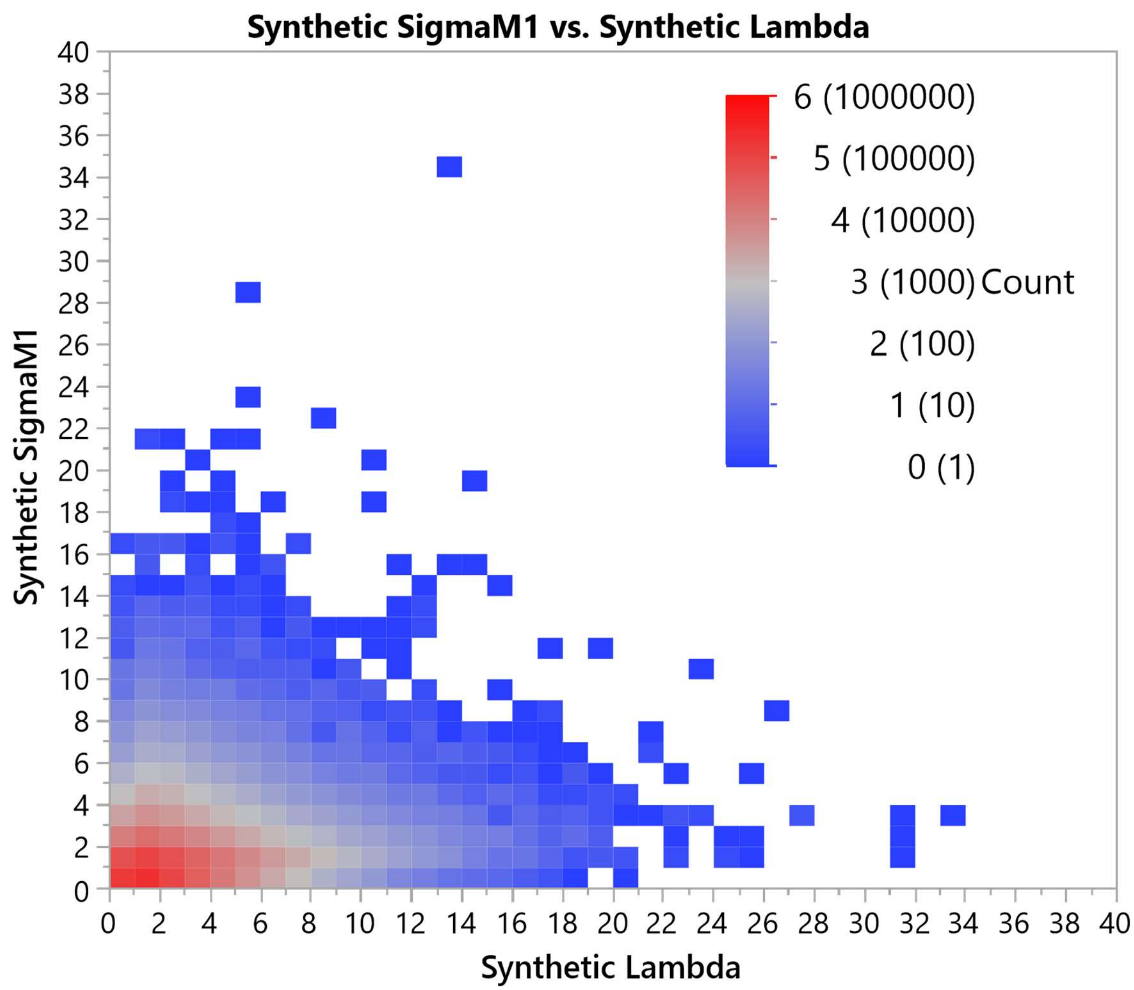
Composite Model Development

- The foregoing analysis demonstrates the generation of plausible synthetic random distribution parameters
- A complete model must address both the Poisson and GammaPoisson possibilities
  - The overall proportion of GammaPoisson best fits within the original data is 35%
  - Modeling the proportion of GammaPoisson best fits by Context does not provide reliable parameters
  - Model for GammaPoisson fraction developed based on Career Stage, Location, and FoP
  - GammaPoisson probabilities tabulated by Context
- Function (algorithm) developed for generating the parameters for a random job requisition with Context as input

Visualization of Synthetic Random Parameter Pairs

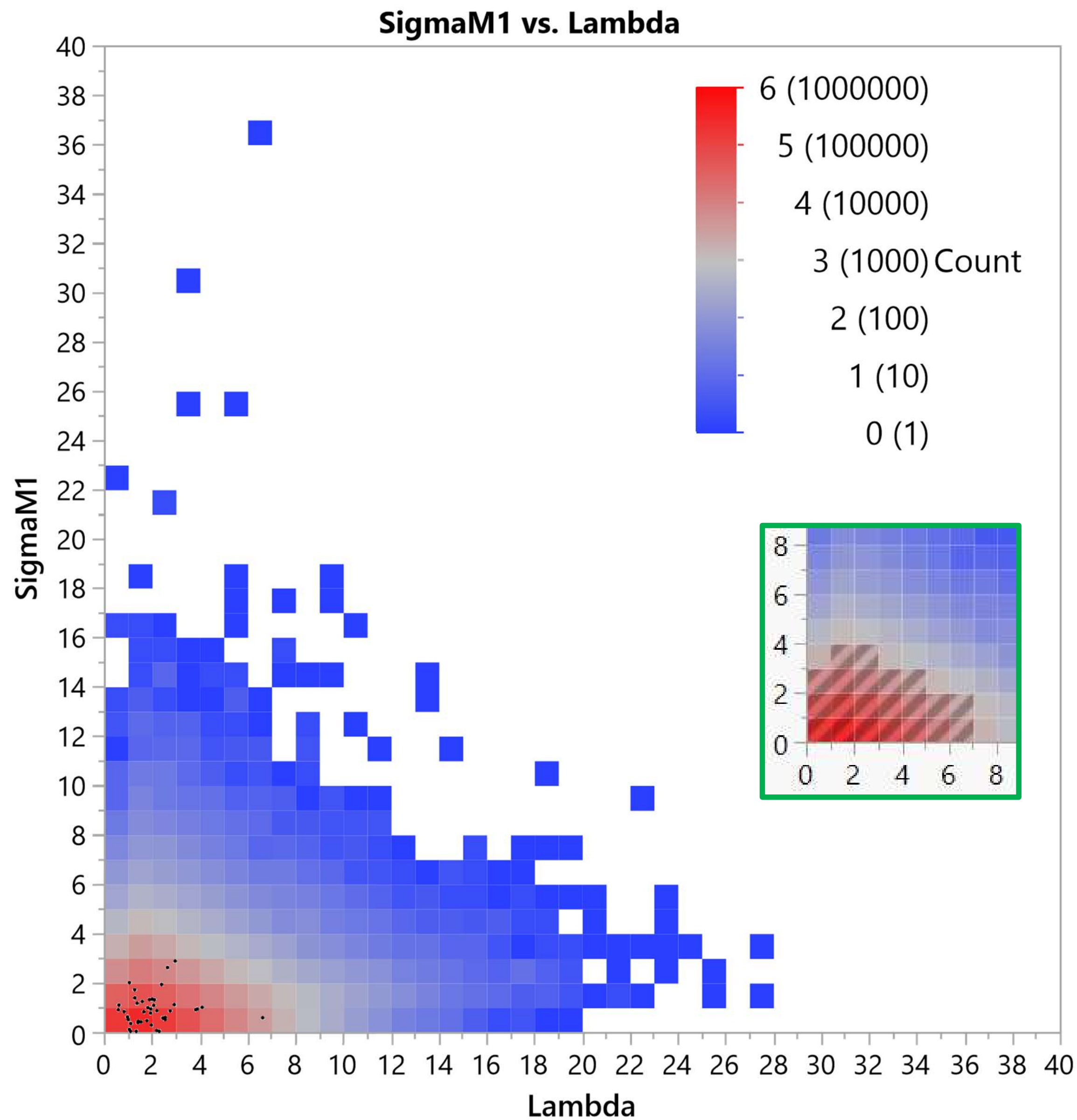
- For the data that best fit the GammaPoisson
- SigmaM1 is correlated to Lambda (0.29)
- Synthetic SigmaM1 is similarly correlated to Synthetic Lambda (0.28)
- Enables plausible visualization of parameter densities using synthetic parameter data (N=1E+6)

Row	Lambda	SigmaM1	Synthetic Lambda	Synthetic SigmaM1
Lambda	1.0000	0.2934	0.3691	0.2659
SigmaM1	0.2934	1.0000	0.2129	0.2188
Synthetic Lambda	0.3691	0.2129	1.0000	0.2784
Synthetic SigmaM1	0.2659	0.2188	0.2784	1.0000



Synthetic Requisition Model Parameters Compared to Parameters for Real Requisitions

- For all broadly visible requisition data
- Poisson outcomes represented as Lambda with SigmaM1 = 0
- Visualization is for a common engineering discipline, early career, located in Site A
- Parameters for real requisitions superimposed on synthetic heatmap (dots, N=43)
- 98% of synthetic density is in the reddish region indicated by crosshatch in inset

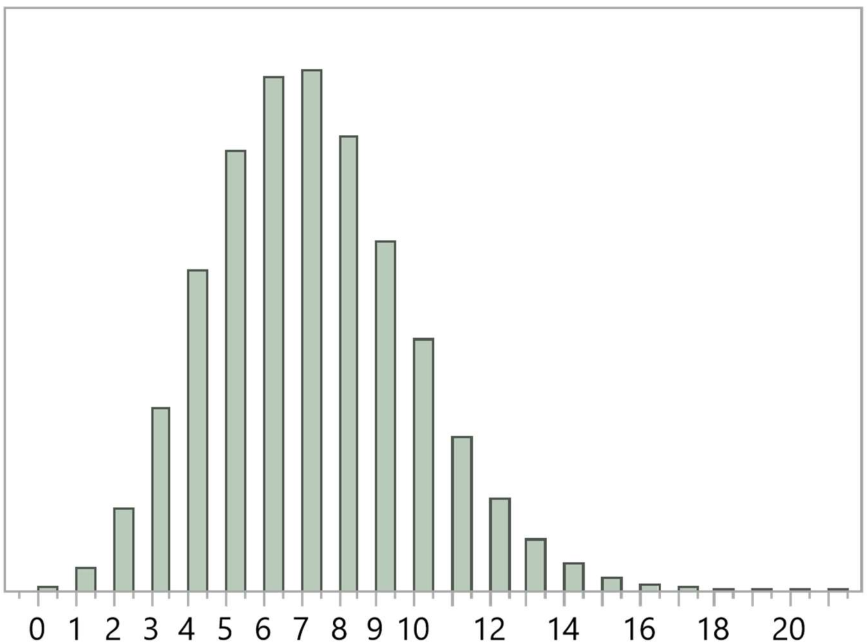


Synthetic parameters offer a reasonable match to actuals for each context while leveraging the complete data set to represent credible extremes

Model Development

Application Count Variability

- The time to obtain enough applications to ensure a reasonably competitive selection for hire should be expected to vary widely
  - Average application rate by requisition for the data shown in this presentation was 1.04/day
  - For λ=7/week, 30% of the time the count will be five or fewer



Remember:

- $Var(Poisson(\lambda)) = \lambda$
- $SD(Poisson(\lambda)) = \sqrt{\lambda}$

The high relative variability of small number statistics can defy expectations based on long-term averages

Pattern Recognition Bias

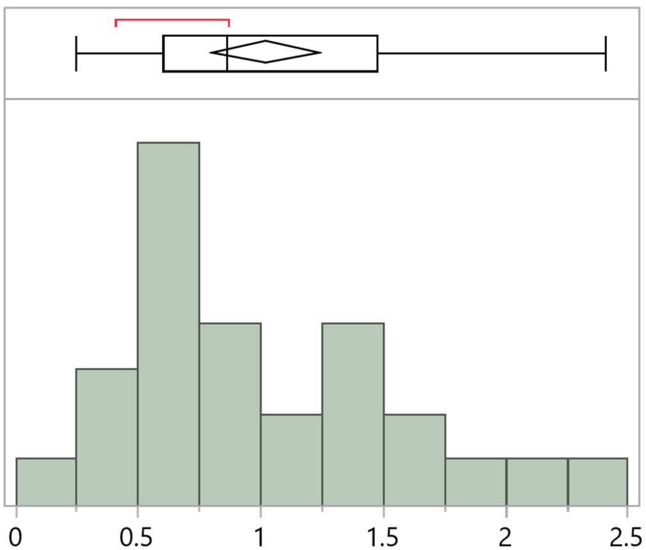
- Any of several cognitive biases characterized by a tendency to imbue meaning to patterns within data that could readily be explained by random action
- The Clustering Illusion – the tendency to erroneously consider the inevitable "streaks" or "clusters" arising in small samples from random distributions to be non-random – is clearly relevant for Poisson distributed data[§]
- The likelihood of a Poisson count generator (λ=7/week) producing a steadily decreasing weekly count over a span of three weeks – {Week1 > Week2 > Week3} – is 12%
- The likelihood of the same generator producing a declining two-week count – {Week1 > Week2} – is 45%
- Pattern recognition bias could lead to perception of scarcity – a finite and small pool of potential respondents
- Consequences may include premature closure of the application window or over-valuation of the applicants in hiring decisions

N.B.: The converse patterns and tendencies are equally likely

§ Clarke, R. D. 1946. "An application of the Poisson distribution." *Journal of the Institute of Actuaries* 72 (3):481.

Application Rate Variability by Field of Practice

- The full range of mean estimated application rate by Field of Practice in our data is nearly an order of magnitude, from 0.245/day to 2.41/day, with a median of 0.858



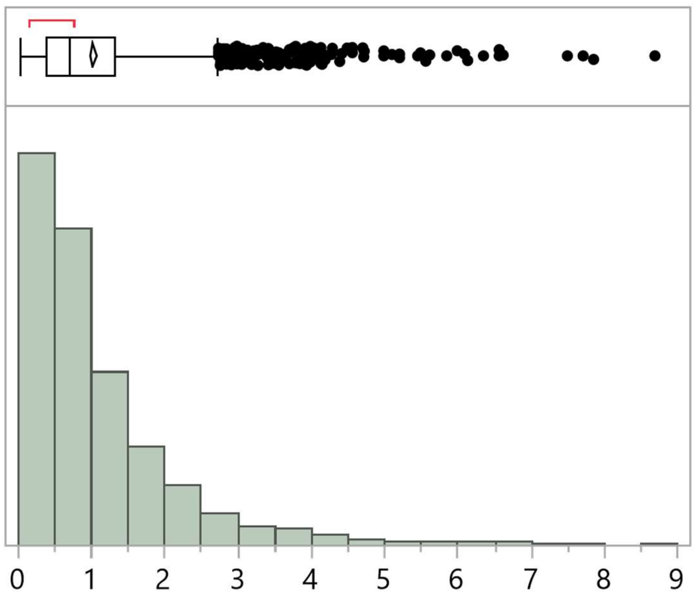
Quantiles		
100.0%	maximum	2.40845440564643
99.5%		2.40845440564643
97.5%		2.40845440564643
90.0%		1.89364340756312
75.0%	quartile	1.47808674516612
50.0%	median	0.85787234471689
25.0%	quartile	0.604214601860075
10.0%		0.433085583029319
2.5%		0.245283016806034
0.5%		0.245283016806034
0.0%	minimum	0.245283016806034

Summary Statistics	
Mean	1.0230732
Std Dev	0.5552113
Std Err Mean	0.1068505
Upper 95% Mean	1.2427075
Lower 95% Mean	0.8034389
N	27

In accordance with intuition and anecdote, some disciplines – *e.g.*, computing fields – are much more challenging to source than others – *e.g.*, technicians

Application Rate Variability by Requisition

- The 95% range of estimated application rates in our data is from 0.147/day to 3.82/day, with a median of 0.714
  - Typical application rates vary from approximately $\frac{1}{5}$ median to 5 times median
- Understanding applicant response as rates and learning more quantitatively how various factors – *e.g.*, field of practice, posting specificity, posting language / framing, advertising, *etc.* – impact those rates may help to improve the effectiveness and efficiency of the talent acquisition business function



Quantiles

100.0%	maximum	8.69230769
99.5%		5.875
97.5%		3.81818181
90.0%		2.21583851
75.0%	quartile	1.33333333
50.0%	median	0.71428572
25.0%	quartile	0.4
10.0%		0.24418605
2.5%		0.14714432
0.5%		0.07983146
0.0%	minimum	0.05194805

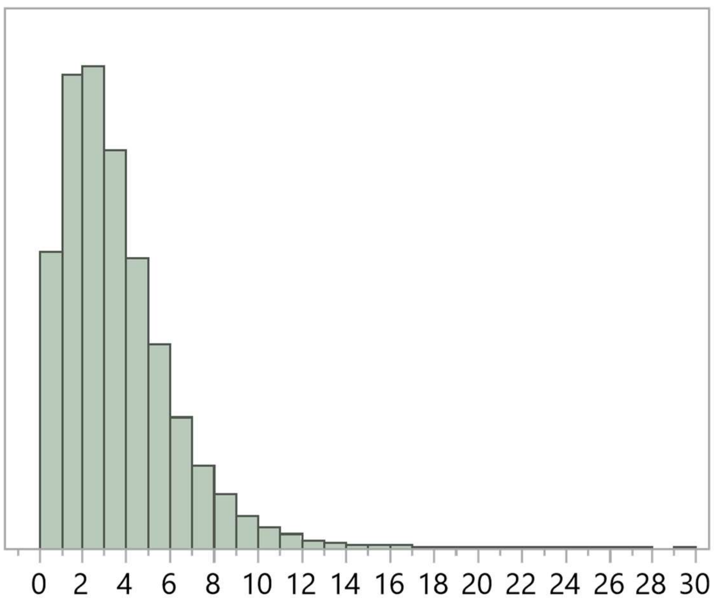
Summary Statistics

Mean	1.0379287
Std Dev	0.9745235
Std Err Mean	0.020001
Upper 95% Mean	1.0771499
Lower 95% Mean	0.9987074
N	2374

Estimated Outcome for a Specific Job Posting

- Simulated distribution of expected total applicants over a two-week period for a job posting for an Established Professional at Site A in discipline FoP09
  - 95% CI ranges from 0 to 9 with median = 2
  - Narrow expectations based on prior hiring experience may be deceptive due to high relative variance

Predicted Two Week Total Applicants



Quantiles

100.0%	maximum	29
99.5%		13
97.5%		9
90.0%		6
75.0%	quartile	4
50.0%	median	2
25.0%	quartile	1
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	3.00647
Std Dev	2.4810904
Std Err Mean	0.0078459
Upper 95% Mean	3.0218479
Lower 95% Mean	2.9910921
N	100000

Limitations of Model and Approach

- As usual, the quality and coverage of the basis data for the model frames the inferences that may sensibly be made
- Infrequently hired fields / rare skill sets – e.g., welding engineers, tribologists – may not be represented if the basis data are collected over a short time frame
  - Representation of unusual (notional outlier) cases hinges on extrapolation through common distribution model
  - Reasonableness of extrapolation depends on capturing a representative range of unusual cases within the basis data
- Conversely, supply, demand, and organizational competitiveness may substantially vary if basis data are collected for a very long time frame
- Modeling approach shown in this presentation does not consider self-cannibalization among applicants
  - If two or more Job Postings are available at the same time within a field, do qualified applicants apply to both or pick one?
  - Model will represent real-world outcome regardless but may not represent scope of opportunity missed

Conclusions

- Employment application response to a job posting tends to be Poisson or GammaPoisson distributed
  - GammaPoisson (coordinated) distribution correlates with high application volume, early career, and non-exempt positions
- Distribution parameters for application response vary substantially
  - Requisition characteristics account for much of this variation – but not all
- Normalization of parameter distributions by requisition characteristics enables fitting to a common profile
- Concise parameter distribution models facilitate generation of synthetic random requisition models
  - Useful for scoping variability of outcomes – expectation setting
- Application arrival models fill an important gap for understanding the complete employee lifecycle
  - Perspective for hiring managers and staffing professionals – counter pattern biases
  - Realistic mechanism for generating applicants in discrete event or agent-based models
  - Method for framing cost per applicant vs. job characteristics, adjustable variables, and external factors