CPSE

2025

EXCELLENCE CONFERENCE

IGNITING INNOVATION

FUELING THE FUTURE OF THE FIRE AND EMERGENCY SERVICE

CPSE
2025 EXCELLENCE CONFERENCE
IGNITING INNOVATION
FUELING THE FUTURE OF THE FIRE AND EMERGENCY SERVICE

Writing a Defensible Outlier Policy

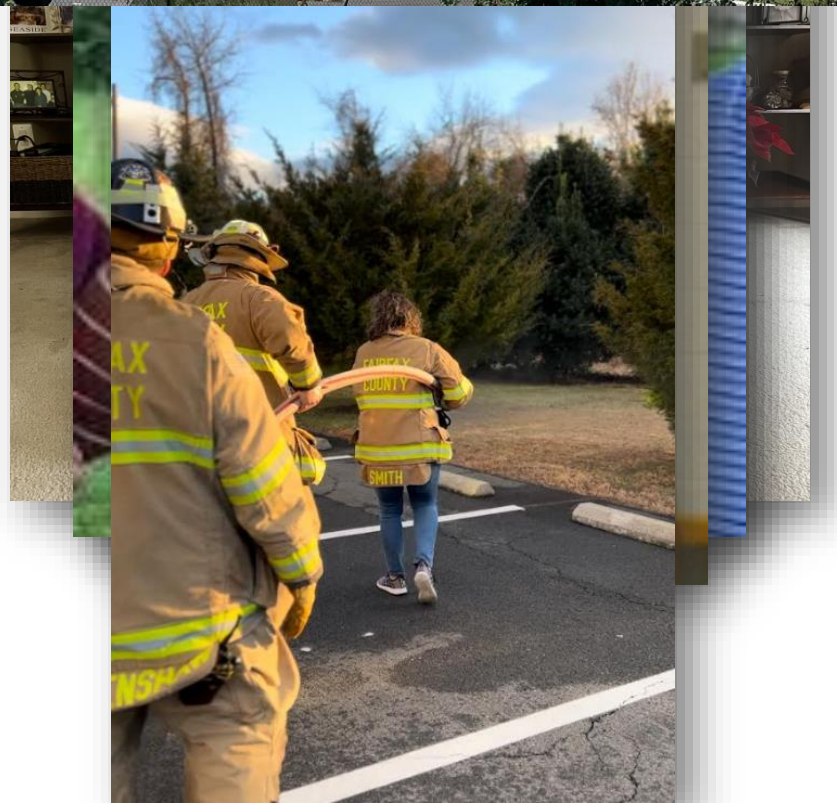
FESSAM Criteria 2C, 2D, 9D

Jessica LeBlanc

Fairfax County Fire & Rescue Department

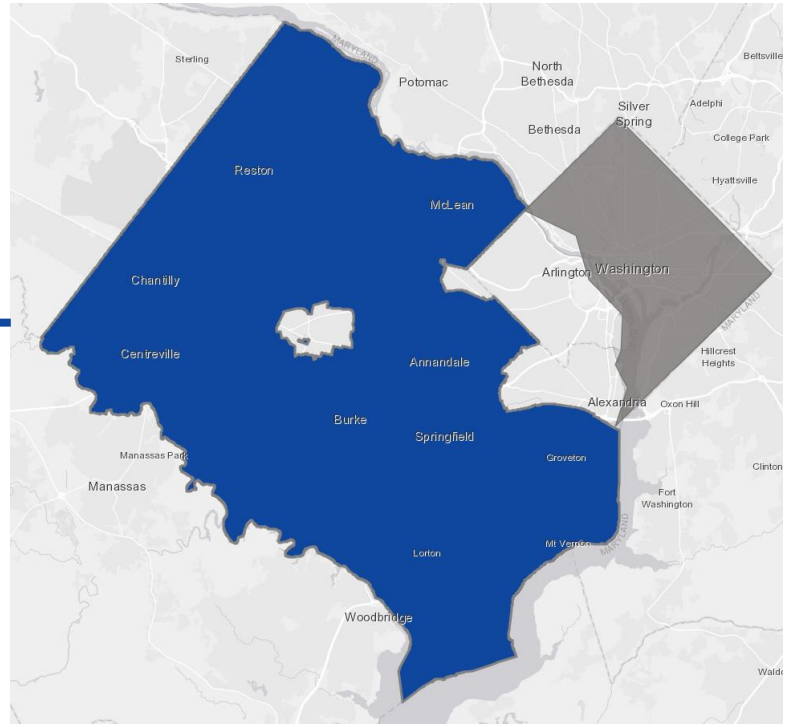
"Gen X Jessica"

- BS in psychology, (almost) MA in criminology
- 21 years in public safety
  - 1 year at NCMEC Exploited Child Unit
  - 14 years as a crime analyst
  - 6 years as a fire/EMS data analyst
- Wanted to be Clarice Starling *(too much school)*
- Wanted to be the Next Food Network Star *(long nights)*
- Wants to be a Taylor Swift groupie *(concept phase)*
- Reality: nerd by day, soccer/dog mom by night
- Best suited behind a computer



Fairfax County, Virginia

Largest jurisdiction in the National Capital Region



1.2 million population
406 square miles

CPSE 2025 EXCELLENCE CONFERENCE

IGNITING INNOVATION

FUELING THE FUTURE OF THE FIRE AND EMERGENCY SERVICE

Fairfax County Fire & Rescue

Personnel

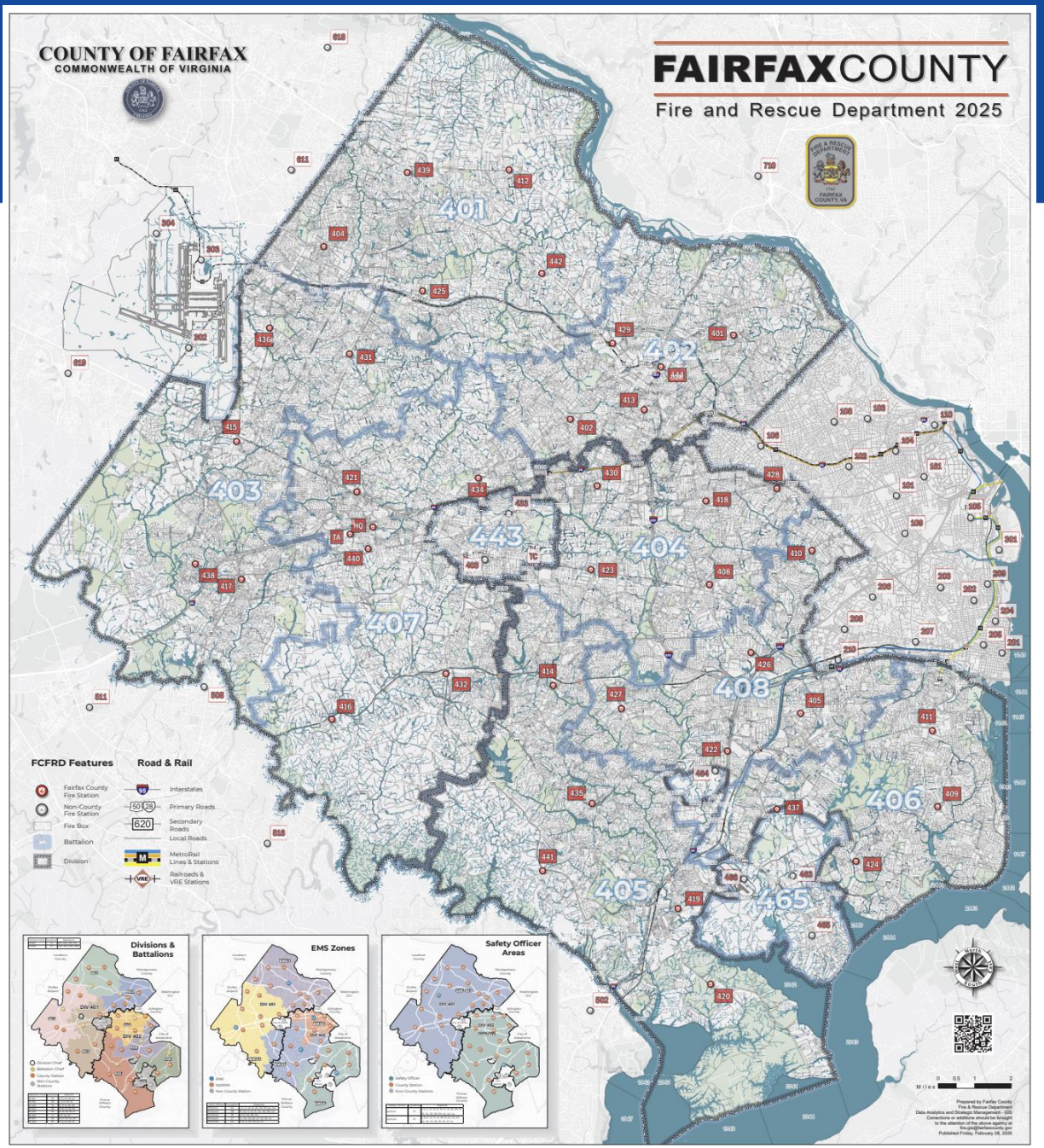
- ~1,400 career uniformed
- ~300 active volunteers
- ~200 civilians

Organization

- 2 divisions
- 8 battalions
- 39 fire stations

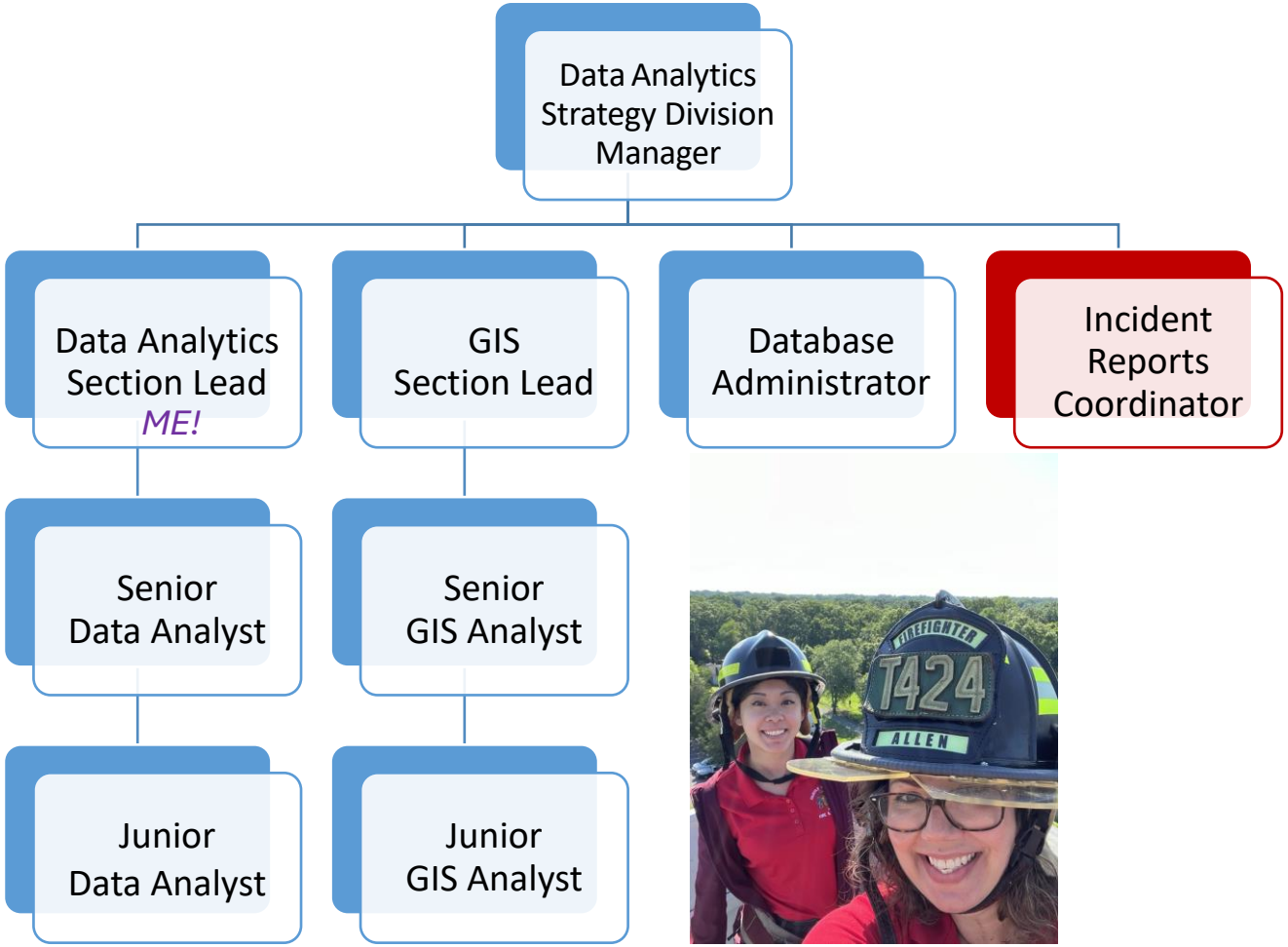
Call Volume Annually

- ~130,000 incidents
- ~250,000 unit responses
- ~70,000 transports



Data Analytics Strategy Management

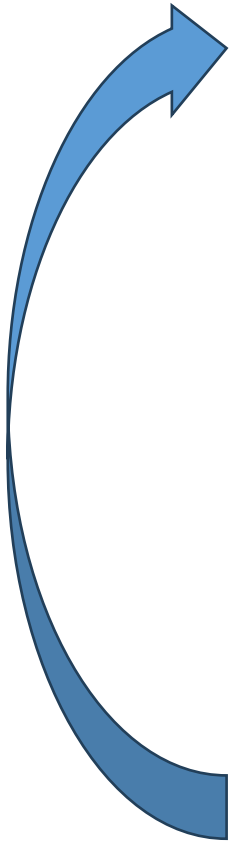
FCFRD



The Plan

- Define “Outlier”
- Visualizing Outliers
- Testing for Confidence
- Standard Operating Procedure Components
- Policy Components

Define "Outlier"



Merriam-Webster Dictionary: Data points significantly beyond the expected or normal range.

Pennsylvania State University: Some observations within a set of data may fall outside the general scope of the other observations. Such observations are called **outliers**.

National Institute of Standards & Technology: An *outlier* is an observation that lies an abnormal distance from other values in a random sample from a population.

Taylor Swift: It's me, hi, I'm the problem, it's me"

FCFRD: Data points significantly beyond the expected or normal range.

Define "Outlier"

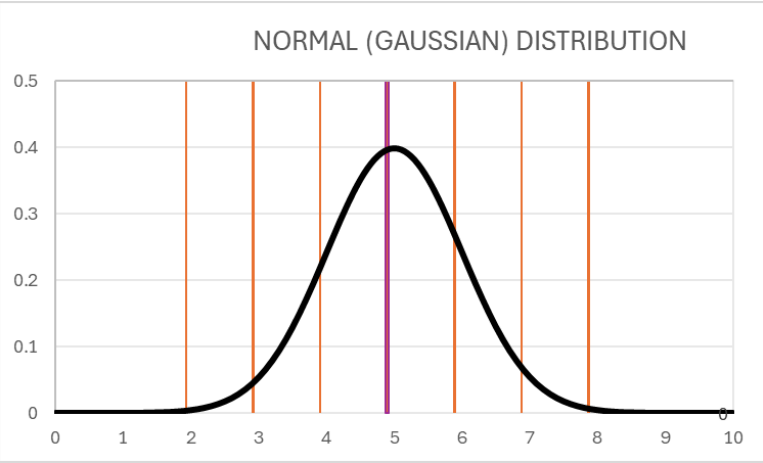
True Outliers

- Actual occurrences that are outside the norm
  - Long response times due to environmental conditions (snow/rain) or mechanical breakdown
  - Long turnouts in stations with sub-optimal design (bunks and bays are far apart)
  - Delays in seeking/requesting mutual aid

Erroneous Outliers

- Created by human error or technological malfunction
  - Timestamps omitted/entered erroneously after CAD downtime
  - Long response time due to failure to mark on-scene
  - Data calculations using the incorrect field
  - Data transfer interruption/failure causes missing values

Visualizing Outliers



Shape (distribution)

Skewness (tails)

Kurtosis (spread)

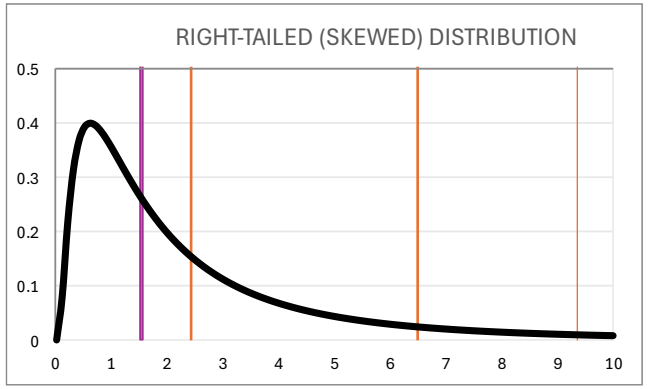
Handy Visualizations (learning style)

Visualizing Outliers

Skewness (tails)

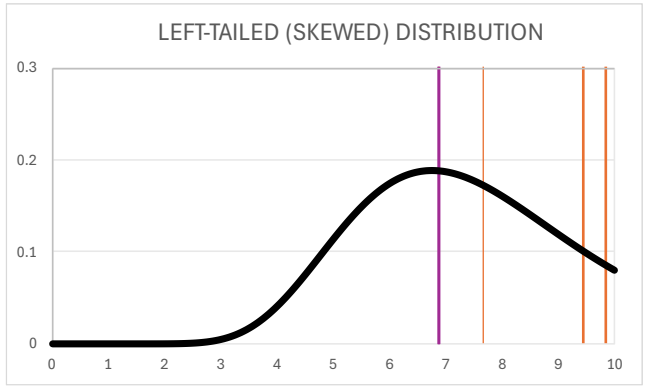
- Call processing
- Turnout

Lots of data clustered at the low end



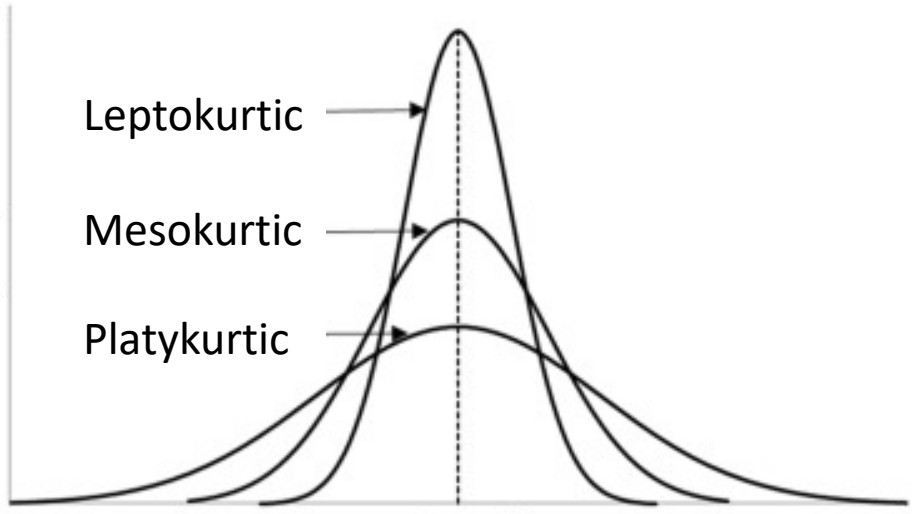
- On-scene time
- Injury recovery

Lots of data clustered at the high end



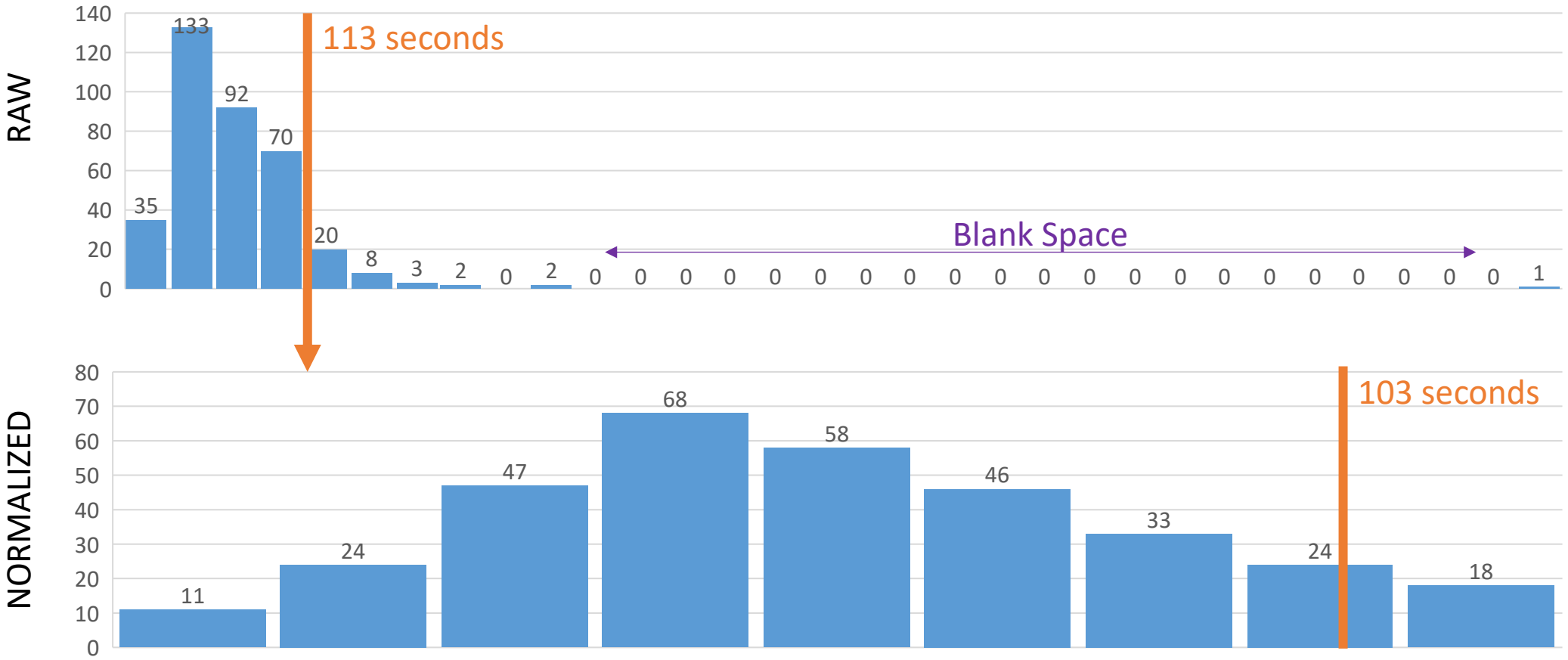
Kurtosis (spread)

- Leptokurtic (+), slender/slight
- Mesokurtic, middle/between (mesoderm)
- Platykurtic (-), flat/broad (platypus)



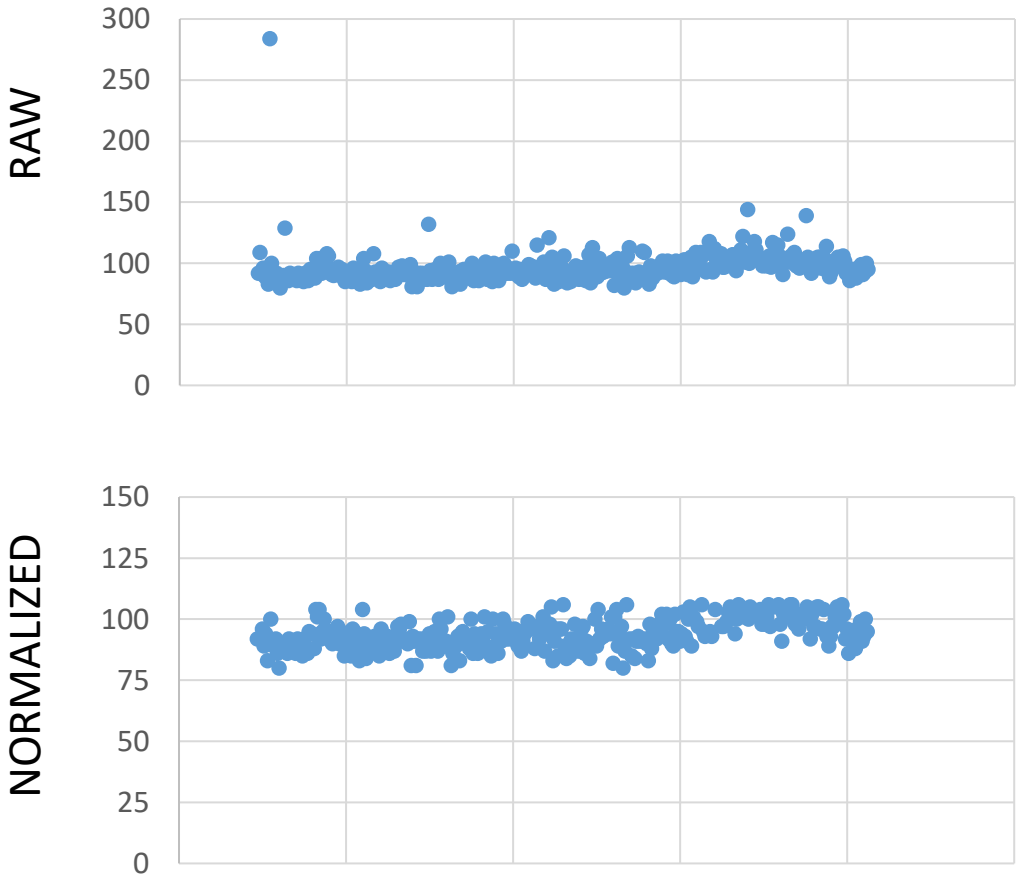
Visualizing Outliers

Histogram

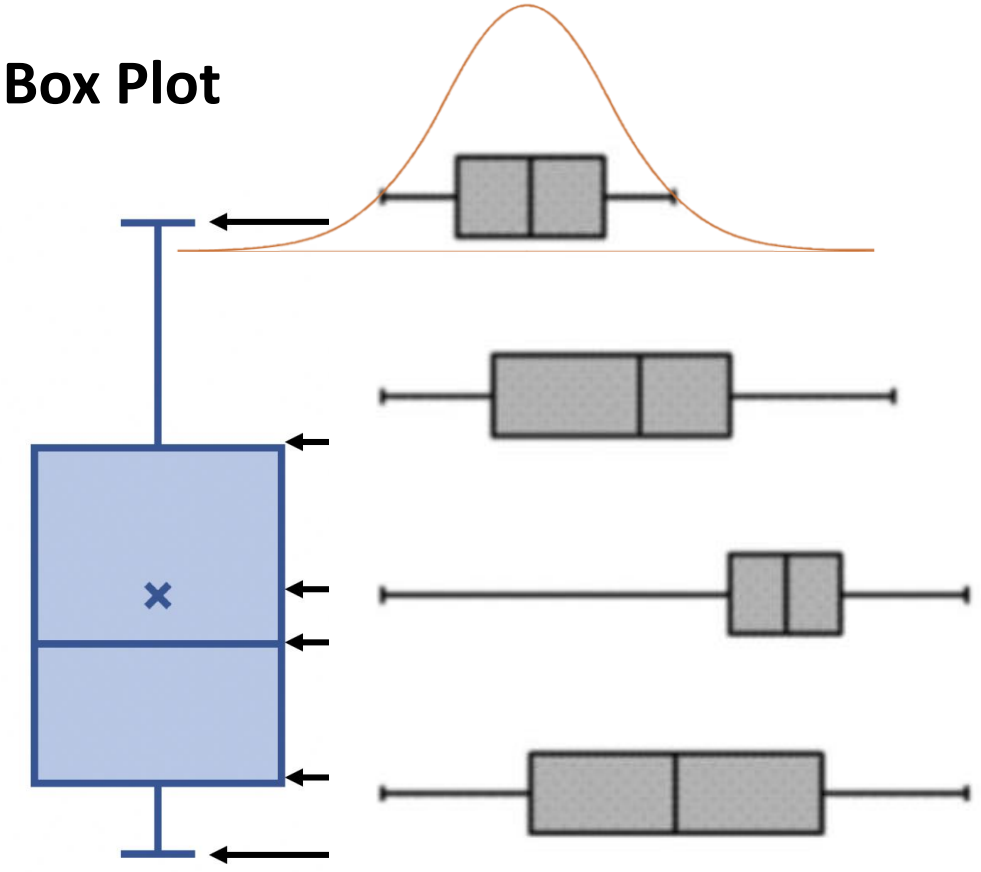


Visualizing Outliers

Scatterplot



Box Plot



Testing for Confidence

How confident are you in the outlier-ness of your outlier?

How do you explain your confidence to others?

What are the limitations of your confidence?

Testing for Confidence

IT'S OPS TALKING:

On average, the daily 90th percentile for call processing is between 93.3 and 94.6 seconds.

Confidence Intervals

- Used when you only have a sample of your data (election results, call volume, etc.) and want to extrapolate (predict) true figures for the whole data set
- Used when you're comparing two samples to each other (before and after or KC Chiefs points with/without TayTay in attendance, etc.)
- You need (easy to grab in Excel)
  - Average
    - =AVERAGE(data range)
  - Standard deviation
    - =STDEV.S(data range)
  - Sample size
    - =COUNT(data range)
  - z value for your confidence level
    - How right do you want to be?
    - 95% = 1.96
    - 99% = 2.58

A	B
Date	All 90th
07/01/19	92
07/02/19	109
07/03/19	92
07/04/19	96
07/05/19	89
07/06/19	94
07/07/19	83
07/08/19	284
07/09/19	100
07/10/19	90
07/11/19	89

A	B	C	D	E	F	G
Date	All 90th	No Outliers 90th			All 90th	No Outliers 90th
07/01/19	92	92		average	96	94
07/02/19	109	-		stdev	13.21549951	6.021055439
07/03/19	92	92		count	366	330
07/04/19	96	96		median	94	93
07/05/19	89	89		mode	92	92
07/06/19	94	94		90th	106	103
07/07/19	83	83				
07/08/19	284	-		high CI	97.77197183	94.55872937
07/09/19	100	100		low CI	95.06409374	93.25945245
07/10/19	90	90				
07/11/19	89	89				

Testing for Confidence

H_0: The full call processing dataset is NOT different from the normalized dataset

H_1: The full call processing dataset is significantly different from the normalized dataset

Z value (how far from average)

- How far "in the Red" is your dataset's average?
- Tells you how extreme a statistic is
- CAUTION: if data is not normally distributed, this value can be misleading (which is why you should take steps to normalize your data first)

T test (probability under H_0)

- If a p value is less than .05, we are 95% confident that the average 90th call processing in the normalized dataset is significantly different than the full dataset
  - *reject the null hypothesis*
- If a p value is more than .05, we are confident that any difference in the average 90th call processing is due to chance – OR – that there is only a 5% chance we'd see an extreme value
  - *cannot reject null hypothesis*

Standard Operating Procedure (SOP) Components

- Purpose
- Data Governance Committee
- Data Analytics Strategy Management Division
- Analysis Standards
- Statistical Definitions

SOP Components

- **Purpose**
- Data Governance Committee
- Data Analytics Strategy Management Division
- Analysis Standards
- Statistical Definitions

FAIRFAX COUNTY FIRE AND RESCUE DEPARTMENT STANDARD OPERATING PROCEDURE		
	SUBJECT: Data Governance and Analysis Standards	**SOP** **ORIGINATOR BUREAU:** Office of the Fire Chief **DIVISION/SECTION:** Data Analytics Strategy Management Division **PAGE** 1 **OF** 2
	CATEGORY: Administration	**SUBCATEGORY:** Reports and Reporting Procedures
	APPROVED BY: Fire Chief John S. Butler **FIRE AND RESCUE DEPARTMENT**	**EFFECTIVE DATE:** March 1, 2025 **REVISION DATE:** N/A
	FORMS REQUIRED: **NOTE:** Current forms are located on the department's SharePoint	

PURPOSE:

To establish and define responsibilities for data governance and analysis standards for the Fairfax County Fire & Rescue Department (FRD).

SOP Components

- Purpose
- Data Governance Committee
- Data Analytics Strategy Management Division
- Analysis Standards
- Statistical Definitions

Data Quality Management:

1. How strongly do you agree that data within the department is consistently accurate and reliable?
   - 1 (Strongly Disagree) 2 3 4 5 (Strongly Agree)
2. To what extent are you satisfied with the completeness of the data available for your job role?
   - 1 (Not at all) 2 3 4 5 (Completely)
3. How frequently do you encounter data inconsistencies or errors that impact your work?
   - 1 (Very Frequently) 2 3 4 5 (Never)
4. To what degree do you feel that data quality issues are promptly addressed and resolved within our organization?
   - 1 (Not addressed at all) 2 3 4 5 (Promptly addressed and resolved)

How is success measured?

SOP Components

- Purpose
- Data Governance Committee
- **Data Analytics Strategy Management Division**
- Analysis Standards
- Statistical Definitions

II. Data Analytics Strategy Management Division (DASM)

a. The DASM has been established as the FRD's primary decision support team. Any analysis for the FRD would benefit from consultation or collaboration with DASM personnel.

b. The composition of the DASM's leadership ("Leads") shall, at a minimum, include:

- DASM Director
- DASM Data Analytics Section Lead
- DASM Geographic Information Systems Section Lead
- DASM Database Administrator
- FRD's Health Insurance Portability and Accessibility Act (HIPAA) Privacy Officer

c. In collaboration with the DGC, the DASM Leads shall define and create policy for procedures, best practices, and minimum standards for analysis in the FRD.

What is the DASM's role in data governance and analysis standards?

SOP Components

- Purpose
- Data Governance Committee
- Data Analytics Strategy Management Division
- Analysis Standards
- Statistical Definitions

III. Analysis Standards

a. All analysis performed within the FRD must be viable and permissible to support business decisions. Viable analysis has a clear and valid methodology with actionable results. Permissible analysis complies with organizational policies, ethical guidelines, and regulatory requirements. As a minimum standard, a viable and permissible analysis conducted within FRD meets the following widely accepted criteria:

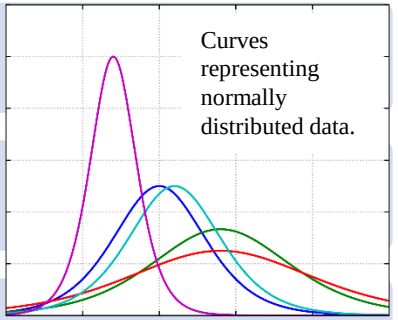
1) clearly defines the problem(s) or research question(s) and includes the specific goals of the analysis,
2) considers relevant prior knowledge and conclusions from analysis conducted either internally within the FRD or externally by other industry stakeholders,
3) cites specific sources of prior research,
4) explicitly describes data collection methods, including sources, and any known data limitations,
5) justifies data transformation methods, detailing the reasoning behind modifications that may influence results or interpretations, ensure data is **CLEAN**
6) identifies and explains analytical techniques used, ensuring methodological transparency,
7) distinguishes analytical observations from business interpretations to reduce bias,
8) provides sound recommendations and/or draws clear evidence-based conclusions,
9) is defensible during a review from subject matter experts, demonstrating rigor and accuracy, and
10) effectively documents and communicates results in a clear, actionable manner.

b. Aggregate, numerical data shall be reported at the median and ninetieth percentile. Average, or mean, shall only be reported for data which is normally distributed and shall be accompanied by the median and ninetieth percentile of the same dataset.

SOP Components

- Purpose
- Data Governance Committee
- Data Analytics Strategy Management Division
- Analysis Standards
- Statistical Definitions

Normal Distribution



Outliers

True

Error-based

90th Percentile

Arithmetic Mean

Median

Discrete vs Continuous Data Point

Policy Components

- Define “outlier”
- Data Generation Processes
- Issues Effecting Data Quality and Integrity
- Detecting Outliers
- Handling Outliers
- Review

Policy Components

- **Define "outlier"**
- Data Generation Processes
- Issues Effecting Data Quality and Integrity
- Detecting Outliers
- Handling Outliers
- Review

Clear, plain language

be Delicate

1. Policy Purpose
   In compliance with Standard Operating Procedure X this policy establishes how outlier data are defined, identified, and handled.

2. Policy Statement
   Standardizing the way in which outlier data are defined, identified, and handled mitigates the potential negative impact resulting from the erroneous or improper use of data in analysis including, but not limited to, producing misleading output and drawing inappropriate or inaccurate conclusions.

4. Definitions
   a. Outliers: Data points significantly beyond the expected or normal range.[i] Unless specified otherwise, an outlier data point is one that is greater than or less than three standard deviations above or below the arithmetic mean of the dataset.
      i. True: Real data representing an extenuating circumstance or occurrence beyond the expected range.
      ii. Error-based: Data caused by human or technical errors which is not reflective of the actual occurrence or phenomenon.

Policy Components

- Define "outlier"
- **Data Generation Processes**
- Issues Effecting Data Quality and Integrity
- Detecting Outliers
- Handling Outliers
- Review

Clear, plain language

3. Known Issues Impacting Data Quality and Integrity

Understanding how data is generated informs how outlier data are defined. Through this understanding, a methodical investigation and analysis to determine whether outlier data points are true or erroneous, or whether they capture a meaningful signal versus irrelevant noise, can be conducted.

Documenting changes to the data generation process provides context when trends in the data shift, helping to understand whether a trend is due to actual change in an occurrence or change to a procedure or business process (e.g. extended call processing times during COVID were caused by the addition of screening questions, not by a change in call taker aptitude, call volume, or emergency type caused by the pandemic). Procedure or business process-based outliers are more easily detected when error-based outliers are removed from the data. Awareness of the impact caused by changes to procedure and business processes informs both future policy considerations and analyses forecasting outcomes based on historical data. Changes to procedural or business processes should be documented via Informational Bulletin for historical record keeping.

Policy Components

- Define "outlier"
- Data Generation Processes
- **Issues Effecting Data Quality and Integrity**
  This is why we can't have nice things
- Detecting Outliers
- Handling Outliers
- Review

Clear, plain language

Data Originating with the Department of Public Safety Communications (DPSC)
Response time data typically originates with the call taker and dispatcher's handling of emergency and non-emergency calls from the public to the DPSC. Although the FRD is accountable for the agency's overall response time (call received to first unit arrived), the FRD does not have direct influence over call processing time (call received to first unit dispatched), which is the result of DPSC's internal business and operating policy and procedures. Known issues affecting data originating with DPSC are identified below.

a. On joint calls with the police department, response time analysis must consider the timestamp when the FRD event entry begins thus capturing the call processing for only the FRD's response. This is particularly critical in cases where the police department is dispatched first.
b. On joint calls in Fairfax County with automatic aid partner response, when an agency other than the FRD arrives first, that agency's response is slowed by a known delay in the CAD2CAD process of sending the request for aid, the agency acknowledging receipt of the request, then permitting the dispatch of their units. This processing delay is typically 30 seconds or less.

Policy Components

- Define "outlier"
- Data Generation Processes
- Issues Effecting Data Quality and Integrity
- **Detecting Outliers**
- Handling Outliers
- Review

Clear, plain language

5. Identifying Outliers

Data tables will be maintained and loaded daily in the FRD's data warehouse capturing each call dispatched in Fairfax County where the dispatched event type was categorized as an emergency medical service or fire emergency. During the process of loading this data, the outlier values for the preceding 90 days and for the entire dataset will be calculated. Columns denoting whether the real value is an outlier will display a 1 and a 0 if not. These data tables will feed a Power BI report which summarizes and trends the data over time.

Additional data tables will be maintained and loaded daily in the FRD's data warehouse capturing each call where an FRD unit was dispatched, both inside and outside Fairfax County, where the dispatched event type was categorized as an emergency medical service or fire emergency. During the process of loading this data, the outlier values for the preceding 90 days and for the entire dataset will be calculated. Columns denoting whether the real value is an outlier will display a 1 and a 0 if not. These data tables will feed a Power BI report which summarizes and trends the data over time. This dataset will allow for the isolation and analysis of calls requiring CAD2CAD transmissions which incur a processing delay.

Exclusions: Calls dispatched in Fairfax County as a public service where emergency response is not required. Included in these call types are Elevator Incident (ELEV), Lock Out (LOCKF), and PUBLIC SERVICE: LIFT ASSIST (LIFT), etc.

Policy Components

- Define "outlier"
- Data Generation Processes
- Issues Effecting Data Quality and Integrity
- Detecting Outliers
- **Handling Outliers**
- Review

6. Handling Outliers

   All datasets should be examined for outliers prior to analysis or reporting. Outliers may be identified through manual or programmatic exploration. Outliers should either be retained, transformed, or excluded.

   For response times, outliers will be handled using one or more of the following methods, listed in order of application. In accordance with the Analysis Standards outlined in SOP X, the handling of any outliers should be documented in the methodology accompanying the analysis and/or its derivative product(s).

   Call processing

   a. *Retain* for response times analysis any outliers identified as True in *4.a.i*.
   b. Continuous collaboration with DPSC to minimize the frequency with which erroneous times are recorded due to known issues impacting data quality and integrity.
      i. On a weekly basis, investigate high outliers to determine whether the call processing time was long due to extenuating circumstances of the call or due to human behavior or technical glitches which may be corrected. When appropriate, changes to the timestamps in the source data may be made to transform reflect true call processing.
   c. *Exclude* from response times analysis any incident where the call processing time is deemed an outlier as defined in *5*.

Policy Components

- Define “outlier”
- Data Generation Processes
- Issues Effecting Data Quality and Integrity
- Detecting Outliers
- Handling Outliers
- **Review**

7. Policy Review
   At a minimum of twice annually, this policy should be reviewed by the DGC to determine its efficacy. Any additions or edits to the policy should be at the direction of the DGC, approved via memo through the Office of the Fire Chief, and announced to the FRD via Informational Bulletin.

Review schedule

- Frequency
- Scope

Authority to modify

- Method
- Chain of command

CPSE

2025 EXCELLENCE CONFERENCE

AFESA

ASSOCIATION OF FIRE AND EMERGENCY SERVICES ANALYSTS

INTEREST FORM

Interest Form

LinkedIn

Questions / Airing of the Grievances (Seinfeld first aired in 1989)

My Contact Information

FRD DASM GitHub

Connect with the DASM

Peer Assessor Continuing Education