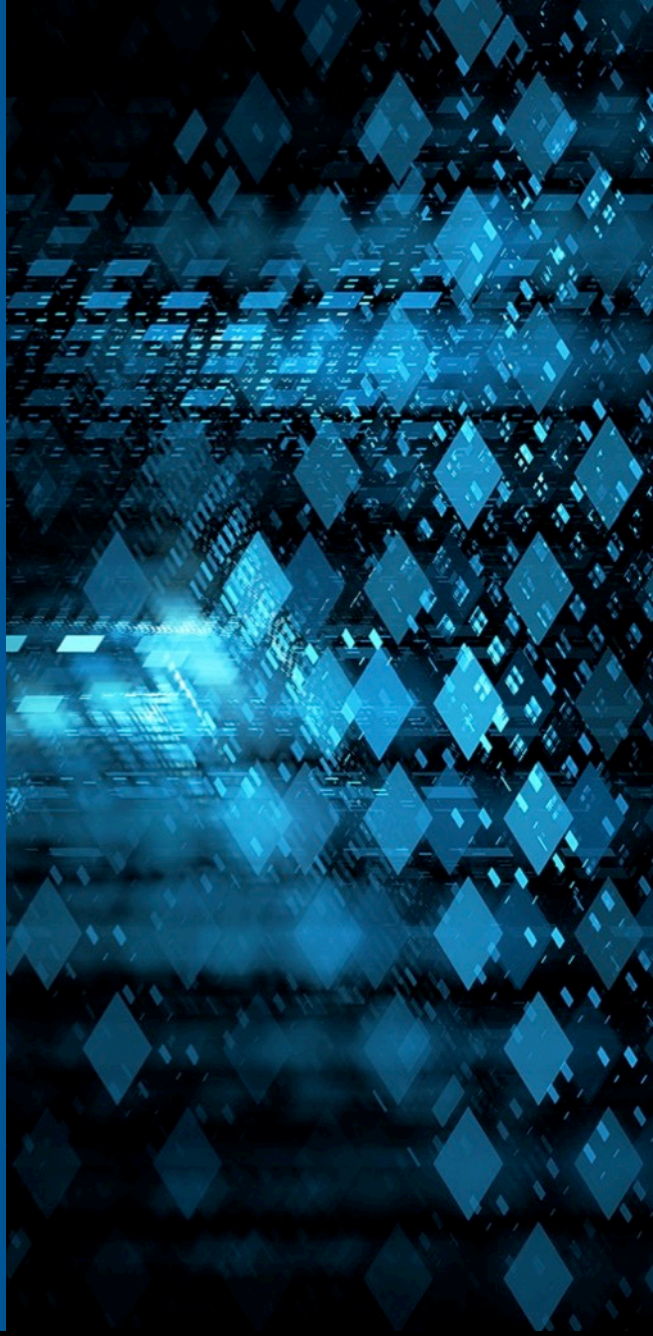
FloCon Analysis Workshop

Angela Horneman

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213



This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM17-0522

Workshop Roadmap

Analysis

Environmental Context

Gathering Data

Microanalysis

Macroanalysis

Reporting and Feedback

Analytic Acumen

Analysis

What is analysis? Formal Definitions

This process as a method of studying the nature of something or of determining its essential features and their relations (*dictionary.com*)

The process of breaking up a concept, proposition, linguistic complex, or fact into its simple or ultimate constituents (*Cambridge Dictionary of Philosophy, 2nd ed., 1999, ed. Robert Audi*)

The isolation of what is more elementary from what is more complex by whatever method (*Dictionary of Philosophy and Psychology, 1925, ed. James Mark Baldwin, Vol. I*)

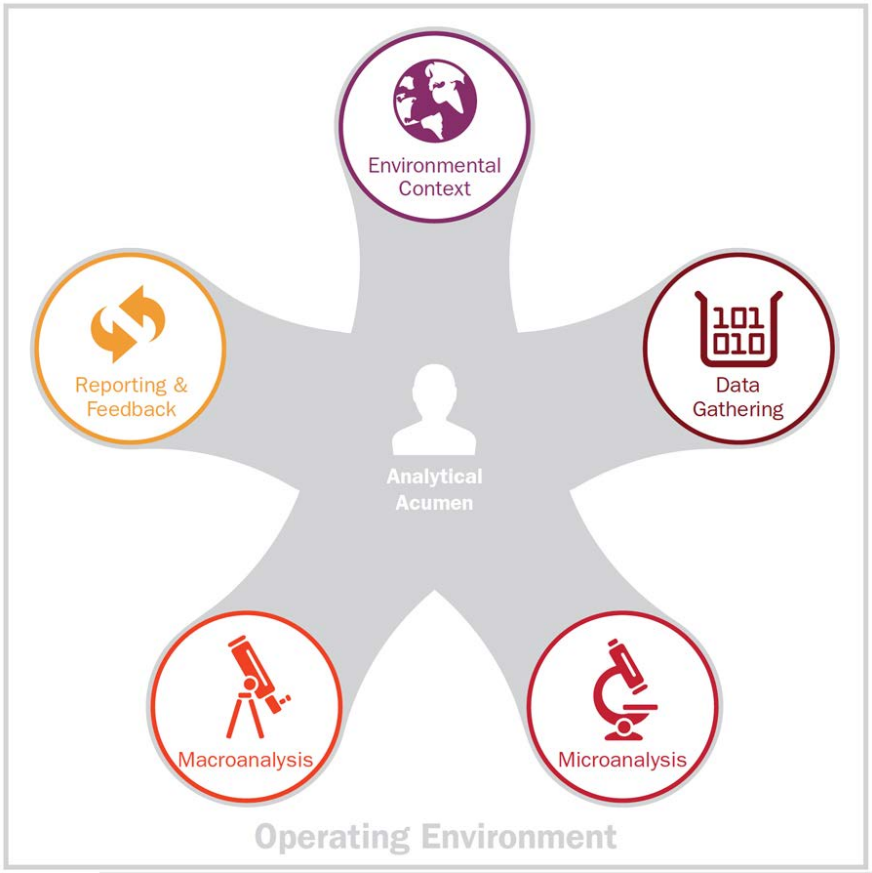
What is analysis? Intelligence Analysis

Intelligence analysis is the application of individual and collective cognitive methods to weigh data and test hypotheses within a secret socio-cultural context. (*CIA, Center for the Study of Intelligence. https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/analytic-culture-in-the-u-s-intelligence-community/chapter_1.htm*)

Intelligence analysis is the process by which the information collected about an enemy is used to answer tactical questions about current operations or to predict future behavior. (*RAND Corp. https://www.rand.org/topics/intelligence-analysis.html*)

What is analysis? More Generally

The process of using data, context, analytical techniques and critical thinking skills to answer a question or test a hypothesis and make the results usable.



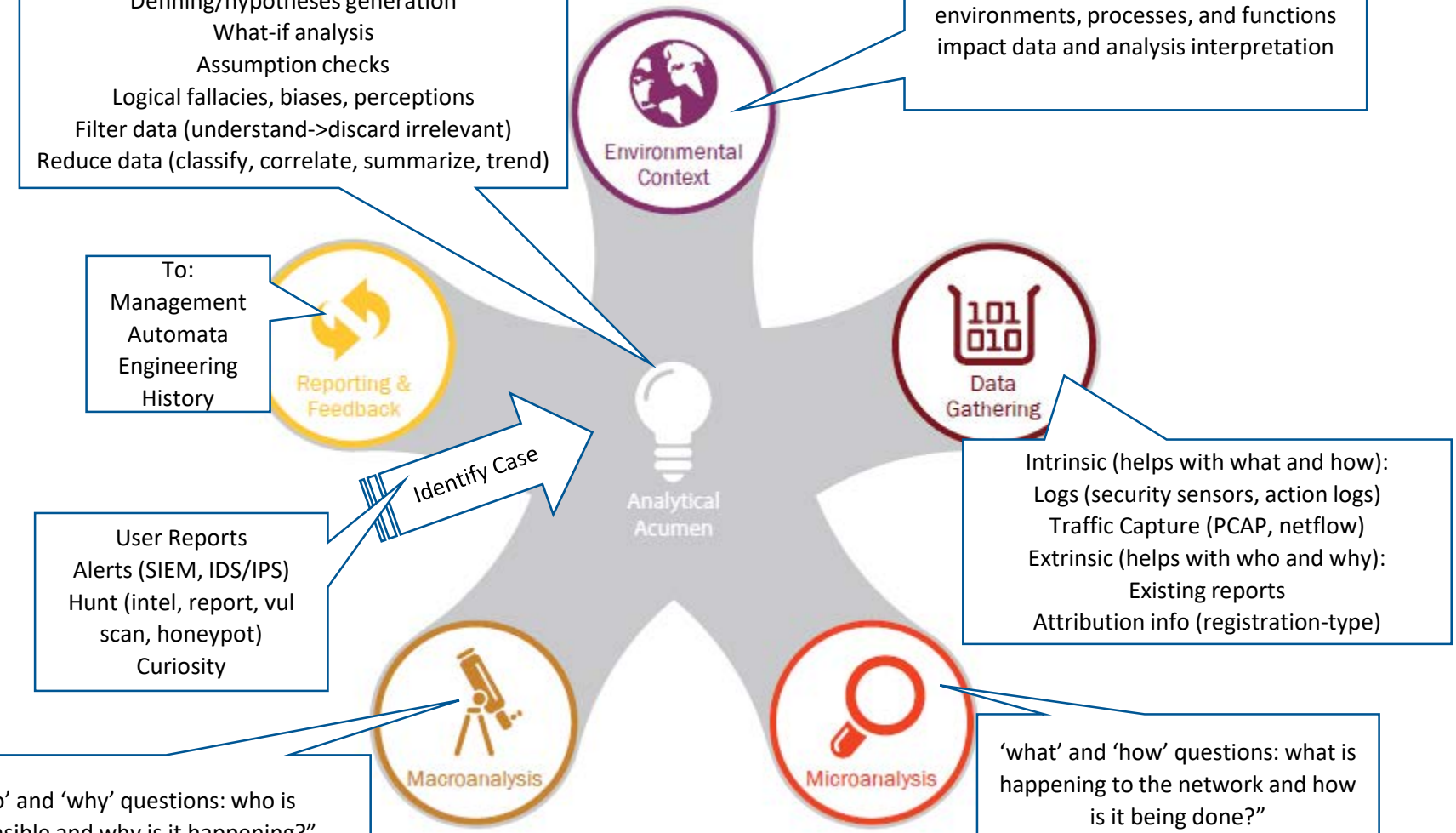
Why this workshop?

Think about how we answer questions and test hypotheses.

Analytical skills and techniques (i.e., critical thinking skills) can improve results.

Understand how data and context impact the process.

What are we covering? Analysis Framework



What are we covering?

Environmental Context

- why context matters
- knowing a cyber environment
- when context is missing

Gathering Data

- What do you need?
- Where do you get it?
- What if you can't find it?

Microanalysis

- finding what and how
  - traffic, logs, and basic statistics

Macroanalysis

- finding who and why
  - intelligence sources and basic fusion

Reporting and Feedback

- sharing is important

Analytic Acumen

- cognitive biases
- hypotheses
- what-if analysis
- Do you really need more data?
- Satisficing is not sufficient.
- Is cyber analysis a puzzle?
- Facts require interpretations; they don't speak for themselves.

Getting Started

Where to begin?

At the end!

To start, answer questions like:

- What do I want to find out?
- What do I think is happening?

Then use the answers to make hypotheses.

Hypotheses

A clear, specific statement of what you are trying to prove or disprove. The statement must be testable.

Bad Example

Hypothesis: This email is bad.

Discussion: What is “bad”? This is not clear, specific, or objective.

Good Example

Hypothesis: This email is trying to get the recipient to install malware through an executable attachment.

Discussion: This is not comprehensive as to all we may consider bad, but is testable.

Comprehensiveness

Favor simple statements over complex ones.

It is fine to have multiple sub-hypotheses that you need to choose between or decide if they are all wrong or all right.

Example:

This email is bad because it

- is trying to get the recipient to install software through an executable attachment
- links to a known malicious URL
- requests the user to provide sensitive data to an unauthorized or spoofed entity

I have my hypotheses, now what?

Make a plan.

Answer questions like:

- Can I directly prove my hypotheses?
- What information do I need to prove my hypotheses wrong?
- What information do I need to support my hypotheses?
- How do I get that information?
- What circumstances and environment factors (context) will influence what I see and how I interpret the information I find?
- What assumptions must I make?

Proving vs. Disproving a Hypothesis

There are many cases where it is more effective and efficient to try to disprove a hypothesis instead of try to prove it.

Example:

The number of attacks against infrastructure companies is the same as those on the commercial sector in general.

There are many cases where it is not really possible for an analyst to prove a hypotheses.

Example:

Attacker ABC is deliberately targeting IoT video cameras.

Discussion:

Proving intentions is very difficult. Without talking to the attacker, it is indistinguishable whether the attacks targeted a specific type of device, the attacks were just the ones caught, or the attacks were only successful on those devices.

Analytical Acumen: Anchoring Bias

People tend to over-rely on a single piece of information, often the first piece of information received. This is a big component of social engineering.

Example:

The financial department gets an email request from an “executive” to wire money to a new vendor for a deal that just closed. The email address appears to be correct for the executive, so the money is wired without verifying the request with the executive by other means.

What are your real-world examples?

Scenario

Scenario: DavyJonesLocker

Background:

You are an analyst responsible for threat intelligence. You receive a report on DavyJonesLocker cryptoware. The provided indicators of compromise are:

- communications to a C2 server at IP address X.X.X.77
- domain name underthesea.xyz is associated with the crime syndicate that authored the malware
- only Microsoft Office documents on Window's 7 machines are encrypted
- uses PsExec for lateral movement
- no current methods for decryption are known

Scenario: DavyJonesLocker

Which hypotheses should we pursue?

1. Our network is not vulnerable to this ransomware.
2. One or more user machines are already infected with this ransomware.

Scenario: DavyJonesLocker 1

Let’s make our plan.

- define the criteria that would tell if our hypothesis is true or false or uncertain
- investigate how the cryptoware is distributed
- check operating system versions of all devices
- check if PsExec is enabled on any device

Scenario: DavyJonesLocker 2

Let's make our plan.

- define the criteria that would tell if our hypothesis is true or false or uncertain
- investigate how the cryptoware is distributed
- check operating system versions of all devices
- check if PsExec is enabled on any device
- check for traffic to the C2
- check for any known files associated with the malware
- check for reports of encrypted or corrupt files

Environmental Context

Why Context Matters

Context provides the details

- needed to gauge potential impact and prioritize investigations
- that let you determine mitigations and recourse
- necessary to identify the scope of an investigation
- necessary to identify the vulnerabilities that allowed an event
- required to understand the types of analyses needed to get insightful results

Knowing a Cyber Environment

Cyber environments consist of

- assets
- people
- policies, protocols, and procedures

Hopefully, these are documented in

- network maps, asset lists
- user lists, user roles
- employee manuals, acceptable use policies, configuration policies, standard operating procedure documents, incident response plans

When Context Is Missing

Things to try when context is missing:

- Ask someone who might know the context.
- Infer it from available information.
- Guess at the possibilities and engage in simple “what-if” analysis.

Analytic Acumen: What-if Analysis

What-if analysis involves testing how different values for a variable change the analysis outcome.

This is useful when the actual value for a variable is unknown.

Example:

Internet traffic to a known-bad URL was detected in network flow. It is unknown if the web proxy allowed the traffic out to its destination or not.

- Possibility A is that the web proxy allowed the traffic, so further investigation is needed.
- Possibility B is that the web proxy blocked the traffic, so further investigation is not needed.

Analytic Acumen: Selective Perception Bias

Expectations do not always match reality and can lead to overlooked information and misinterpretation.

Be cautious about how your expectations influence what you see.

Example:

Analysts receives an alert from a virus scanner about a file with a detection of EICAR Test File. Analysts expectation is that the virus scanner find malware, so the detection is escalated. Reality is that EICAR Test File is a testing mechanism for security appliances. Similarly, PUPs often trigger similar reactions.

Scenario

Scenario: DavyJonesLocker 1

Finding our context:

- Where do we find the information on operating systems?
- How do we find out about PsExec on hosts?

Scenario: DavyJonesLocker 2

Finding our context:

- Where do we find the information on operating systems?
- How do we find out about PsExec on hosts?
- What do we need to know about the network to check for C2 traffic?
- How would we check for reports of encrypted or corrupt files?

Data Gathering

What do you need?

Evidence to support or disprove the hypotheses.

In the cyber realm this may include information on

- how appliances, services, and threats operate.
- a device's (or user's) activities.
- relevant policies, allowed activities, and expected uses.

Where do you get it?

How appliances, services, and threats operate

- domain knowledge, specifications (like RFCs)
- user manuals, white papers, observations
- device and appliance configurations/environment setup
- threat reports (e.g., intelligence reports, malware analysis results)

A device's (or user's) activities

- logs
- eye-witness accounts

Relevant policies, allowed activities, and expected uses

- organizational policies, management expectations

What if you can’t find it?

Just like for missing context:

- Ask someone who might know this information.
- Infer it from available information.
- Guess at the possibilities and engage in simple “what-if” analysis.

Analytic Acumen: Do you really need more?

The desire for more information is a common theme among analysts, but more is not always better.

Questions to ask:

- Do I truly understand the data I already have?
- Is the data I need missing or do I just need to find it in what I have?
- Do I need more data or different data?

Analytic Acumen: Information Bias

People tend to seek information even when that information will not change the end results.

Analysts need to focus on the information that will change an interpretation or decision.

Example:

An analyst gets an alert about traffic from a specific IP address. He or she starts by looking up the geolocation of the address, even though that is not a criteria for determining maliciousness.

Relevant or Not?

Should a manager reprimand an employee who visited a blocked website?

1. The website was registered in a foreign country.
2. The website's default language was not English.
3. The website is categorized as unknown.
4. The employee was supposed to be taking care of patients.

Do I need to take an umbrella tomorrow?

1. Tonight's sky is red.
2. Leaves are upside down.
3. Farmers almanac states this month will be rainier than normal.
4. It is currently raining.
5. It rained on this day last year.
6. I only park in garages.

Should I click the link?

1. Sender is my grandmother.
2. Email appears to be a chain letter.
3. Mouse-over points to a tinyurl.
4. You know everyone else in the recipient group.

Should I buy a lottery ticket?

1. I found a penny heads up.
2. I have a few extra bucks.
3. I could play my lucky numbers.
4. I rigged the system.

Should I ignore this certificate warning?

1. It occurs on a page on our work domain.
2. It is a self-signed certificate warning.
3. It uses RC4.
4. The webpage was not blocked by the web proxy.

Scenario

Scenario: DavyJonesLocker

Gather your data.

Microanalysis

What is microanalysis?

Microanalysis is the process of trying to figure out what occurred and how it happened. Or similarly, if something occurred at all.

Example scenario	What needs to be determined
Alert about use of privileged account on a sensitive server	Was the use authorized? If not, what did the user do? How did the account get access?
Email submitted to abuse mail box	Did the email result in infection for the submitter or any other recipient?

Common Microanalysis Techniques

Direct investigation

- looking at an asset or various logs to find direct evidence

Computational analysis

- using statistics and other computational methods to find anomalies or patterns as evidence

Direct Investigation

In the cyber realm, direct investigation may involve

- Checking device security or application logs
  - web proxy/firewall
  - server/PC
- Looking through network traffic capture
  - network flow
  - full packet capture
- Examining files
- Forensic analysis of a device
- Talking to end users

Phishing Email Investigation

Investigating Phishing Attempts

1. Look at the message
2. Research links and files
3. Check if others received similar messages
4. Check for traffic to suspicious domains

Computational Methods

Types of computational methods

- statistical: using statistics to gain insight from data
- machine learning: transforming or analyzing the data to find something you didn't know before
- data mining: extracting something you know is there from a large dataset

Methods and concepts have varying levels of complexity

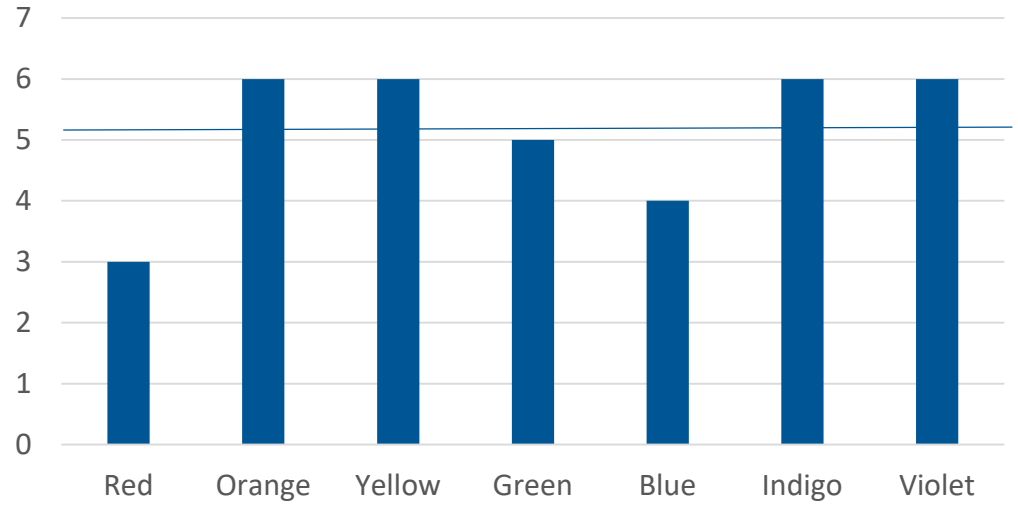
- means, normal distributions, standard deviations
- clustering
- trend analysis

Statistics for Simple Anomaly Detection

Mean

Simple average

1. Add up the items.
2. Divide the result by the number of items.



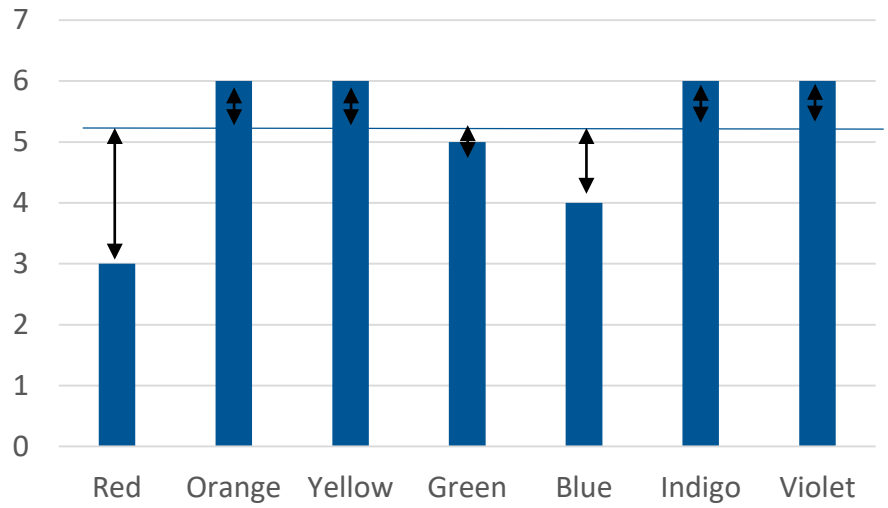
Example: compute the mean number of letters in the colors of the rainbow

(3 + 6 + 6 + 5 + 4 + 6 + 6) ÷ 7 = 5.14…

Variance

Measures how different the values are from the mean, on average.

More technically, variance is the mean of the squared difference from the mean of all items.



Example: variance of number of letters in the colors of the rainbow

$[(3 - 5.14)^2 + (6 - 5.14)^2 + (6 - 5.14)^2 + (5 - 5.14)^2$

$+ (4 - 5.14)^2 + (6 - 5.14)^2 + (6 - 5.14)^2] \div 7 =$

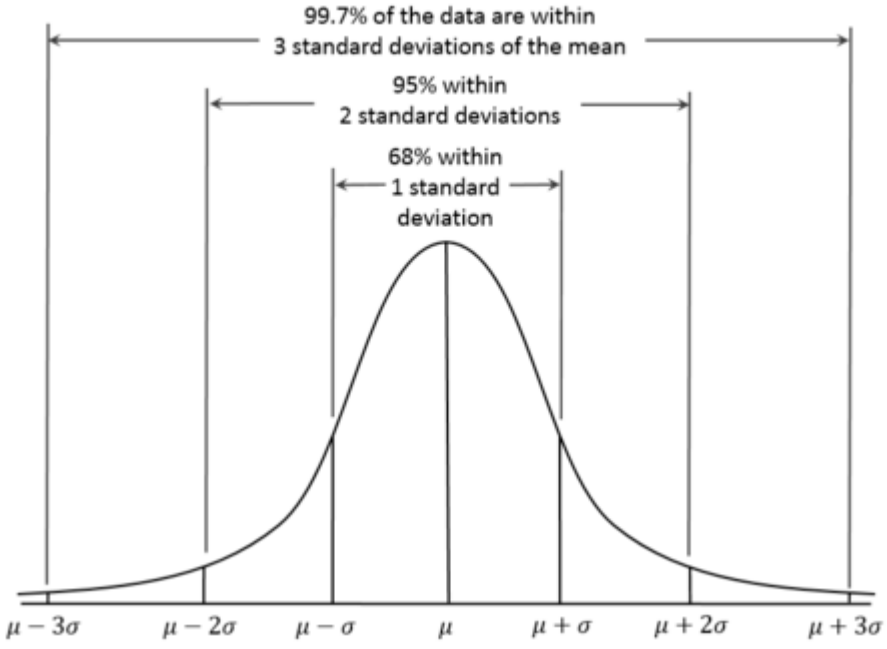
$[-2.14^2 + .86^2 + .86^2 - .14^2 - 1.14^2 + .86^2 + .86^2] \div 7 =$

$[4.5796 + .7396 + .7396 + .0196 + 1.2996 + .7396 + .7396] \div 7 =$
$8.8572 \div 7 = 1.265\ldots$

Why do we care?

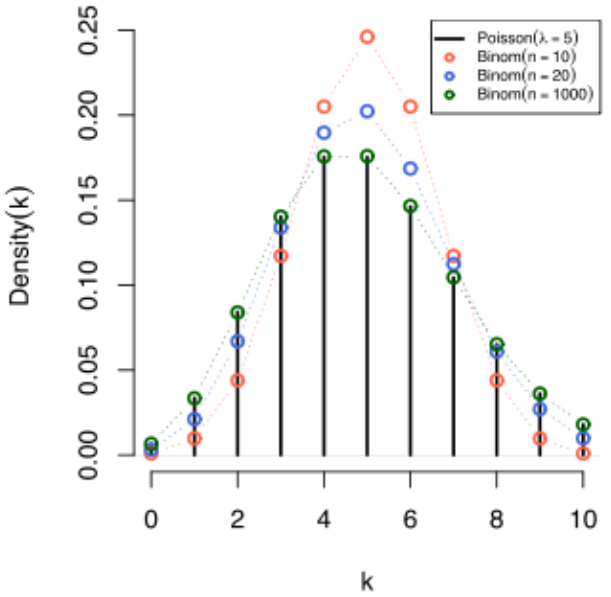
Values occurs in predictable frequencies.

https://en.wikipedia.org/wiki/68%E2%80%9395%E2%80%9399.7_rule



This is true even if your data is not a normal distribution.

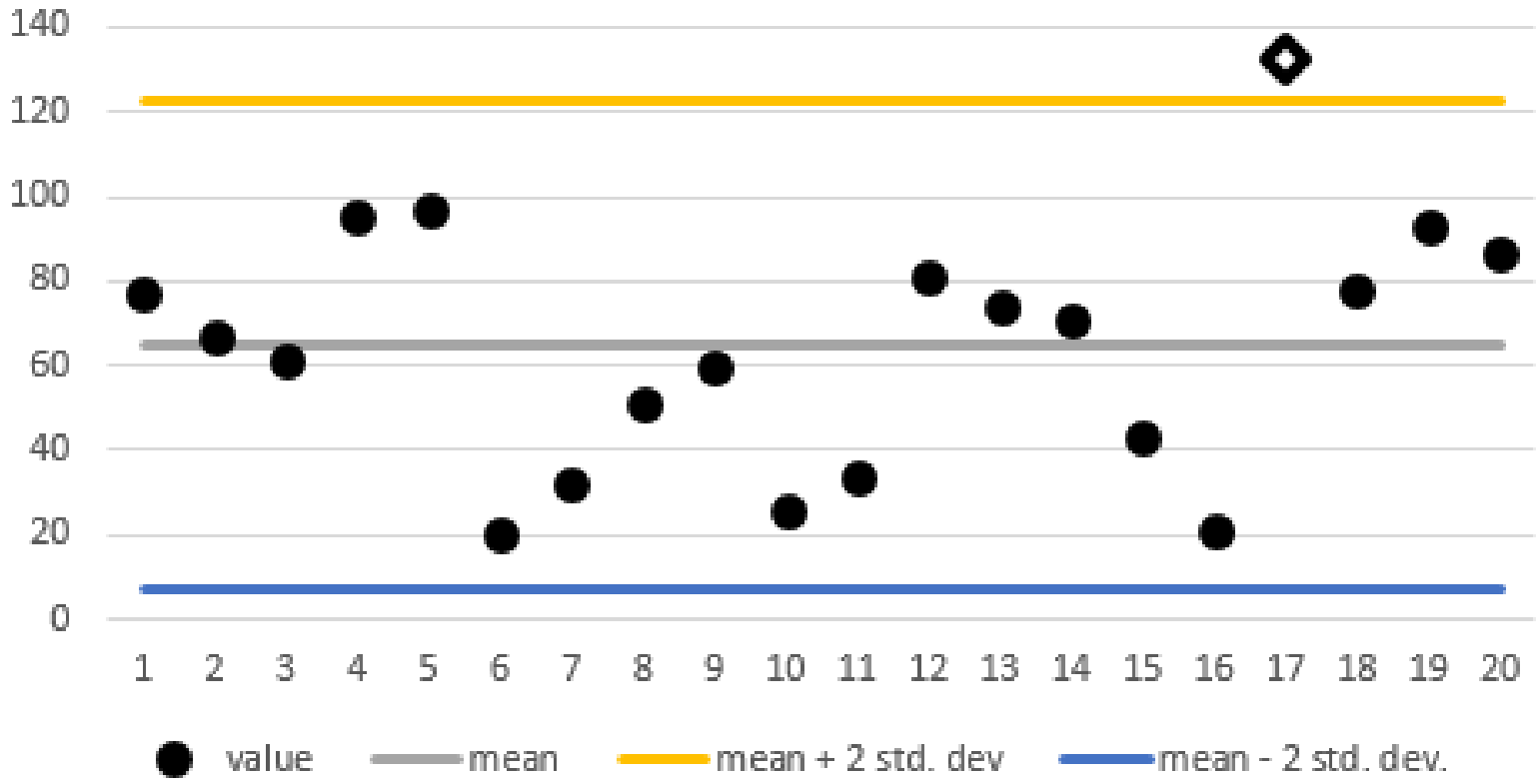
https://en.wikipedia.org/wiki/Poisson_distribution



This means that we can say how likely a value is to occur.

This lets us identify unusual events—anomalies.

For example, in normally distributed data, the diamond is expected to occur less than 5% of the time.



Analytic Acumen: Satisficing is not [Always] Sufficient

Satisficing is the strategy that looks for the first “satisfactory” solution, answer, or decision—as opposed to the optimal one.

There are many instances where this is a reasonable strategy:

- choosing what to eat for lunch
- investigating and handling low-impact threats

But following this strategy can have undesired consequences:

- In hiring, it can lead to poorly-formed teams that lack diversity and have incompatible personalities.
- In malware response, it can lead to reoccurring or spreading infections.

Analytic Acumen: Confirmation and Conservatism Biases

Confirmation: people tend to only consider information that confirms their already-held beliefs.

Conservatism: people prefer evidence that supports their current beliefs, even when new evidence suggests a change is needed.

Analysts need to actively look for information that would make them change their minds.

Example:

An organization experiences a cyber campaign that they believe is by a certain nation-state. They research indicators that they see in the attack and find that some of them are associated with attacks by that nation-state. They ignore indicators that have no association with that nation-state or disregard information that these indicators are also used by many other actors.

Scenario

Scenario: DavyJonesLocker 1

Analyze your data.

What, if any, hosts are vulnerable?

Why are they vulnerable?

Scenario: DavyJonesLocker 2

Analyze your data.

What, if any, hosts are compromised with DavyJonesLocker? How certain are you?

Is lateral movement possible? Is it occurring? How certain are you?

Macroanalysis

What is macroanalysis?

Macroanalysis is the process of adding perspective, context, and depth to analysis. Often, this involves trying to figure out who (or what) did something and why they did it.

Example scenario	What needs determined
Unauthorized use of a privileged account on a sensitive server	Who used the account? What was their end goal?
Spear phishing email sent to a researcher	Who sent the email? What did they hope to gain by it?

Common Macroanalysis Techniques

Intelligence research

- using existing reports on similar events to provide insight

Data fusion

- pulling together related information from various sources to provide more complete data

Intelligence Research

In the cyber realm, intelligence research sources include

- blogs (e.g., insights.sei.cmu.edu, isc.sans.edu)
- government information (e.g., www.us-cert.gov, www.enisa.europa.eu)
- vendor reports
- social media sites
- darkweb sites

Data Fusion

Fusing small pieces of data from various sources often provides better understanding.

In the cyber realm, these types of information are often useful

- registry and whois data (www.iana.org, www.iana.org/whois)
- domain and IP address history (www.robtex.com)
- known bad data bases or black lists (www.virustotal.com)

Analytic Acumen: Mosaic Theory or Not?

Mosaic theory of analysis

- Analysis is like a puzzle.
- You gather as many little pieces as possible and put them together to see the picture.

How analysts really work

- Analysis is like a medical diagnosis.
- You look at a few pieces, come up with a theory, and work from there.

Analytic Acumen: Recency Bias

People tend to think of the latest as the best.

Be careful not to disregard information as invalid just because it is "old."

Example:

An analyst sees an alert on traffic that matches a signature for a piece of malware that was prevalent three years ago.

Scenario

Scenario: DavyJonesLocker 1

Analyze your data.

- Who is spreading the ransomware?
- Why do we suspect the adversary will target us?

Scenario: DavyJonesLocker 2

Analyze your data.

- What can we tell about who is spreading the ransomware?
- Why do you think they are doing so?

Reporting and Feedback

Sharing is important.

If you are a one-man shop and do not need to work with anyone else, maybe you do not report.

Otherwise, incident ticket documentation, IR reports, and talking to other analysts or managers are all forms of reporting.

When reporting:

- Know your audience.
- Make your point clear and provide evidence to support it.
- Acknowledge the limitations of your analysis and any other possibilities.

Analytic Acumen: Facts Require Interpretation

Contrary to popular belief,

facts do not speak for themselves.

Everyone interprets what they see based on their knowledge, experience, and biases.

Analysts must be aware of how these three things influence their interpretations and account for them in their findings and reporting.

Analytic Acumen: Blind-spot Bias

People tend to not recognize their own biases.

Analysts should have someone review their findings, especially for those that are very important or go to upper management or external entities.

Example:

An analyst selected a new security appliance and is conducting a pilot test. Other analysts should review the test to ensure that the analyst is not influenced by personal biases such as choice-supportive bias, pro-innovation bias, and halo effect.

Scenario

Scenario: DavyJonesLocker

Who makes up our target audiences?

What do we share with each audience?

How do we share it?

References

Lee, Robert M. & Bianco, David. Generating Hypotheses for Successful Threat Hunting. *SANS Institute Website.* https://www.sans.org/reading-room/whitepapers/threats/generating-hypotheses-successful-threat-hunting-37172

Lubin, Gus & Lebowitz, Shana. 58 cognitive biases that screw up everything we do. *Business Insider Website.* http://www.businessinsider.com/cognitive-biases-2015-10/