Analyze

Define The Current Process Capability

Analyzing Data Graphically
- Stability
- Shape
- Central tendency
- Variation / Spread

Process Capability
- Discrete Data
- Continuous Data

Define The Performance Objective

Benchmarking

Identify the Key Drivers of Variation

- Focus Groups
- Value Add/Non-Value Add Analysis
- Fishbone
- Z Test
- Hypothesis Testing
- Correlation
- Regression

In the Analyze phase, we:

1. **Define the current baseline process capability.**
2. **Define Performance Objective.**
3. **Analyze historical data to identify the Key Drivers of Variation.**

This phase is crucial as it clarifies the current performance level of the process and identifies the primary factors responsible for performance variation. Identifying these key drivers of variation enables us to focus on them in subsequent project stages to achieve the desired improvements.

1) Define the current baseline process capability:

Process Capability Definition:

Process capability refers to the capacity of a process to consistently produce defect-free products or services within a controlled production or service environment.

Calculating Process Capability

In this stage, we determine the process's ability to meet performance standards by calculating the sigma level. This is done using the data collected during the Measure phase.

Establishing a Baseline

Defining process capability is essential for establishing measurable goals. To set achievable improvement targets, we must first understand our current position. Baseline process capability provides a starting point from which we can measure progress and set improvement objectives.

Key Factors in the Current Baseline Process Capability:

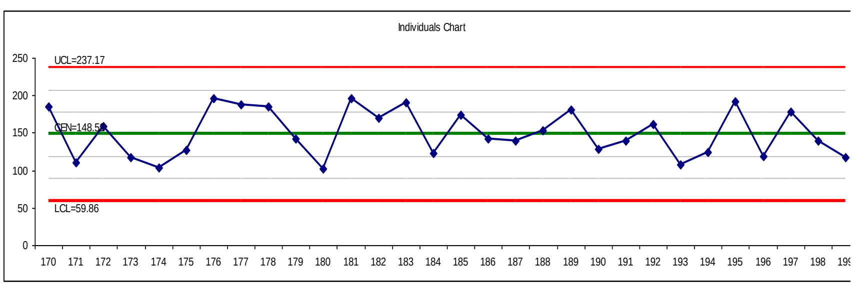
When establishing the baseline for process capability, it's crucial to assess four key factors:

1. **Stability:** Ensure the process is stable and predictable over time without excessive variation or fluctuations.

The objective of this step is to ensure the use of valid data for establishing our process baseline. It's important to be vigilant for the following:

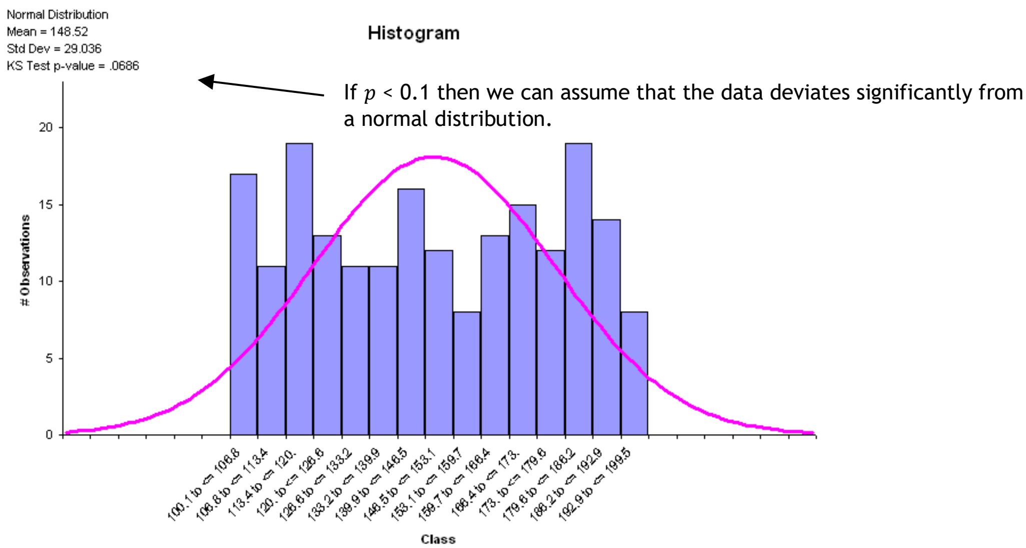
Trends: Look out for significant trends such as program ramps or gradual shifts in process behavior over time.

Special Causes: Identify one-off events or unique occurrences that can influence process performance, such as product modifications or launches initiated by the client. By recognizing and addressing trends and special causes, we can maintain the integrity of our baseline data and accurately assess process capability. This awareness enables us to make informed decisions for process improvement and optimization.



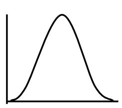
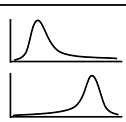
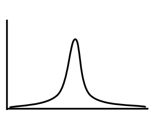
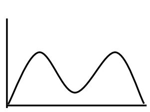
2. **Shape:** To evaluate the distribution shape of process outputs, you can use various statistical tools and techniques. Here are some commonly used methods:

Histograms: Plotting histograms of your process data is a simple yet effective way to visually assess the distribution shape. Histograms show the frequency distribution of data points within specific intervals (bins). By examining the shape of the histogram, you can get a sense of whether the data is symmetric (like a bell curve for normal distribution), skewed to the left or right, or exhibits other patterns.



3. **Centering:** Determine the alignment of process outputs with the target or desired value. This assesses whether the process is consistently hitting the intended mark.

4. **Spread:** Measure the variability or spread of process outputs to gauge how much the results deviate from the target value.

Shape	**Center** (central tendency)	**Spread** (variation)
 normal	Mean ($\bar{x}$)	Standard Deviation (s)
 skewed	Median ($\tilde{x}$)	Stability Factor (SF) = Q1/Q3
 long-tailed	Median ($\tilde{x}$)	Span (5-95)
 bimodal	The different processes must be stratified before calculating descriptive statistics.	

By considering these factors, we can effectively establish a comprehensive baseline for process capability, enabling informed decision-making and targeted improvements.

Calculating Current Process Capability:

To measure the Current Process Capability, we need to follow these steps:

1. **Collect Data:** Gather data on a key process characteristic or output that you want to assess for capability.

2. **Calculate the Process Mean ($\bar{X}$) and Standard Deviation (σ)** of the collected data. The mean ($\bar{X}$) represents the average value, and the standard deviation (σ) measures the variation or dispersion of the data around the mean.

3. **Define Specification Limits** for the process characteristic. These are typically the upper and lower limits within which the process output should ideally fall to meet customer requirements.

4. **Calculate Process Capability Indices:** Use the following formulas to calculate process capability indices:

$$Cp = \frac{USL - LSL}{6\sigma}$$

where USL = Upper Specification Limit, LS= Lower Specification Limit, and σ = Standard Deviation

2) Define Performance Objective:

Defining a performance objective is crucial for setting clear improvement goals and focusing efforts toward achieving specific outcomes. Here's how you can effectively define a performance objective:

1. **Statement of Improvement:** Clearly articulate the expected reduction in defects or improvement in performance that you aim to achieve through the project. For example, specify the percentage reduction in defects, increase in productivity, or enhancement in quality metrics.

2. **Quantifiable Target:** Express the improvement goal in measurable terms. Use quantitative metrics such as reducing defects by a certain percentage (e.g., 20% reduction in defects), improving cycle time by a specific duration (e.g., 15% reduction in cycle time), or increasing customer satisfaction scores by a defined rating (e.g., achieve a customer satisfaction score of 90%).

3. **Statistical Definition:** Support the performance objective with statistical parameters or benchmarks. Use statistical methods to define the baseline performance and set a statistically significant target for improvement. This could involve analyzing historical data, conducting process capability studies, or setting performance thresholds based on industry standards.

4. **Financial Benefits:** Estimate the potential financial benefits associated with achieving the performance objective. Quantify the cost savings, revenue increase, or return on investment (ROI) expected from the improvements. This helps in aligning the project with organizational objectives and justifying resource allocation.

5. **Clarity and Focus:** Ensure that the performance objective is specific, achievable, relevant, and time-bound (SMART). This clarity will guide project planning and implementation, enabling teams to work towards a tangible and meaningful goal.

By defining a performance objective with these considerations, you provide a clear direction for improvement efforts, facilitate measurement of progress, and establish a basis for evaluating project success. This approach ensures that improvement initiatives are aligned with strategic goals and yield measurable benefits for the organization.

3) Identify the Key Drivers of Variation:

Identifying the drivers of variation involves pinpointing the key factors (X's) that significantly impact a particular metric or outcome (Y) within a process. This step is crucial in process improvement and problem-solving for several reasons:

Focus on Root Causes: By identifying the drivers of variation, you shift the focus from merely observing outcomes (Y's) to understanding and optimizing the inputs and process variables (X's) that influence these outcomes. This allows you to address the root causes of issues rather than just treating the symptoms.

Efficient Resource Allocation: Knowing the critical factors that affect your desired metric helps allocate resources more efficiently. Instead of addressing all variables indiscriminately, you can prioritize efforts and resources on those X's that have the most significant impact on improving outcomes.

Targeted Improvement Strategies: Once you identify the drivers of variation, you can develop targeted strategies to optimize these factors. This might involve adjusting process parameters, implementing controls, or making changes to inputs to minimize variability and improve overall performance.

Data-Driven Decision Making: The identification of significant X's is based on statistical analysis of historical data. This data-driven approach provides objective insights into which variables are most influential, enabling informed decision-making in process improvement initiatives.

Continuous Improvement: Understanding the drivers of variation is integral to continuous improvement efforts. By continually monitoring and optimizing these critical factors, organizations can sustainably enhance process efficiency, product quality, and overall performance over time.

Identifying the drivers of variation is essential because it shifts focus toward understanding and optimizing the inputs and process variables that impact outcomes. This approach enables targeted improvements, efficient resource allocation, and data-driven decision-making to achieve sustainable process optimization and quality enhancement.

We identify the key drivers of variation using the following qualitative & quantitative tools in this step:

1. **Focus Groups**
2. **Root Cause Analysis / Fishbone**
3. **Outlier Analysis**
4. **Value Add / Non-Value Add Analysis**
5. **Graphical tools:**
   - Pareto, Scatter Plot, Bar Charts, Histograms, Box Plots
6. **Statistical tests:**
   - Z-Test, ANOVA, Correlation, Regression, Chi-Square

1) Focus Groups:

- The focus group should represent a diverse cross-section of the call center, including L1s, L2s, Supervisors, Operations Managers, Learning Team Resources, and others.

- Focus groups are instrumental in identifying potential causes of problems and gathering and prioritizing ideas for solutions. Causes can be structured using a Fishbone Diagram (see the next section).

- Share identified potential solutions with the focus group to gather feedback.

- Probe for suggestions on the preferred vehicle(s) for implementation:
  - Training
  - Process
  - Job aids
  - Online tools

- Checklists
- Quality assurance processes
- Etc.

Seek input on the specific content, timing, and required resources for each vehicle.

The focus group can transition into the project team for future phases.

Once details such as vehicle, content, and other specifics are finalized, the focus group should review them to ensure alignment with objectives before implementation.

2) Root Cause Analysis / Fishbone:

To conduct a Root Cause Analysis using Fishbone Diagrams and the "5 Whys" technique, follow these steps:

a) Fishbone Diagram Development:

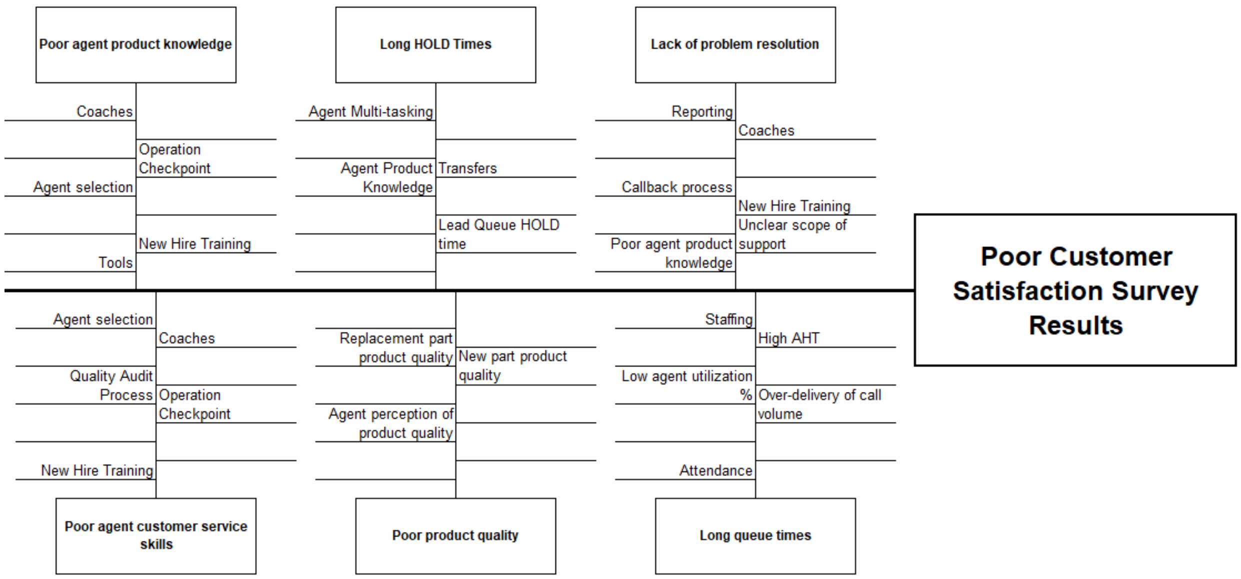
Gather insights from the focus group regarding the current problem or issue.

Use a Fishbone Diagram (also known as Ishikawa or Cause-and-Effect Diagram) to visually map out potential causes. The main categories typically include:

Methods: How the process is performed.
Machines: Equipment and tools involved.
Materials: Raw materials or resources used.
Manpower: People involved in the process.
Measurement: Metrics and methods used for evaluation.
Environment: Conditions where the process occurs.

Organize these categories as branches stemming from the main problem statement on the fishbone diagram.

Example of a fishbone diagram:



b) Utilize the "5 Whys" Technique:

Select one of the identified causes from the Fishbone Diagram.

Ask "why" to understand the cause at a deeper level. Repeat this process up to five times (or more if necessary) to uncover underlying reasons behind each cause. The goal is to move beyond superficial reasons to identify root causes.

Example:

Problem: The Customer Satisfaction (CSAT) rate is low.
Why is the CSAT low? (1st Why)
Because agents are not using empathy in their communication.
Why are agents not using empathy? (2nd Why)
Because they don't feel the customer's frustration.

Continue this process to reach deeper insights into root causes.

c) Prioritize Causes Using Nominal Group Technique (NGT):

After identifying potential causes through the Fishbone Diagram and "5 Whys," prioritize them using NGT.

Gather a group of stakeholders or subject matter experts.

Have each participant independently rank the identified causes based on their perceived impact or importance.

Compile and analyze the rankings to determine the most significant or likely root causes.

d) Validate Causes Through Statistical Analysis:

Once causes are prioritized, validate them through further analysis such as statistical methods.

Use data analysis techniques (e.g., hypothesis testing, regression analysis) to verify the relationship between identified causes and the observed problem.

Ensure that the root causes identified align with the data and evidence collected during the analysis phase.

By following these systematic steps, you can effectively identify and validate root causes of the problem, enabling targeted interventions and solutions to address underlying issues and improve overall process performance. Each technique complements the other in providing a comprehensive understanding of the problem's origins and contributing factors.

3) Outlier Analysis:

- For each metric improvement project, a report should be generated to highlight outliers, both top-performing and bottom-performing.
- Each metric improvement "Bridge Plan" should encompass a sub-project aimed at enhancing the performance of bottom X% outliers. These sub-projects often yield the most significant impact on overall metric improvement.
- Moreover, analyzing the practices of top performers is essential to identify "best practices" that can be adopted by bottom-performing outliers, or potentially across the entire site or campaign.

4) Value Add / Non-Value Add Analysis:

The initial step is to create a highly detailed process map.

Conduct a time and motion study to measure the duration spent on each step.

Next, assess each step based on the following criteria:

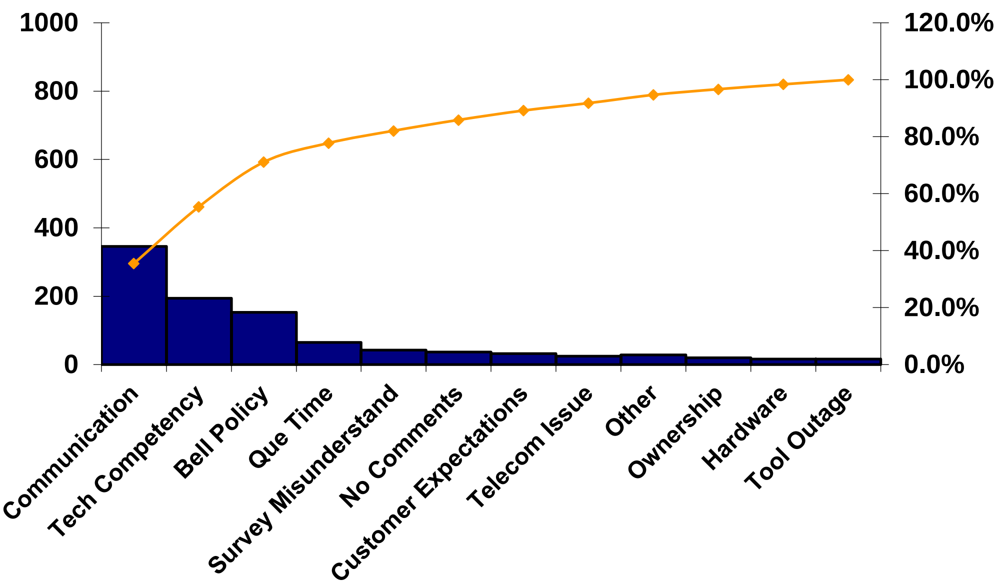
- Determines if the step adds value. Note that all "Moments of Truth" are considered value-adding.
- Identifies if the step is a Value Enabler.
- Determines if the step is Non-Value Adding.

In our projects, we should focus on reducing the NVA steps.

5) **Graphical tools** (Pareto, Scatter Plot, Bar Charts, Histograms, Box Plots):

Pareto:

- Origin: The Pareto principle, named after Italian economist Vilfredo Pareto, highlights that roughly 80% of the wealth is owned by 20% of the population.
- What 20% of sources are causing 80% of the problems (80/20 Rule)?
- Where should we focus our efforts to achieve the greatest improvements?



Histograms:

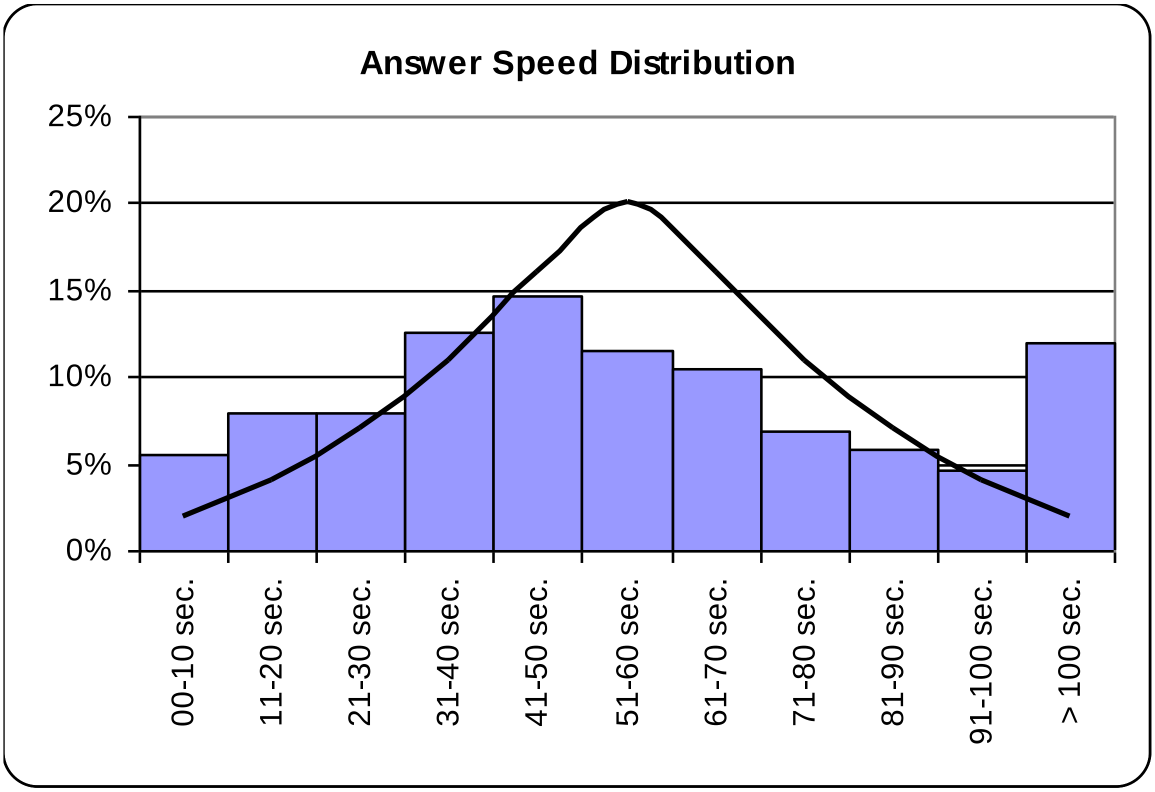
The chart below illustrates the distribution of answering speeds in the Escalation queue. Although the average answering speed is currently 50 seconds, agents have been expressing concerns about the prolonged Average Speed of Answer (ASA), leading to extended wait times for customers and increased dissatisfaction.

The blue bars in the chart represent the actual distribution of answering speeds. This distribution does not follow a "normal" curve.

Notably, there's a significant concentration of answering speeds exceeding 100 seconds. This extended wait time affects 14% of agents calling the Escalation Queue, resulting in customer dissatisfaction.

Despite the average meeting the goal, this distribution highlights the reason behind complaints about long hold times experienced by customers and agents.

The black bell-shaped line represents a "normal" distribution curve.

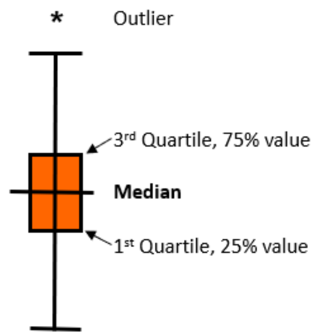


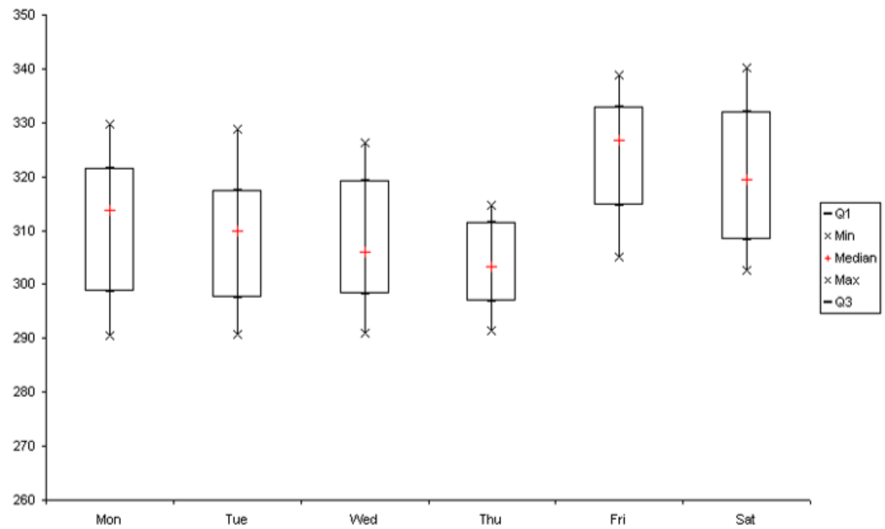
Box Plots:

A box plot is a graphical representation of the variation of a variable (Y) across different categories or groups (X). The size of the box in a box plot indicates the extent of variation in the data for each category of X:

- A larger box indicates higher variation (lower consistency) in the data for that category of X.
- A smaller box indicates lower variation (higher consistency) in the data for that category of X.

Box plots are versatile and can be used to visualize the distribution of both normal and non-normal data effectively.





6) **Statistical tests** (Correlation, Regression, Bar Charts, Z-Test, ANOVA, Chi-Square):

	Continuous Y	**Discrete Y**
Continuous X	Scatter Plot Correlation Linear Regression	Convert X to discrete data and use Chi-Square
Discrete X	Histograms Box Plots Z-tests 1 way ANOVA Mood's Median	Pareto Chi-Square

Correlation Analysis:

Correlation analysis is a method used to identify relationships between different metrics or variables. Correlation can be either positive or inverse (negative):

- **Positive correlation:** When one metric increases, the other metric also increases. Conversely, when one metric decreases, the other metric decreases.
- **Inverse (negative) correlation:** When one metric increases, the other metric decreases. Conversely, when one metric decreases, the other metric increases.

If a correlation exists between two metrics, it suggests a relationship where changes in one metric may lead to changes in the other metric. This understanding can inform projects aimed at improving specific metrics to drive improvements in related areas.

The correlation formula calculates an R-value, which ranges between -1 and 1:

- An R-value closer to -1 indicates a strong inverse correlation.
- An R-value closer to 1 indicates a strong positive correlation.
- An R-value of 0 suggests no correlation between the metrics.

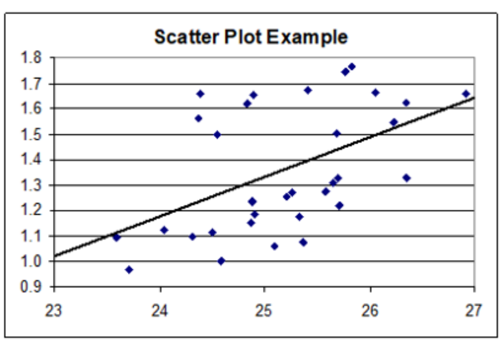
To interpret the R-value:

The absolute value of the R-value (|R|) is compared against a minimum R-value threshold based on the sample size to determine if a statistically significant correlation exists.

You can easily calculate the correlation using Microsoft Excel by using the "CORREL" formula, which computes the correlation coefficient (R-value) between two sets of data.

Correlation

R Value	**0.7704**
Sample Size	25
T-value	2.0687
Minimum R Value	**0.3961**
Correlation	**Yes**



Data to Correlate

Metric 1	Metric 2
23.04	1.12
23.31	1.10
24.37	1.08
24.09	1.06
23.59	1.00
22.71	0.97
21.22	0.98
22.59	1.10
23.51	1.11
23.86	1.15
24.33	1.18
23.91	1.19
23.88	1.24
24.71	1.22
24.22	1.25
24.65	1.31
24.58	1.27
24.27	1.27
25.35	1.33
25.70	1.33
25.69	1.50
26.23	1.55
24.04	1.12
24.31	1.10
25.37	1.08

The correlation calculator above displays correlation statistics comparing the Average Handle Time (AHT) of agents with their average number of Escalation requests per day. The calculator indicates that a correlation exists because the absolute value of the correlation coefficient (R-value) is greater than the minimum R-value threshold based on the sample size. Therefore, based on this correlation:

As the number of Escalation requests increases, the AHT of the agent also increases, or vice versa. This relationship suggests that changes in one metric tend to coincide with changes in the other metric. However, it's important to note that correlation does not imply causation; an increase in one metric may not necessarily cause an increase in the other.

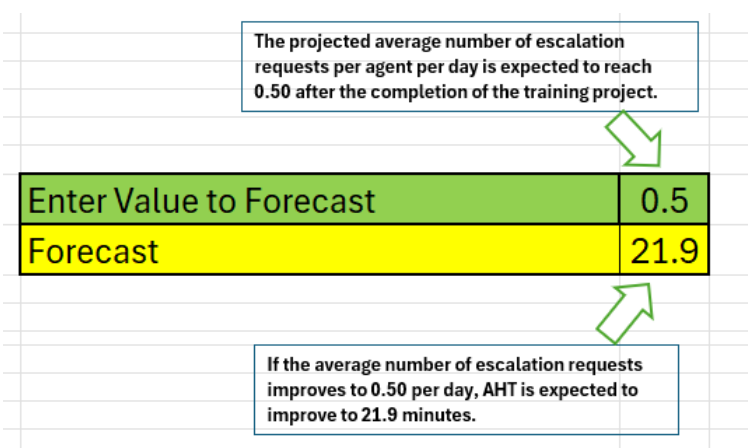
To address the goal of improving AHT without compromising quality, focusing on reducing the number of Escalation requests could be an effective strategy. This could involve targeted training to address the primary causes of Escalation requests, aiming to decrease their frequency and subsequently improve overall AHT performance.

Regression Analysis:

When two sets of data exhibit a correlation, Regression Analysis can be employed to predict the performance of one metric based on the performance of another metric. This regression calculation can be easily conducted using the 'FORECAST' formula in Microsoft Excel.

The Regression Analysis below uses the example previously demonstrated in the Correlation example.

Given that the R-Value was 'significant' and a correlation exists, the Regression formula can be utilized to forecast Average Handle Time (AHT) based on a project aimed at reducing the average number of Escalation requests to 0.50 per agent per day. This will be achieved by training agents on effective troubleshooting techniques for symptoms commonly escalated to the Escalation group.

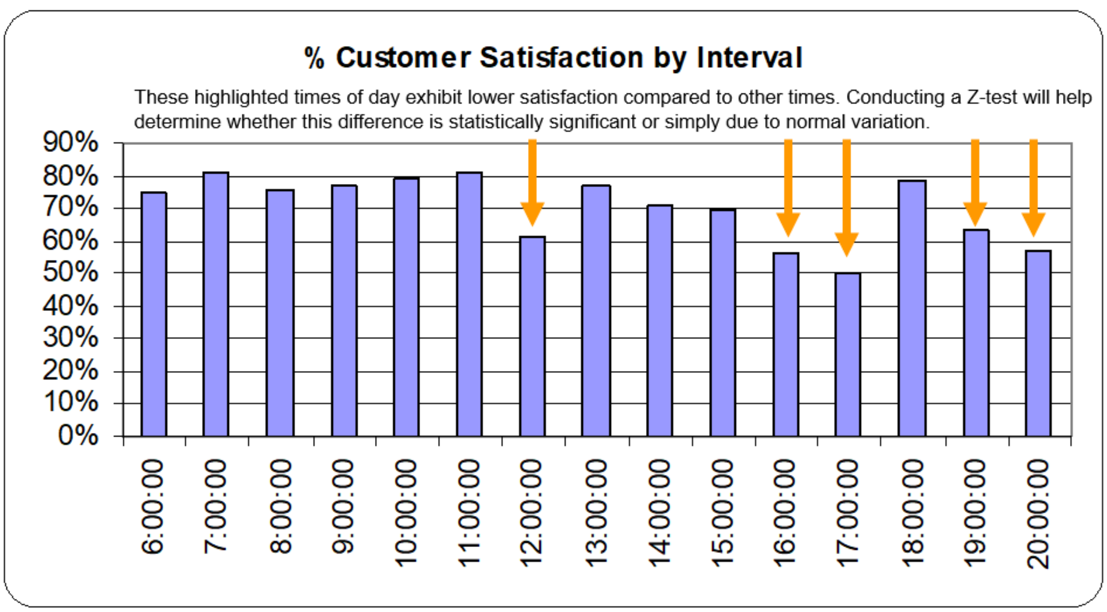


Bar Charts:

- Bar charts are effective for visualizing metric performance across various subsets of the population such as teams, tenure bands, product lines, time of day, and day of the week. These charts can be easily created using Microsoft Excel.

- A bar chart can highlight specific subsets of the population that might benefit from improvement initiatives. However, it's important to ensure that any observed differences between the target subset and the overall population are statistically significant before investing time, money, and resources into improvement efforts.

- Differences in metric performance between a subset and the population can sometimes be attributed to normal variation. Before targeting a subset based on a bar chart analysis, it's advisable to conduct a Z-test to determine statistical significance. Z-tests are explained in the next section.



Z-Tests:

Z-tests determine whether a sample or subset of the population is statistically different from the rest of the population.

A significant Z-score from a Z-test can identify areas that may benefit from improvement initiatives. Conversely, it can also highlight subsets that perform significantly better, prompting further investigation to uncover best practices.

Z-Test for a Sample Mean

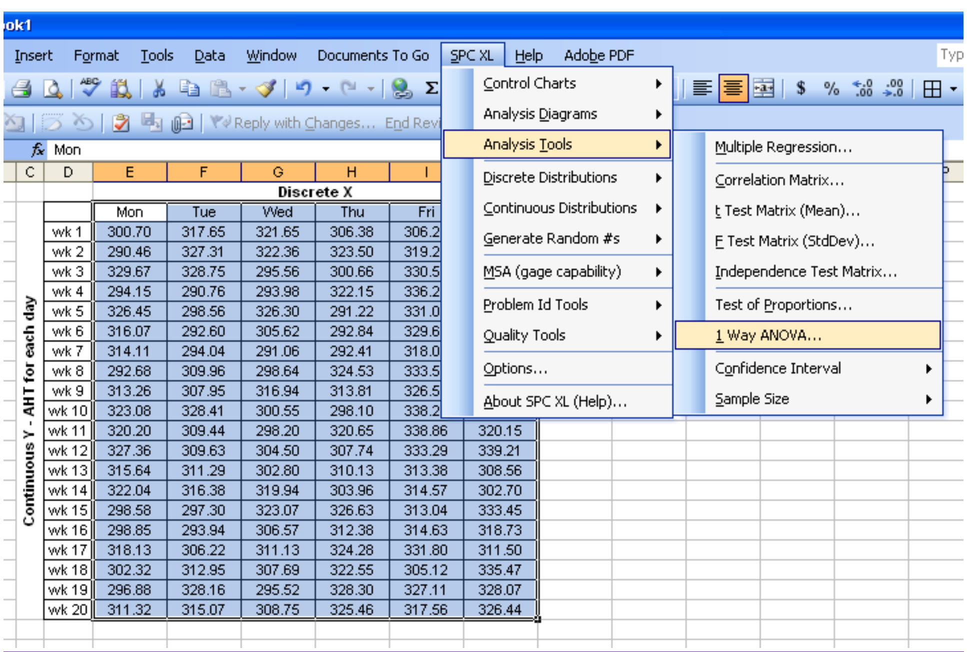
Enter Sample Mean	24.50	AHT of the pilot group after conducting the pilot
Enter Population Mean	28.40	AHT of the pilot group before conducting the pilot
Enter Population Standard Deviation	17.40	Standard deviation of the pilot group before conducting the pilot
Enter the Sample Size	1200	Number of calls handled by the pilot group during the pilot
Z Value	**-7.76**	Note: A Z-value of < -1.96 means that the pilot group significantly lowered their AHT during the pilot. This was not due to random variation.

Z-Test for a Sample Proportion

Enter Sample Proportion	59.00%	CSAT of the pilot group during the pilot
Enter Population Proportion	71.00%	CSAT of the control group during the pilot
Enter the Sample Size	50	Number of CSAT surveys received by the pilot group during the pilot
Z Value	**-1.31**	Since the Z-score is between -1.96 and +1.96, there was no significant difference between the pilot group and the control group

ANOVA – Continuous Y, Discrete X

ANOVA calculations can be performed using SPC Excel, as demonstrated in the example below:



	Mon	Tue	Wed	Thu	Fri	
wk 1	300.70	317.65	321.65	306.38	306.2	
wk 2	290.46	327.31	322.36	323.50	319.2	
wk 3	329.67	328.75	295.56	300.66	330.5	
wk 4	294.15	290.76	293.98	322.15	336.2	
wk 5	326.45	298.56	326.30	291.22	331.0	
wk 6	316.07	292.60	305.62	292.84	329.6	
wk 7	314.11	294.04	291.06	292.41	318.0	
wk 8	292.68	309.96	298.64	324.53	333.5	
wk 9	313.26	307.95	316.94	313.81	326.5	
wk 10	323.08	328.41	300.55	298.10	338.2	
wk 11	320.20	309.44	298.20	320.65	338.86	320.15
wk 12	327.36	309.63	304.50	307.74	333.29	339.21
wk 13	315.64	311.29	302.80	310.13	313.38	308.56
wk 14	322.04	316.38	319.94	303.96	314.57	302.70
wk 15	298.58	297.30	323.07	326.63	313.04	333.45
wk 16	298.85	293.94	306.57	312.38	314.63	318.73
wk 17	318.13	306.22	311.13	324.28	331.80	311.50
wk 18	302.32	312.95	307.69	322.55	305.12	335.47
wk 19	296.88	328.16	295.52	328.30	327.11	328.07
wk 20	311.32	315.07	308.75	325.46	317.56	326.44

The table above displays the Average Handle Time (AHT) for a team across different days of the week over 20 weeks. AHT (Y metric) is continuous, and the Day of the week (X) is discrete. Therefore, ANOVA (Analysis of Variance) is appropriate for this analysis.

After running ANOVA, you obtain a p-value. In hypothesis testing, if the p-value is less than 0.05 (typically chosen as the significance level, denoted as α), then you reject the null hypothesis (Ho).

The null hypothesis (Ho) in this context states that there is no statistical difference in AHT based on the day of the week. Essentially, it assumes that any observed differences in AHT across different days are due to random variability or chance.

By conducting ANOVA and examining the resulting p-value, you assess whether the observed differences in AHT among different days of the week are unlikely to be solely due to random variation. If the p-value is sufficiently small (less than 0.05), it suggests that the differences in AHT are statistically significant, and you may reject the null hypothesis in favor of the alternative hypothesis (Ha), which proposes that there is a meaningful relationship between AHT and the day of the week.

Hypothesis testing, particularly through ANOVA and interpretation of p-values, helps to guide decisions about whether observed differences or effects are likely to be genuine or simply due to chance.

Chi-Square

Chi-square (χ^2) is a statistical test used to determine whether there is a significant association between categorical variables. It is particularly useful for analyzing data that consist of counts or frequencies within different categories.

Here are key points about the Chi-square test and its applications:

- **Types of Chi-Square Tests:**

Chi-Square Test of Independence: This test assesses whether two categorical variables are independent of each other or if there is an association between them.

Chi-Square Goodness of Fit Test: This test compares observed categorical data to expected data to see if they fit a specific distribution or model.

- **Conditions for Use:**

Categorical Data: Chi-square is used when working with categorical data, where observations are placed into categories or groups.

Independence: For the Chi-square test of independence, the data should be collected from random samples, and the observations within each cell of the contingency table should be independent of each other.

Expected Frequencies: The expected frequency count for each cell in a contingency table should be at least 5 for the test to be valid (this is known as the "5 or more rule").

- **Steps in Conducting a Chi-Square Test:**

Formulate Hypotheses: Set up the null hypothesis (H_0) and alternative hypothesis (H_a) to test the association between categorical variables.

Collect Data: Gather categorical data and organize it into a contingency table.

Calculate Expected Frequencies: Compute the expected frequencies for each cell in the contingency table under the assumption of independence.

Compute Chi-Square Statistic: Calculate the chi-square test statistic based on the observed and expected frequencies.

Determine Degrees of Freedom: Determine the degrees of freedom for the chi-square distribution, which depends on the dimensions of the contingency table.

Interpret Results: Compare the computed chi-square statistic to a critical value from the chi-square distribution (or use a p-value). If the computed chi-square value exceeds the critical value (or if the p-value is less than a chosen significance level, commonly 0.05), then you reject the null hypothesis, meaning the categories are significantly different form each other, else assume no statistical significance.

Example:

Chi-sq	**15.70**	
DoF	**6**	
p value	**0.02**	⇐ A p-value of < 0.05 signifies a difference in values that is not simply related to random variation. (The probability of it being random is less than 5%).

		Discrete y		Proportion	
		Satisfied	Dissatisfied	Satisfied	Dissatisfied
Discrete x	6 AM - 7 AM	48	112	30.0%	70.0%
	7 AM - 8 AM	26	82	24.1%	75.9%
	8 AM - 9 AM	12	47	20.3%	79.7%
	9 AM - 10 AM	5	9	35.7%	64.3%
	10 AM - 11 AM	4	7	36.4%	63.6%
	11 AM - 12 PM	22	44	33.3%	66.7%
	12 PM - 1 PM	21	18	53.8%	46.2%

Disclaimer:

This article provides an introductory overview of the Analyze phase within the DMAIC (Define, Measure, Analyze, Improve, Control) methodology. Successful execution of DMAIC projects often requires the use of specialized software tools such as Minitab, SPC Excel, and others. While these tools are essential for data analysis and project management in Six Sigma and process improvement initiatives, this article does not delve into specific software applications or their functionalities.

The focus of this article is to provide a conceptual understanding of the Analyze phase, including key principles, techniques, and approaches used to analyze data and identify root causes. Practical implementation of DMAIC projects may involve utilizing appropriate software tools to facilitate statistical analysis, process mapping, and visualization of results.

Readers interested in implementing DMAIC projects are encouraged to explore and leverage relevant software tools that support data-driven decision-making and continuous improvement efforts within their organizations.