LTIMindtree

Point of View

# Enabling Faster Business Decisions

## Democratization of Statistical and Predictive Analysis

**Author:** Shreya Tiwari, Sanjivani Jadhav

# Introduction

As an organization or system, we continuously need to track project parameters and take corrective action if any outliers are spotted. Of late, managers have been improvising their processes and focusing on data collection to perform various analyses, such as trend analysis, root cause analysis, and more, to avoid recurring issues.

However, for a manager to make critical decisions, it is beneficial to perform predictive and statistical analysis, benchmarking their performance. By benchmarking, managers can create a wealth of information for their organization that can be critical while responding to RFPs and managing contract renewals. The entire process requires a lot of study, analysis, and understanding of statistics, which is highly technical and effort-intensive, making it a significant challenge for organizations. The top-down approach works well for a strategic initiative, but it can be augmented with a bottom-up strategy to drive continuous performance improvement through statistical analysis. And that is where methodologies, such as CMMI and Six Sigma, have created a lot of impact in organizations to create high maturity models.

# Defining High Maturity Models

Before we delve into how we implement a bottom-up approach to democratize statistical analysis, enabling faster decisions and deploying high maturity models, let us quickly look at what is considered a high maturity model according to CMMI.

CMMI- The Capability Maturity Model Integration, a performance improvement approach, is a process model that provides a clear definition of what an organization should do to promote behaviors that lead to improved performance. CMMI can be used as a benchmark to measure the maturity of an organization's software process. Levels 4 and 5 of the CMMI are considered '**High Maturity**' and are predominantly characterized by quantitative improvement.

**High maturity means:**

- The process is data-driven, managed, and controlled by quantitative objectives.
- The correct measurement data is selected.
- The focus is on continuous process improvement, which includes analyzing variations, having the ability to derive the impact of a change, and knowing what to do with data to support effective decision making, thereby improving performance, stability, and predictability.

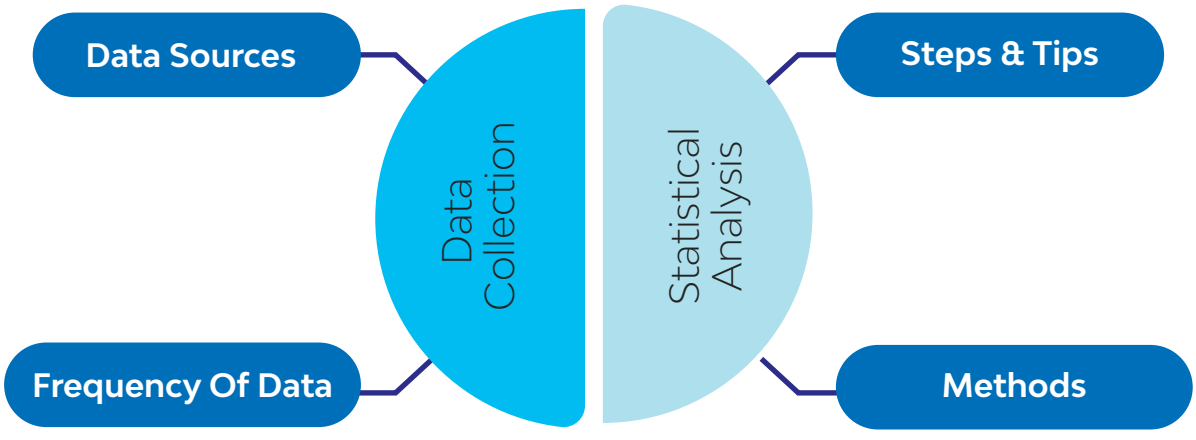
As stated above, being data-driven and continuously improving processes requires a lot of proactive and predictive analysis, which effectively applies statistical concepts. So, it becomes essential to discuss statistical analysis, why it is crucial, and how it helps managers make faster decisions.

# What is statistical analysis?

> *Statistics is about the development of methods for the collection and analysis of data in order to answer specific questions in an unbiased way, so that the conclusions depend only on the data and not on any preconceived ideas.*
>
> ***- Bryan Manly,** author of **Multivariate Statistical Methods, Fourth Edition***

## Statistical tests are essential for a CMMI High Maturity (HM) compliant project or organization, helping in:

- Discovering crucial measures within the data, such as the 'mean' or average.
- Summarizing and presenting the data in a graph or chart to present key findings.
- Calculating if the data is clustered or spread out, allowing us to make educated guesses, assumptions, and hypotheses.
- Making future predictions based on past behavior.



# Why should a project manager opt for statistical analysis? And how does statistical analysis help in enabling faster decisions?

Statistics enable organizations to predict future trends, optimize operations, and gain actionable insights. Most business-based decisions need to be backed by metrics, facts, or figures supporting the organization's aims, goals, or initiatives, providing a stable backbone for management reports and business operations.

**Let's look at some examples where a project manager would leverage statistical analysis:**

- Know how much variation the project process can handle without severing SLAs or missing deadlines.
- Learn what is required to achieve Six Sigma benefits.
- Understand the productivity of the team and analyze their performance against defined benchmarks.
- Perform 'what-if analysis' and assess the impact of changes.

**Further, let's discuss the steps involved in the statistical analysis process, critical for driving informed decision-making:**

## Raw Data

Raw data gathered from disparate sources, not received in a standardized template, and has not been processed yet by a machine or human. This data is further processed and analyzed to gain in-depth insights

## Histogram

Raw data gathered from disparate sources, not received in a standardized template, and has not been processed yet by a machine or human. This data is further processed and analyzed to gain in-depth insights

## Stability Test

Process stability for the selected parameter (e.g., story point, mean time to resolve/respond, effort variance, and more) is tested with the help of control charts and boxplots.

01 **Control charts:** Used to study how a process changes over time, to identify trends and shift in data. It is also used for outlier analysis, which is the removal of out-of-control points.

02 **Boxplot:** Used to assess and compare distribution characteristics, such as median and range, and identify outliers.

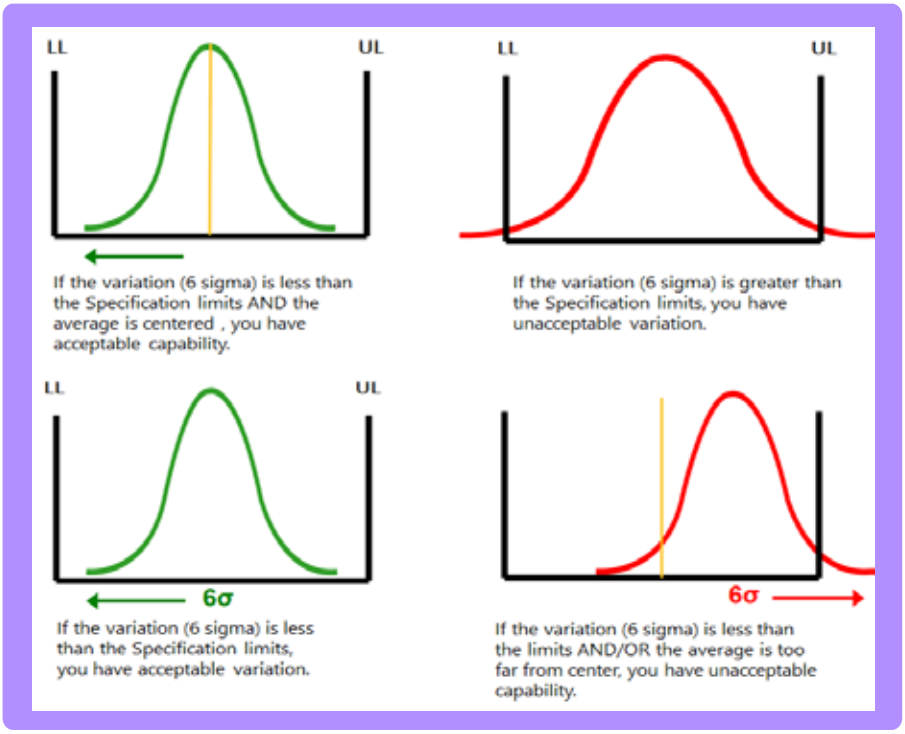
**The Goodness of Fit (GoF) Test:** This test helps determine which distribution does the data fit. The GoF Test also gives percentiles that indicate the percentage of data points (story point, MTTR, etc.) that fall below a specific data point. For example, if a story point 3.4 lies at the 75$^{th}$ percentile, it is higher than 75 percent of other story points or lies in the third quartile.

**Hypothesis Test:** Hypothesis testing is an approach for analyzing data. It helps managers understand if the effect they think they observe in the data is real. Hypothesis testing is vital in quality improvement to assess if the change made to the process creates a meaningful difference in the output.

**Capability Test:** It helps assess whether a process is statistically able to meet a set of specifications. If the process is stable and all the data points are well within the specification limit, then the process is capable.

As per the Six Sigma Institute,

Process Sigma is a measurement yardstick to evaluate the output of a process against the set performance standard. The higher the Process Sigma, the better is the process capability.




**LL:** Lower Limit
**UL:** Upper Limit.

**Performance Test:** This test compares the previous project performance with the current performance. This comparison aims to understand whether there is any improvement in the performance or deterioration.

**Predictive Analysis:** The historical data, i.e., data collected through past performances, is used to build a predictive model, which is applied to the current data to predict future performance or to suggest actions to take to drive optimal outcomes.

**Hence, by following the steps mentioned above, a Project Manager can derive valid inferences using insights from the output of the process, which can help identify areas of improvement and make mid- to long-term decisions that improve project performance and create business value.**

# Monitoring and sustaining project performance:

For a project to sustain its performance, we believe it needs to establish some "baseline" against which to monitor efforts during the project's duration. The purpose of monitoring is to recognize the long-term trends in performance, deducing where we will end up staying at the present performance level or how long it will take for us to reach there. Monitoring calls for baselining or benchmarking the data of a project.

A **baseline** is a fixed reference point that standardizes a project's performance at any given point of time to measure and compare the project's progress against that reference point. That helps to improve processes by easily smashing bottlenecks, spotting potential problems, and identifying improvement areas.

A baseline serves as the basis for further development of the project. For baselining or benchmarking the data, there are various tools available in the market, which are **menu-driven** and **code-driven.** To conduct a benchmark analysis using such tools, we should keep the following in mind:

- The user is trained in in statistics and has programming skills or statistical tools to perform and infer the analysis. That implies there is a dependency on statistical subject matter experts.
- Tools come with high license costs leading to limited licenses at the organization level, thereby limiting usage.
- Statistical analysis is an iterative process, and when done, using tools will require a lot of effort.
- Managers will spend significant effort on complex baseline steps.

- There can be several data errors, and there will be a requirement of considerable manual effort for data consolidation.
- Unless everyone gains statistical skills, a bottom-up approach is not achievable.

These pointers above delay the entire benchmark analysis process or may act as hurdles in analyzing the project performance. Now let's talk about the next possibility.

## What if we automate the entire baselining process with predictive analysis?

In our opinion, we can build an automated tool where the user has to provide raw data and go through a few steps to categorize the data (e.g., divide data depending on the phase, period, technology, target, lower limit, upper limit, etc.). Next, in just one click, the user can generate graphs, statistical insights, and visual features to understand the project performance. That would help in deriving rational inferences and further discover underlying patterns and predict project performance.

As stated in 'The Age Of Analytics: Competing In A Data-Driven World' report by Mckinsey & Company:

*Analytics capabilities are already leading to new business models and reshaping industry competition. These capabilities have become a differentiating factor in industry competition, as leading players use data and analytics to grow revenue, to enter or even create new markets, to change the nature of their relationship with customers, and to increase organizational efficiencies. Organizations that are lagging will need to adapt quickly before the gap grows wider.*

# Conclusion

Gone are those days when managers relied on intuition, their experience in similar circumstances, or others' advice to make big decisions. Today, backing every business decision requires reliable and accurate empirical data. That is where statistical analysis plays a crucial role in running businesses effectively and bolstering management capabilities. It can help improve any project's performance or find out the optimum performance capacity, enable efficient management of work and employee performance, limit the wastage of resources, and more. And automating statistical analysis will facilitate a robust bottom-up approach for decision making, which is the best way to accelerate an organization's growth strategy.

**Together, the high maturity processes deliver improved business capabilities, helping enterprises achieve their quantitative objectives for quality and process performance and placing them in the catbird seat.**

# References:

**01** ***Mckinsey & Company*** *(2016) THE AGE OF ANALYTICS: COMPETING IN A DATA-DRIVEN WORLD, McKinsey Global Institute: Available at: https://www.mckinsey.com/~/media/McKinsey/Industries/Public%20and%20Social%20Sector/Our%20Insights/The%20age%20of%20analytics%20Competing%20in%20a%20data%20driven%20world/MGI-The-Age-of-Analytics-Full-report.pdf*

**02** ***Six Sigma DMAIC Process -*** *Measure Phase - Process Capability, Available at: https://www.sixsigma-institute.org/Six_Sigma_DMAIC_Process_Measure_Phase_Process_Capability.php*

# Author

**Shreya Tiwari**
Software Engineer, LTIMindtree

Shreya has 2+ years of experience in software development and is currently working with LTIMindtree's Automation IP Development team. Being an integral part of the BGenie team, she has gained significant insights into exploratory analysis and explored various aspects of data science and machine learning using R. Her other areas of expertise include Java, MySQL, and Liferay. Further, Shreya is continuously working towards enhancing her knowledge of machine learning and other intelligent technologies.

**Sanjivani Jadhav**
Software Engineer, LTIMindtree

Sanjivani has 2+ years of experience in software development and is a part of LTIMindtree's Automation IP Development team. Being an integral part of the BGenie team, she has gained hands-on knowledge of statistical analysis and honed her statistical skillset to deliver superior value to clients. Java, MySQL, Spring Hibernate, and Liferay are her other areas of expertise, and she is continuously strengthening her statistical computing and graphics capabilities in the R environment.