Quality Control in Beverage Production – Cap Defects Analysis Using an Integration of R-Studio in Power BI

Espitia-Rodriguez, Arcesio. [1[0009-0006-9490-2620]] Patiño Rodriguez, Carmen Elena. [2]

[1] University of Antioquia, Medellín 050010, Colombia
arcesio.espitia@udea.edu.co
[2] University of Antioquia, Medellín 050010, Colombia
elena.patino@udea.edu.co

Abstract. This case study, developed in the Standardization and Quality Control 2025-1 course, analyzed the capping and sealing process in beverage production under Quality 4.0. Using SIPOC, key variables were defined: defective caps (discrete) and closing turns (continuous). Data collection stage enabled BI control charts, p-charts, and Gage R&R by attributes with risk analysis method for assess measurement reliability. From 1,978 caps across 56 batches with 3 operators, high defect rates linked to raw material variability, poor inspection, and unstandardized procedures were found. Corrective actions included stricter inspections, supplier renegotiation, and process standardization. Integrating R-Studio with Power BI enabled real-time dashboards for defect monitoring, supporting data-driven decisions. This approach demonstrates how statistical methods and BI tools enhance quality control, offering potential for predictive analytics, anomaly detection, and broader industrial applications.

Keywords: Quality 4.0, Statistical Process Control, Digitalization, Control Charts, Power BI, R-Studio

1 Introduction

Global manufacturing industries are increasingly adopting the principles of Quality 4.0, which emphasize digital integration, advanced analytics, and real-time monitoring. Traditional statistical tools, such as Statistical Process Control (SPC), remain indispensable but require alignment with modern Business Intelligence (BI) solutions to address the current complexity of industrial operations.

The beverage industry is particularly sensitive to packaging and sealing processes, where cap defects can compromise product safety, consumer trust, and regulatory compliance. While control charts, Gage R&R, and measurement system analysis have historically been applied, their integration into real-time BI systems remains underexplored.

This study advances the field of Quality 4.0 by demonstrating a reproducible, statistically rigorous integration of R-based quality control methods within an enterprise business intelligence platform (Power BI). Unlike traditional SPC workflows that are executed exclusively in statistical software or custom dashboards, the proposed framework embeds R's classical quality control tools, such as p-charts, operating characteristic curves, and attribute-based R&R, in a real-time, industry-standard BI environment. This fusion enables organizations to adopt advanced statistical techniques without migrating to R-native ecosystems such as Shiny.

From a scientific standpoint, the contribution lies in showing how R's computational capabilities can be operationalized as an analytical engine inside Power BI, creating a hybrid architecture that supports real-time SPC monitoring, reproducible analyses, and traceable statistical computation. The work also formalizes the theoretical underpinnings of the analyses used, linking Statistical Process Control, Measurement System Analysis, and reliability modeling, with the technological enablers of Quality 4.0 (connectivity, interoperability, and digital integration).

This positions the study not merely as an applied case, but as a methodological template for how statistical computing languages can be embedded inside BI ecosystems to expand analytical depth while maintaining industrial usability.

2 Bibliometric analysis

In this Section, we provide a bibliometric analysis of the reviewed papers. To structure the bibliographic search in Scopus, the query was divided into three main criteria blocks that reflect the conceptual, technological, and industrial dimensions of the research. The first block addresses the classical statistical quality methods that form the theoretical backbone of quality control. The second block emphasizes the integration of business intelligence and digital tools, particularly Power BI and R-Studio, as enablers of Quality 4.0. Finally, the third block narrows the focus to the industrial application context, with special emphasis on the beverage and packaging sector where capping and sealing defects occur. Table 1 summarizes the query strings and scope of each block.

Table 1. Query string for bibliography search

Criteria block	Query string	Scope of search
Quality & Statistical Methods	"Quality 4.0" OR "digital quality management" OR "statistical process control" OR "SPC" OR "Gage R&R" OR "measurement system analysis" OR "Cohen's Kappa" OR "control charts"	Captures literature related to quality management paradigms, SPC techniques, and measurement system analysis as classical foundations of quality control.
Business Intelligence & Digital Tools	"Power BI" OR "business intelligence" OR "dashboard" OR "real-time monitoring" OR "data visualization" OR "predictive quality" OR "machine learning" OR "R-Studio" OR "R language"	Focuses on modern BI and digital analytics tools (Power BI, R, dashboards, ML) enabling Quality 4.0 integration and real-time process monitoring.

Industrial Application Context	"beverage industry" OR "food and beverage" OR "manufacturing process" OR "packaging" OR "sealing process" OR "capping" OR "quality defects" OR "supply chain quality"	Narrows down to industrial applications, especially the beverage/food sector, packaging, capping/sealing defects, and supply chain quality issues.

To better understand the scientific landscape surrounding Quality 4.0, SPC, and digital integration (Power BI, R, machine learning), a bibliometric analysis was conducted using VOSviewer. The analysis focused on keyword co-occurrence extracted from Scopus-indexed publications between 2000 and 2024.

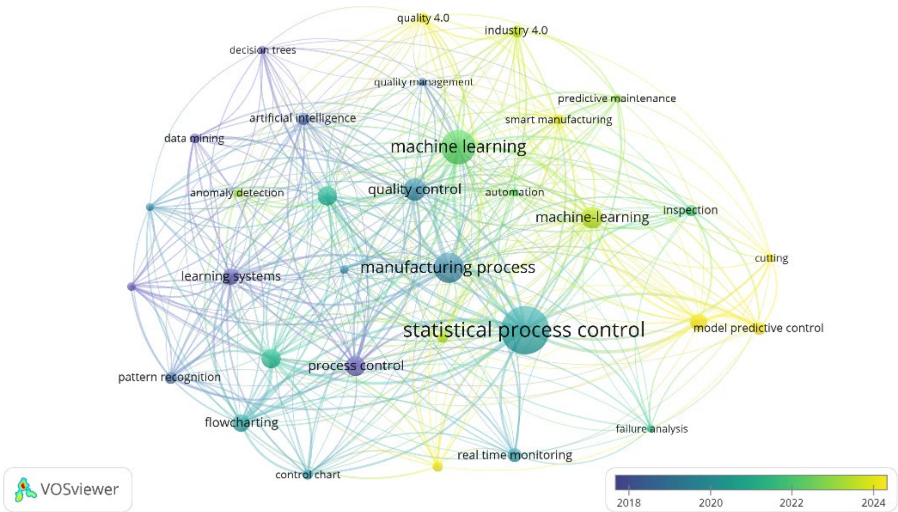


Fig. 1. Bibliography analysis

The bibliometric evidence reinforces that SPC remains the backbone of quality management research; there is a clear trend toward AI, predictive models, and real-time monitoring, aligning with Quality 4.0 principles; the integration of BI tools (Power BI + R-Studio) into manufacturing quality remains underexplored, validating the novelty of this study.

This bibliometric analysis was included to establish the conceptual landscape that frames this study, specifically highlighting the growing convergence between classical SPC, digital quality systems, and analytics platforms. By mapping the evolution of Quality 4.0 literature, the section contextualizes the relevance of integrating R into BI environments and clarifies the research gap that motivates this work.

3 DMAIC Methodology and Data Validation

The statistical procedures implemented in this study build directly on foundational SPC theory from Montgomery (2019) and Wheeler (2010), ensuring that the analytical workflow adheres to established best practices. The p-chart methodology follows the binomial-based control structure for attribute data, using subgroup sizes and defect counts to estimate process stability and inherent capability. The OC curve computation relies on the theoretical relationship between sample size, process proportion nonconforming, and the probability of detecting assignable causes.

Measurement System Analysis follows ISO 22514-7 and AIAG MSA guidelines, using inter-observer agreement (Cohen's Kappa) as the inferential measure of attribute repeatability and reproducibility. The computational execution of these procedures is handled entirely through R, ensuring that the statistical calculations follow open-source standards rather than proprietary BI algorithms.

The project was developed following the DMAIC methodology (Define, Measure, Analyze, Improve, Control) according to the following guidelines: (Park, 2003):

- **Define**. Identification of the process or product that needs improvement.
- **Measure**. Identify those characteristics of the product or process that are critical to the customer's requirements for quality performance, and which contribute to customer satisfaction.
- **Analyze**. Evaluate the current operation of the process to determine the potential sources of variation for critical performance parameters.
- **Improve**. Select those product or process characteristics which must be improved to achieve the goal. Implement improvements.
- **Control**. Ensure that the new process conditions are documented and monitored via statistical process control methods (SPC).

Depending on the outcome it may become necessary to revisit one or more of the preceding phases.

Together, these foundations situate the study within classical SPC and MSA theory while leveraging modern data workflows enabled by Quality 4.0 technologies.

3.1 Define and Measure Phase: Variables and Baseline

Beverage production is divided into four major macro-processes, which are indicated simultaneously in Figure 2.

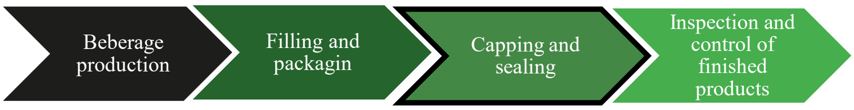


Fig. 2. Beverage production macroprocess

The Definition and Measurement phase began with a SIPOC diagnosis, which describes the process carried out in the covering and sealing of the subsequent identification of the quality variables of interest.

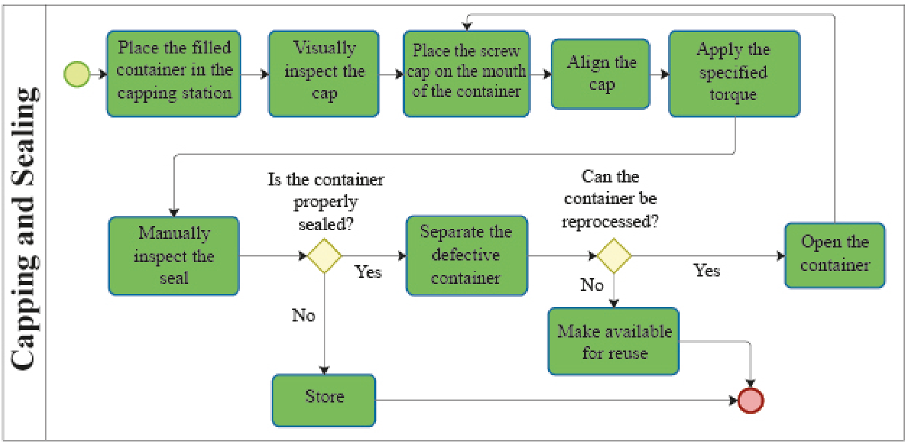


Fig. 3. Capping and sealing process

- **Discrete variable (Attributes):** Cap not properly fitted (Yes/No). This variable affects the Reliability quality dimension.
- **Continuous variable:** Closing torque (Nm) or Number of turns to close (Continuous).

The analysis focuses on the capping and sealing process, illustrated in the flowchart, which details the sequence from placing the filled container to verifying the sealing quality and managing defective units. Two critical variables are monitored to ensure process reliability. The discrete variable, Cap not properly fitted (Yes/No), reflects whether the sealing meets the required reliability standard, serving as an attribute-based indicator of product conformity. Meanwhile, the continuous variable, closing torque (Nm) or Number of turns to close, provides quantitative control over the precision of the capping operation. Together, these variables enable a comprehensive quality evaluation of the process, linking operational consistency with reliability outcomes as shown in the procedural flow.

To establish the baseline and monitor the proportion of defective units, a set of exploratory visualizations was developed to describe the initial defect patterns observed during the preview stage. These graphical tools (boxplot, histogram, and dotplot) were generated using R within the Power BI environment to visualize the dispersion, central tendency, and distribution of cap defect data across the first 16 batches. Together with the descriptive data table, these visualizations provide an initial diagnostic view of process stability, operator consistency, and potential improvement areas before implementing more advanced statistical controls.

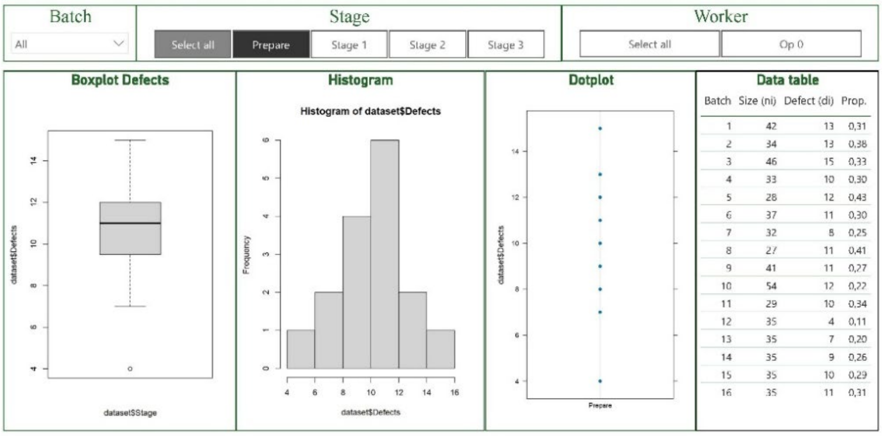


Fig. 4. Distribution chart for the proportion of defective caps in the preview stage

In the first stage of data collection, where an average of 36.13 caps were analyzed for each of the 16 batches, a moderately consistent process with some variability in defect proportions was revealed, centered around a median of approximately 0.32. The histogram indicates a unimodal, slightly right-skewed distribution, suggesting sporadic increases in defects tied to material or operational inconsistencies. Early batches show higher defect ratios (up to 0.43), while later ones exhibit improvement, potentially due to operator adaptation or initial process adjustments.

Overall, the process displays emerging stability with indications of learning and quality enhancement, warranting continued monitoring through control charts to confirm sustained improvement.

3.2 Validation of the Measurement System (R&R by Attributes)

The rapid advancement of technology has left a knowledge gap for manufacturing and process industries that do not have state-of-the-art technologies or that, by their nature, do not produce the quantity and variety of data needed to apply models that require high volumes of data.

Validation of the inspection system was crucial to ensure that decisions (accept/reject) were accurate, reducing the subjectivity associated with attribute-based evaluation. A Gage R&R analysis by attributes was performed using the risk analysis method, supported by contingency tables and Cohen's Kappa index.

Fig. 5. Risk analysis method

The study included two operators, evaluating 45 parts with five measurements (or trials), resulting in a total of 225. Cohen's Kappa index is a measure of agreement that compares the observed agreement with that which could occur by chance. In this case, the results of the R&R analysis were summarized in a contingency table shown in Table 1, the observed agreement was Po = 0.947, and the Kappa agreement was kappa = 0.787

Table 2. Contingency table

Contingency table		**Observer 2**		**Total**
		Satisfied	Not Satisfied	
Observer 1	Satisfied	186	8	194
	Not Satisfied	4	27	31
Total		35	190	225

4 Analysis and Results

The results of the monitoring stage carried out by the two operators in charge provided a preliminary P-chart based on the 21 samples inspected, which did not identify any patterns of special causes (out-of-control points or trends). However, the control chart for the proportion of defective caps indicated a high average percentage of defective caps. This implies that, although the process is stable, it is inherently incapable (consistently poor) due to the variability of the raw material.

At the same phase, the main assignable causes of variation were identified, including shortages in raw material supply, inadequate inspection processes—which justified the application of the R&R study—and the presence of unstandardized procedures.

Data Analytics (DA) is identified as a key enabler and fundamental success factor for the transformation of manufacturing processes, especially in the context of Quality 4.0, for this one in the next session are described the steps to control and stabilize the procedure of described.

5 Improvement, Control, Quality 4.0

In order to improve the performance of the capping process, various interventions were implemented to reduce process variability. These included, but were not limited to, increased inspection procedures prior to sealing, renegotiation with suppliers to ensure higher quality materials, as well as the adoption of waste practices and standardization of procedures to reduce variability.

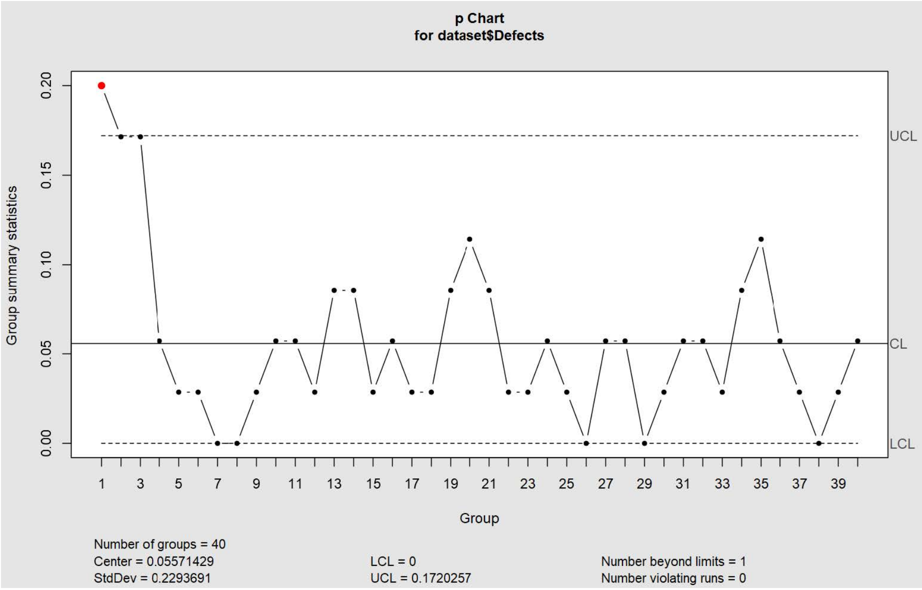


Fig. 6. P-chart proportion of defective caps monitoring stage

During the monitoring stage, 40 groups were sampled with an average batch size of 35 units. The process center is located at 0.055 (5.5%), with significant variability (σ = 0.22). One point is observed outside the control limits (group 1, in red), indicating the presence of a special cause of variation. In addition, there is a low number of cases in the proportion of defective items in the groups, which suggests that an increase in the size of these monitoring groups is necessary.

Integration and Continuous Monitoring (Control)

The control phase focused on implementing a continuous monitoring system. To do this, R-Studio and Power BI were integrated under the Quality 4.0 paradigms. The integration of these two tools made it possible to:

1. Dynamic visual representation of process metrics and defect trends.
2. A real-time monitoring dashboard, accessible from multiple devices (mobile-friendly).
3. Facilitate faster, more informed decision-making by combining business intelligence with interconnectivity.

This application aligns with the concept that computer systems specialized in statistics facilitate descriptive analysis and the application of methods. Framework illustrating the integration of classical quality tools such as Statistical Process Control, operations curves and so on, with modern digital platforms (R-Studio & Power BI) for real-time monitoring.

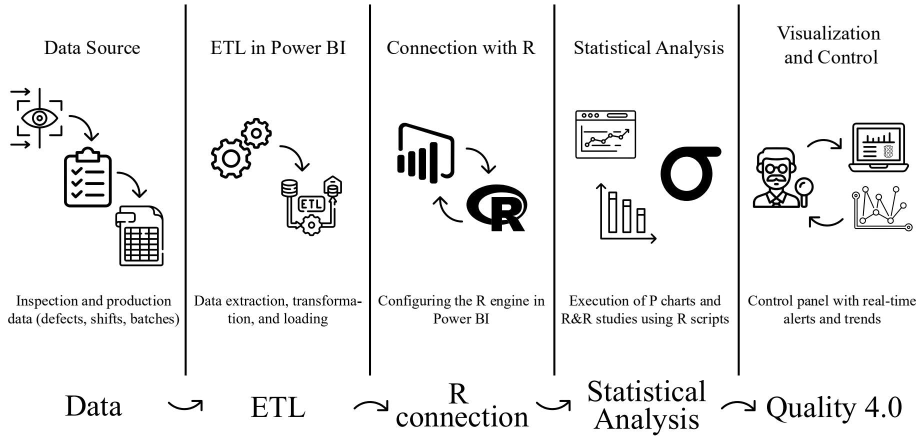


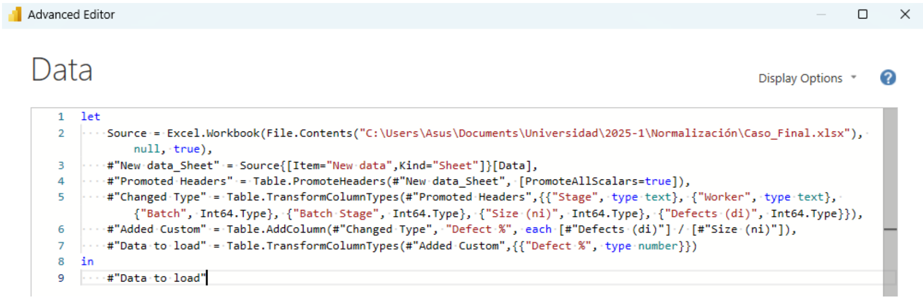
Fig. 7. Power BI–RStudio integration flow for statistical monitoring

Figure 6 illustrates the sequential workflow used for the data integration and statistical control process in the beverage capping study. It begins with the data acquisition and preparation phase, where inspection records are cleaned and structured in Power BI through ETL (Extract, Transform, Load) operations. The process continues with the connection and configuration of R within Power BI, enabling the execution of R scripts for statistical computation. Subsequently, the control chart generation and defect analysis are performed using R code directly embedded in Power BI visuals, allowing dynamic interaction with the underlying data. Finally, results are visualized in real-time through dashboards, enabling ongoing process monitoring and facilitating rapid decision-making.

Data Source and ETL

In Power BI, a data source is the entity from which data is extracted for creating reports and visualizations. It can be a file, webpage, database, or another app/service. The ETL process refers to a series of steps to extract data from numerous sources, transform that data according to business rules or analytical needs, and finally load it into a system where it can be used for analysis and decision-making.

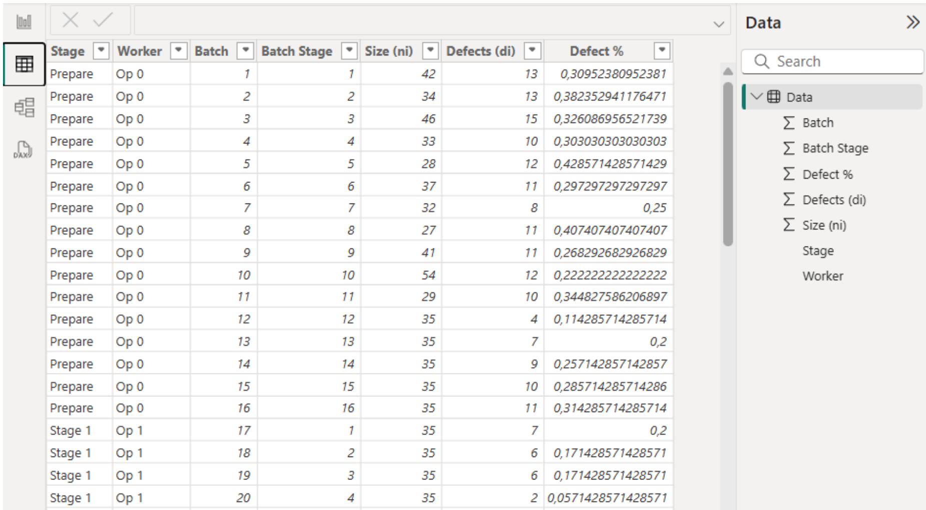
For this case study, we used a flat Excel file containing information about the stage, workers, group labels, and defects, i.e., the number of parts inspected, and the total number of defective parts found in that specific group. During the transformation process, the columns of the file used were configured and a new column was added to calculate the proportion of defective parts in each group. Finally, the final data is loaded to create the visualizations in Power BI.




```
let
    Source = Excel.Workbook(File.Contents("C:\Users\Asus\Documents\Universidad\2025-1\Normalización\Caso_Final.xlsx"),
        null, true),
    #"New data_Sheet" = Source{[Item="New data",Kind="Sheet"]}[Data],
    #"Promoted Headers" = Table.PromoteHeaders(#"New data_Sheet", [PromoteAllScalars=true]),
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"Stage", type text}, {"Worker", type text},
        {"Batch", Int64.Type}, {"Batch Stage", Int64.Type}, {"Size (ni)", Int64.Type}, {"Defects (di)", Int64.Type}}),
    #"Added Custom" = Table.AddColumn(#"Changed Type", "Defect %", each [#"Defects (di)"] / [#"Size (ni)"]),
    #"Data to load" = Table.TransformColumnTypes(#"Added Custom",{{"Defect %", type number}})
in
    #"Data to load"
```

Fig. 8. Data extraction, transformation and loading script in Power Query (Power BI)

Figure 8 shows the Power BI Advanced Editor containing an M language script used to perform the ETL (Extract, Transform, Load) process. In this code, data is imported from an Excel file, column headers are promoted, data types are assigned, and a calculated column named "Defect %" is created by dividing the number of defects by the sample size. This script ensures that the dataset is properly structured and ready for analysis within Power BI. The first rows of the data table loaded into the model in Power BI are shown in Figure 9. The data includes key variables such as stage, worker, batch, sample size (ni), defects (di), and defect percentage (%).



Stage	Worker	Batch	Batch Stage	Size (ni)	Defects (di)	Defect %
Prepare	Op 0	1	1	42	13	0,30952380952381
Prepare	Op 0	2	2	34	13	0,382352941176471
Prepare	Op 0	3	3	46	15	0,326086956521739
Prepare	Op 0	4	4	33	10	0,303030303030303
Prepare	Op 0	5	5	28	12	0,428571428571429
Prepare	Op 0	6	6	37	11	0,297297297297297
Prepare	Op 0	7	7	32	8	0,25
Prepare	Op 0	8	8	27	11	0,407407407407407
Prepare	Op 0	9	9	41	11	0,268292682926829
Prepare	Op 0	10	10	54	12	0,222222222222222
Prepare	Op 0	11	11	29	10	0,344827586206897
Prepare	Op 0	12	12	35	4	0,114285714285714
Prepare	Op 0	13	13	35	7	0,2
Prepare	Op 0	14	14	35	9	0,257142857142857
Prepare	Op 0	15	15	35	10	0,285714285714286
Prepare	Op 0	16	16	35	11	0,314285714285714
Stage 1	Op 1	17	1	35	7	0,2
Stage 1	Op 1	18	2	35	6	0,171428571428571
Stage 1	Op 1	19	3	35	6	0,171428571428571
Stage 1	Op 1	20	4	35	2	0,0571428571428571

Fig. 9. Structured dataset loaded and processed in Power BI

Connection with R

The following configuration window represents the initial setup of the R integration within Power BI, a fundamental step that enables the execution of statistical analyses and the creation of advanced visualizations directly within the BI environment. Through this interface, the R home directory and the preferred R Integrated

Development Environment (IDE)—in this case, RStudio—are defined, ensuring that Power BI correctly identifies the installed R engine.

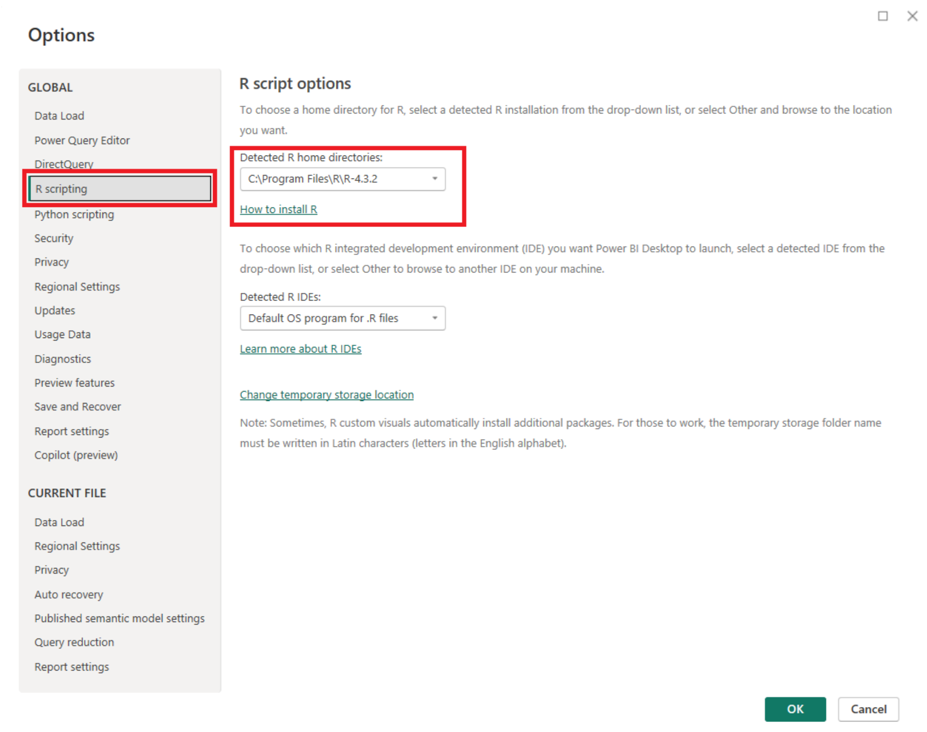


Fig. 10. Configuration of R scripting environment in Power BI

This integration enables the creation of data-driven control charts and quality diagnostics directly within Power BI, enhancing analytical depth, reproducibility, and the visualization of complex statistical models in a clear, dynamic, and research-oriented environment.

Statistical Analysis

Once the data has been loaded into Power BI and the integration with RStudio has been completed, the statistical analysis process begins executing R scripts for statistical computation.

*OC Curves Script***.** Creating statistical graphs in Power BI using R integration. The red boxes highlight key components: on the right, the "R script visual" icon is selected from the visualization pane, and relevant variables (*Batch*, *Size*, and *Defects*) are assigned to the Values section to feed the R script. At the bottom, the R script editor contains the code that loads the qcc library, defines a *p*-type control chart, and generates the Operating Characteristic (OC) curve using the oc.curves() function. The resulting plot, shown at the top, represents the probability of detecting process variations for a sample size of 35, integrating R's analytical power directly into Power BI's visualization environment. This group size parameter can be made dynamic using DAX language.

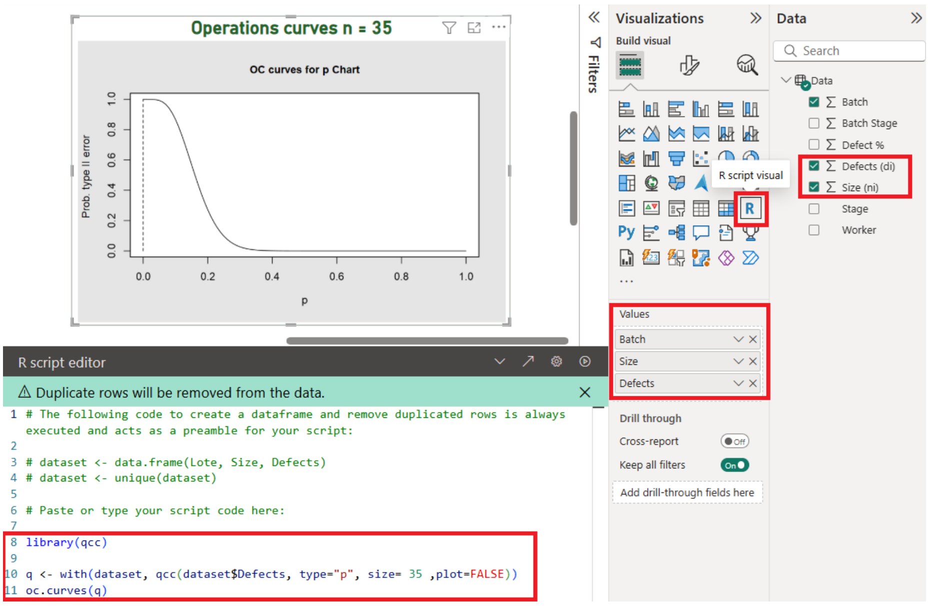


Fig. 11. Generation of OC Curves in Power BI Using R Script

*P-Chart Script***.** The figure shows a screenshot of the R script integrated into Power BI, used to generate a p-type control chart using the qcc library. This chart allows you to monitor the proportion of defective units in the process, taking the number of defects and sample size as variables. Its implementation within Power BI facilitates real-time statistical quality analysis, integrating visualization and process control directly into management dashboards.

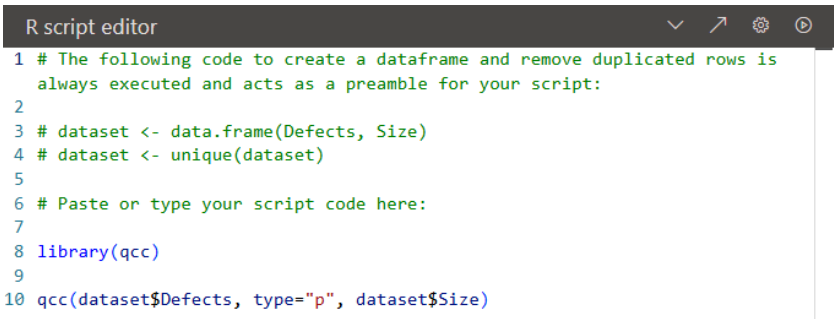


Fig. 12. Configuration of R script for p-chart generation in Power BI

Visualization and Control Dashboard

Figure 13 illustrates how statistical monitoring and BI dashboards enable you to obtain predictive information and make decisions more quickly, combining the capabilities of Power BI with the statistical tools of R-Studio.

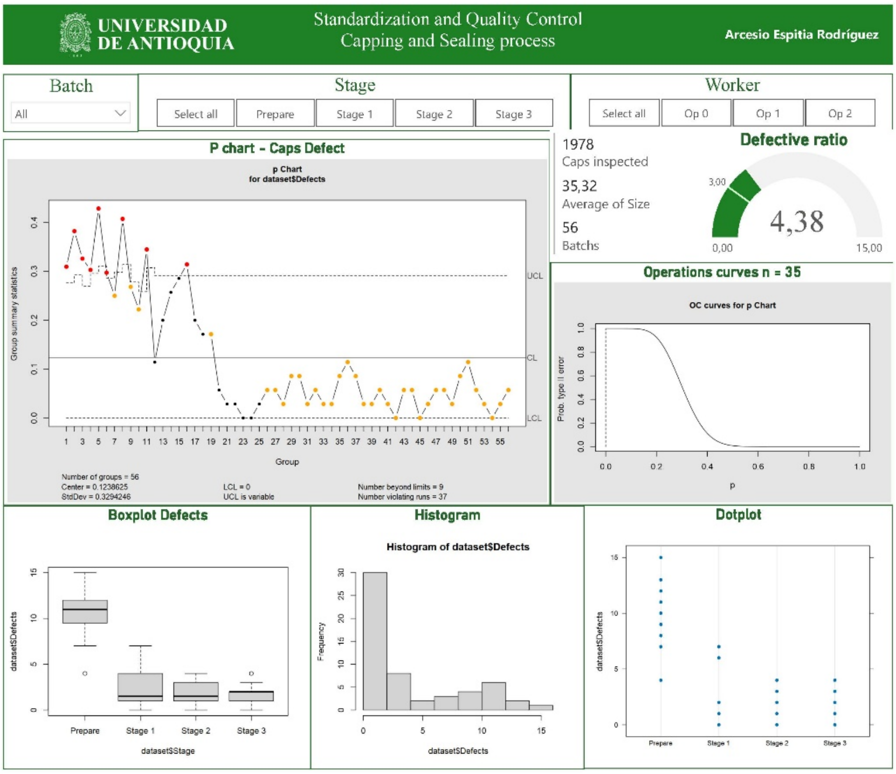


Fig. 13. Power BI dashboard for standardization and quality control in the capping and sealing process

This dashboard integrates Power BI with R scripting to visualize and analyze the capping and sealing process through control charts, descriptive statistics, and defect distribution graphs. It provides a comprehensive overview of process stability, highlighting variability across stages and identifying potential sources of defects. Future studies should integrate real-time data monitoring and augmented intelligence systems, combining machine learning predictive models with human expertise within Power BI to anticipate deviations before they occur and support proactive decision-making. Additionally, incorporating torque values and sealing parameters into the analysis would enhance process capability evaluation and drive continuous improvement through a more intelligent, data-driven approach to quality control.

6 Data availability

To ensure methodological transparency, the full R scripts used in this study, including p-chart generation, OC curve computation, Gage R&R analysis, data preprocessing and PBIX file, are provided as supplementary material in a public repository at https://github.com/arcesioespitia-cyber/Defects-Analysis-Using-an-Integration-of-R-Studio-in-Power-BI.git. These scripts reproduce every figure and result presented in the manuscript, allowing practitioners to adapt and apply the workflow to their own industrial data.

7 Conclusions

The novelty of this work lies in operationalizing a statistically rigorous SPC workflow within Power BI, utilizing R as the embedded computational engine. While Shiny and R Markdown offer R-native deployment solutions, they are rarely adopted in industrial environments governed by Microsoft BI ecosystems. By demonstrating that full SPC pipelines can be executed directly inside Power BI, this study bridges a critical usability gap in industrial quality analytics.

The application of the DMAIC methodology proved effective in diagnosing and improving the capping process. The attribute-based R&R study was essential for validating the reliability of the measurement system, particularly in relation to visual inspection. Furthermore, the integration of Business Intelligence platforms (Power BI) with statistical tools (R) positions this approach as a versatile digital solution that bridges traditional statistical methodologies with modern business intelligence. Looking forward, this combined application demonstrates significant potential for predictive quality control and the early detection of anomalies through the incorporation of machine learning models, thereby aligning with the principles of Quality 4.0.

References

1. Choudhury, A. (2022). Machine Learning and AI for Quality Improvement: Applications in Industrial Processes. Journal of Intelligent Manufacturing, 33(2), 465–478.
2. Davenport, T. H., & Ronanki, R. (2018). Artificial Intelligence for the Real World. Harvard Business Review, 96(1), 108–116.
3. ISO 9001:2015. *Quality Management Systems — Requirements*. International Organization for Standardization.
4. Microsoft Corporation. (2024). *Power BI Documentation: Integrate R Scripts in Power BI Desktop*. Available at: https://learn.microsoft.com/power-bi/
5. Montgomery, D. C. (2019). *Introduction to Statistical Quality Control* (8th ed.). John Wiley & Sons.
6. R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Available at: https://www.r-project.org/
7. Scrucca, L. (2004). qcc: An R Package for Quality Control Charting and Statistical Process Control. R News, 4(1), 11–17.

8. Shmueli, G., Bruce, P. C., & Patel, N. R. (2021). *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R*. Wiley.
9. University of Antioquia. (2025). *Standardization and Quality Control – Capping and Sealing Process Project Report* (Unpublished academic work).
10. Wheeler, D. J., & Chambers, D. S. (2010). *Understanding Statistical Process Control* (3rd ed.). SPC Press.

Acknowledgments

We would like to thank the University of Antioquia, the Faculty of Engineering, and the Department of Industrial Engineering for providing the resources and allowing us to carry out the studies conducted in this case.