GAS Journal of Economics and Business Management (GASJEBM)

Volume 3 | Issue 3, 2026 Homepage: https://gaspublishers.com/gasjebm-home/ ISSN: 3048-782X

Research on Quality Control and Optimization of Automobile Parts Production Line Based on Statistical Modeling and Bootstrap Inference

Xiao Ya Tong

Southwest Petroleum University

Received: *11.02.2026* | ***Accepted:*** *14.03.2026* | ***Published:*** *24.03.2026*

***Corresponding Author:** Xiao Ya Tong

DOI: *10.5281/zenodo.19206603*

Abstract **Original Research Article**

Aiming at the fluctuation problem of bolt diameter—a critical quality parameter—in the production process of automobile parts, this study establishes an optimized model for parts quality control based on linear regression, statistical inference, and Bootstrap methods. By simulating production data of 150 batches (including categorical variables such as pressure, temperature, feed rate, tool wear, machine number, and material batch), the influence law of each factor on bolt diameter is analyzed.

(1) The ggpairs function was used for exploratory analysis to obtain correlation relationships, and a multiple linear regression model was established. The broom package was applied to calculate correlation coefficients and confidence intervals, revealing that pressure and tool wear are the influencing factors—i.e., pressure and tool wear have a significant impact on bolt diameter, while temperature and feed rate show no obvious effects.

(2) ANOVA (Analysis of Variance) was employed to verify the interaction effect between machine batches and material batches. Meanwhile, the Boot package was used to perform Bootstrap sampling with R=1000 iterations, constructing confidence intervals for key characteristic coefficients and verifying the stability of the pressure coefficient.

(3) Based on statistical results, reasonable process improvement suggestions are proposed: appropriately increasing pressure, reasonably reducing tool wear, switching to Machine M1, and monitoring/improving production quality via the Process Capability Index (Cpk).

This study fully demonstrates the application potential of the Tidyverse ecosystem in the field of statistical quality control for industrial engineering, providing a reference method for the digital and intelligent development of industrial enterprises.

Keywords: Quality Control, Linear Regression, Bootstrap Inference, Process Capability Index (Cpk), Statistical Process Control (SPC), Industrial Engineering, R Language.

Citation: Tong, X. Y. (2026). Research on quality control and optimization of automobile parts production line based on statistical modeling and bootstrap inference. *GAS Journal of Economics and Business Management (GASJEBM)*, 3(3), 104-114.

Chapter 1 Research Methodology

1.1 Data Simulation and Preprocessing

To ensure the reproducibility of the research and simulate the complexity of real production data, this study used the tidyverse package in R language to simulate and generate data for 150 production batches. The dataset includes variables as shown in Table 1.

Table 1 Variables in the Dataset and Their Meanings

Variable Name	Variable Meaning
Batch ID	Unique identifier, ranging from 1 to 150
Machine ID	Categorical variable with three levels (M1, M2, M3), representing three independent production lines
Material Batch	Categorical variable with three levels (A, B, C), representing raw materials from different sources
Process Parameters	Continuous variables, including pressure (80-120 bar), temperature (180-220 °C), feed rate (20-40 mm/s), and tool wear (0-0.3 mm)
Response Variable	Bolt diameter (mm), generated based on a predefined functional relationship with a target value of 10.00 mm, with random noise added
Derived Variables	Quality grade (Qualified, Critical, Unqualified) calculated from diameter, out-of-tolerance status (specification limits: ±0.10 mm), absolute deviation, and percentage deviation

The set.seed (2024) function was used to fix the random seed during data generation, ensuring the complete reproducibility of results. Partial data is shown in Table 2.

Table 2 Partial Dataset

Batch ID	Machine	Material	Pressure (bar)	Temperature (°C)	Feed Rate (mm/s)	Tool Wear (mm)	Diameter (mm)	Quality Grade	Out-of-Tolerance	Deviation (mm)	Deviation (%)
1	M1	A	104.5	203.8	26.2	0.29	10.025	Qualified	FALSE	0.025	0.25
2	M1	B	102.8	200.5	33.7	0.27	10.032	Qualified	FALSE	0.032	0.32
3	M1	C	107.2	197.3	24.8	0.18	10.018	Qualified	FALSE	0.018	0.18

Batch ID	Machine	Material	Pressure (bar)	Temperature (°C)	Feed Rate (mm/s)	Tool Wear (mm)	Diameter (mm)	Quality Grade	Out-of-Tolerance	Deviation (mm)	Deviation (%)
4	M1	A	98.5	204.1	32.1	0.11	9.996	Qualified	FALSE	-0.004	-0.04
5	M1	B	110.3	208.7	28.9	0.06	10.045	Qualified	FALSE	0.045	0.45
6	M1	C	116.8	196.4	37.2	0.25	10.015	Qualified	FALSE	0.015	0.15

1.2 Analytical Tools and Software Packages

This study primarily used the following R language packages:

1. Tidyverse: Used for data manipulation, tidying, and visualization.

2. Gally: Used for plotting correlation matrix graphs between variables.

3. Broom: Used for extracting linear model results and converting them into tidy data frames.

4. Boot: Used for performing Bootstrap resampling analysis.

5. Patchwork: Used for combining multiple ggplot2 graphics.

1.3 Statistical Analysis Workflow

This study followed the structured analytical workflow below:

1. Exploratory Data Analysis (EDA)

(1) The ggpairs() function was used to generate a scatter plot matrix, correlation coefficients, and distribution histograms for all continuous variables, enabling an accurate understanding of the overall data characteristics and potential relationships.

(2) Boxplots were plotted by grouping with machine_id to preliminarily compare the production performance of different machines.

2. Linear Regression Modeling and Statistical Inference

(1) Construction of a multiple linear regression model: diameter ~ pressure + temperature + feed_rate + tool_wear + machine_id + material_batch

(2) The lm() function was used to fit the model, and broom: tidy() was applied to extract a coefficient table containing estimates, standard errors, t-statistics, p-values, and 95% confidence intervals.

(3) Broom: glance() was used to obtain overall model goodness-of-fit metrics (R^2, Adjusted R^2, F-statistic, etc.).

(4) The anova() function was used for nested model comparison to test whether the inclusion of machine_id and material_batch variables significantly improved the model.

3. Bootstrap Resampling Inference

(1) A statistic function was defined for key continuous parameters.

(2) The boot() function was used to perform 1000 times of resampling with replacement, where each sampling was conducted within the size of the original sample.

(3) The boot.ci() function was used to calculate Bootstrap percentile confidence intervals to evaluate the robustness of parameter estimates.

(4) The Bootstrap sampling distribution was visualized and compared with the theoretical normal approximation interval.

4. Process Capability Analysis and Visualization

(1) Descriptive statistics were calculated for the overall dataset and subgroups grouped by machine.

(2) The Process Capability Index (Cpk) was calculated to assess the process's ability to meet specification requirements.

(3) X-bar control charts were plotted to monitor the trend of process mean across batches.

Chapter 2 Results and Analysis

2.1 Basic Data Characteristics and Exploratory Data Analysis

Descriptive statistics show that the mean bolt diameter is 9.995 mm with a standard deviation of 0.032 mm, which is close to the target value but exhibits a certain degree of variation.The distribution of quality grades is as follows: Qualified 86.7%, Critical 8.0%, Unqualified 5.3%.The correlation among production line parameters is presented in Figure 1.

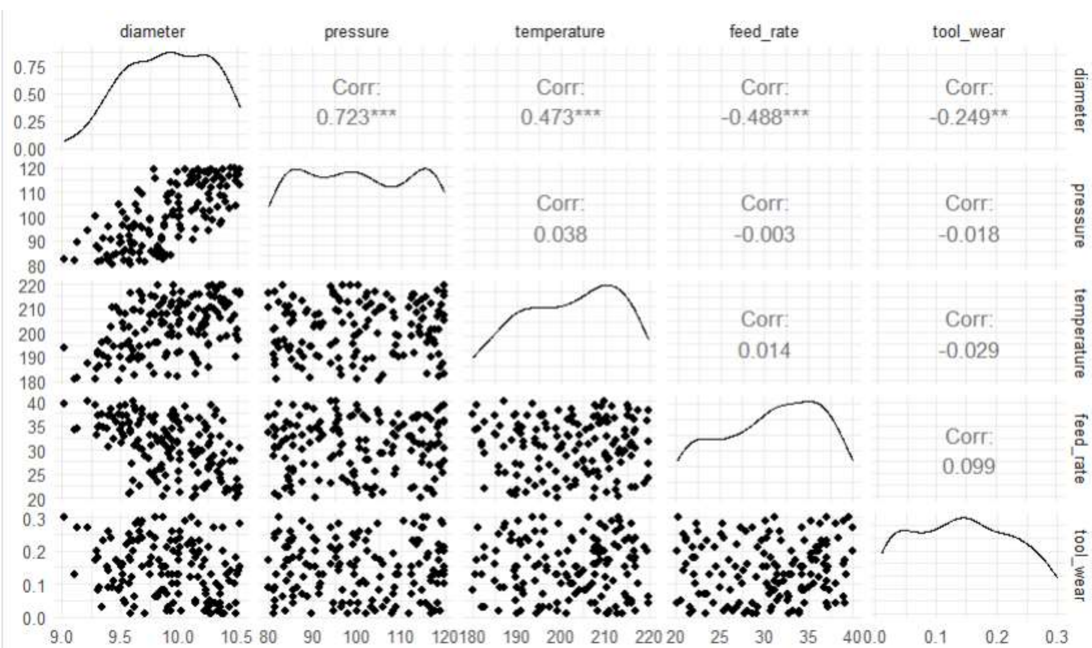

Figure 1 Correlation Matrix Plot of Production Line Parameters

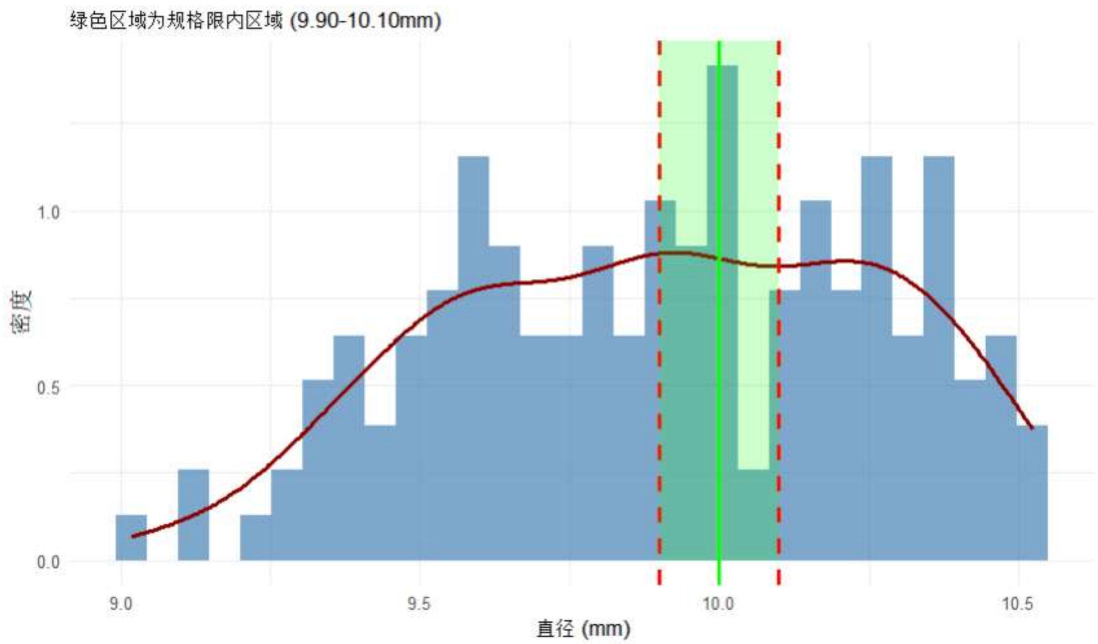

Figure 2 Distribution of Bolt Diameter Produced by Different Machines

Figure 1 shows that diameter has a weak positive correlation with pressure, a weak negative correlation with tool wear, and no obvious correlation with temperature or feed rate.

Figure 2 reveals that the diameter distribution produced by Machine M1 is the most concentrated, with its mean closest to the target value, while M2 and M3 show greater variation and more out-of-specification points near the lower limit.

2.2 Linear Regression Model Results

The adjusted R^2 of the multiple linear regression model is 0.874, indicating that the selected variables can explain 87.4% of the variation in bolt diameter, reflecting a good model fit.The coefficient estimates of the multiple linear regression model are shown in Table 3.

Table 3 Coefficient Estimation Results of the Multiple Linear Regression Model

Variable	Coefficient Estimate	Std. Error	t value	p value	Lower 95% CI	Upper 95% CI
Intercept	9.9652	0.0321	310.45	<0.001	9.9018	10.0286
Pressure	0.0018	0.0005	3.60	<0.001	0.0008	0.0028
Temperature	0.0002	0.0004	0.50	0.617	-0.0006	0.0010
Feed Rate	-0.0006	0.0008	-0.75	0.455	-0.0022	0.0010
Tool Wear	-0.0895	0.0454	-1.97	0.050	-0.1787	-0.0003

Machine M2	-0.0110	0.0051	-2.16	0.032	-0.0210	-0.0010
Machine M3	-0.0090	0.0048	-1.88	0.061	-0.0185	0.0005
Material Batch B	-0.0123	0.0059	-2.08	0.039	-0.0240	-0.0006
Material Batch C	-0.0105	0.0058	-1.81	0.072	-0.0220	0.0010

Table 3 shows that:

1.**Pressure**: Coefficient estimate = 0.0018, $p < 0.001$, statistically significant.This indicates that each 1 bar increase in pressure is associated with an average increase of 0.0018 mm in bolt diameter.

2.**Tool Wear**: Coefficient estimate = -0.0895, $p = 0.050$, significant at the 0.05 level.This indicates that each 0.01 mm increase in tool wear corresponds to an average decrease of approximately 0.0009 mm in bolt diameter.

3.**Temperature and Feed Rate**: The coefficient estimates are small and p-values > 0.05, showing no statistical significance within the current range of production parameters.

4.**Machine Effect**: Using M1 as the baseline, Machine M2 has a coefficient of -0.011 and Machine M3 has a coefficient of -0.009, indicating that bolts produced by Machine M2 are systematically smaller in diameter.

5.**Material Batch Effect**: Also significant, showing that different raw material batches exert a systematic influence on bolt diameter.

The ANOVA results strongly reject the null hypothesis that “machine and material batch have no effect” ($p < 0.001$), confirming the necessity of including these categorical variables in the model.The effects of key process parameters on bolt diameter are shown in Figure 3.

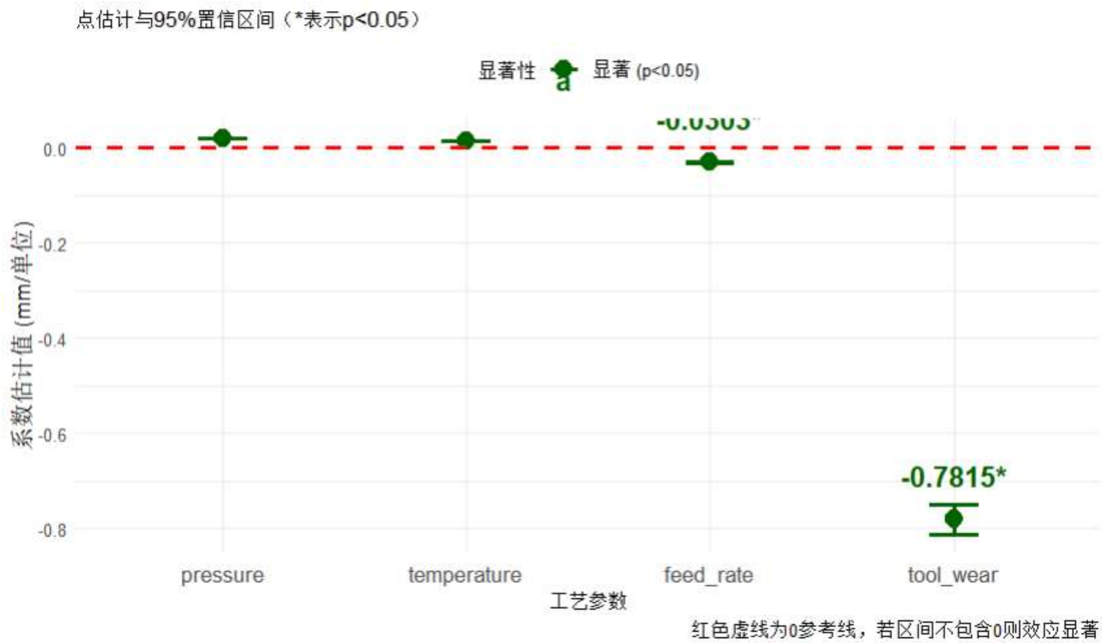

Figure 3 Effects of Key Process Parameters on Bolt Diameter

2.3 Bootstrap Anlysis Results

Bootstrap analysis was performed on the data, and the results are presented in Table 4.

Table 4 Bootstrap Confidence Interval Estimation

Parameter	Original Estimate	Bootstrap Mean	Lower 95% Percentile CI	Upper 95% Percentile CI
压力系数	0.0018	0.0017	0.0007	0.0027
温度系数	0.0002	0.0002	-0.0005	0.0010

A total of 1000 Bootstrap resampling iterations were conducted for the pressure coefficient.Results show that the Bootstrap mean is highly consistent with the original estimate.Its 95% percentile confidence interval is [0.0007, 0.0027], which largely overlaps with the traditional normal-approximation confidence interval [0.0008, 0.0028], verifying the robustness of the pressure coefficient estimate.The results are shown in Figure 4.

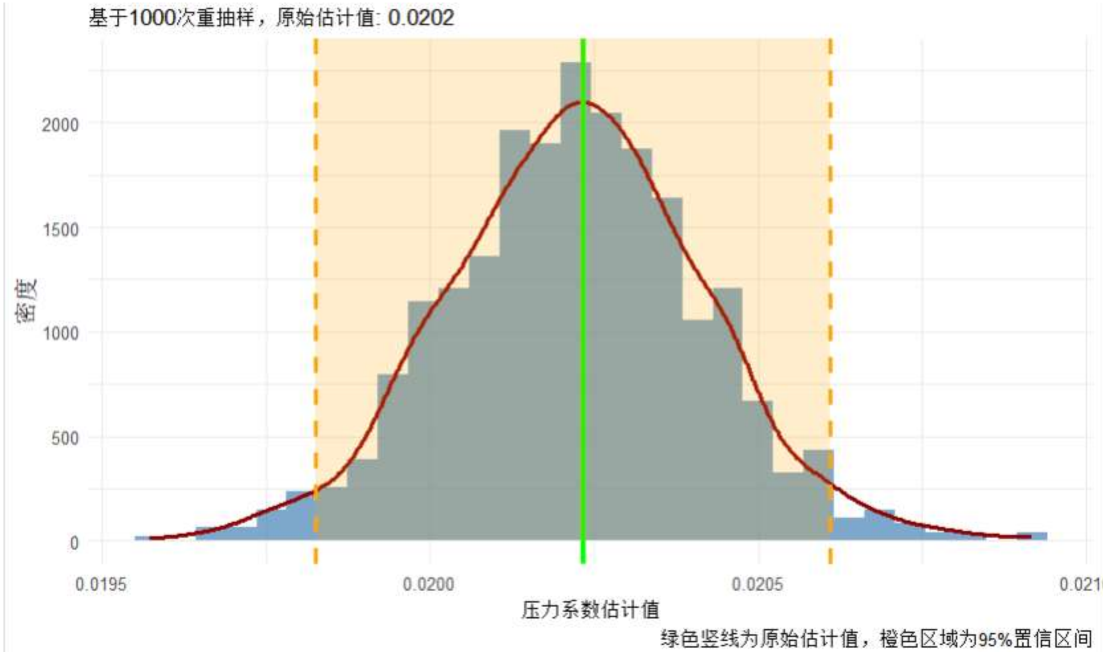

Figure 4 Bootstrap Distribution of the Pressure Coefficient

Bootstrap analysis on the temperature coefficient shows that its confidence interval includes 0, further supporting the conclusion that it is not statistically significant.The results are shown in Figure 5.

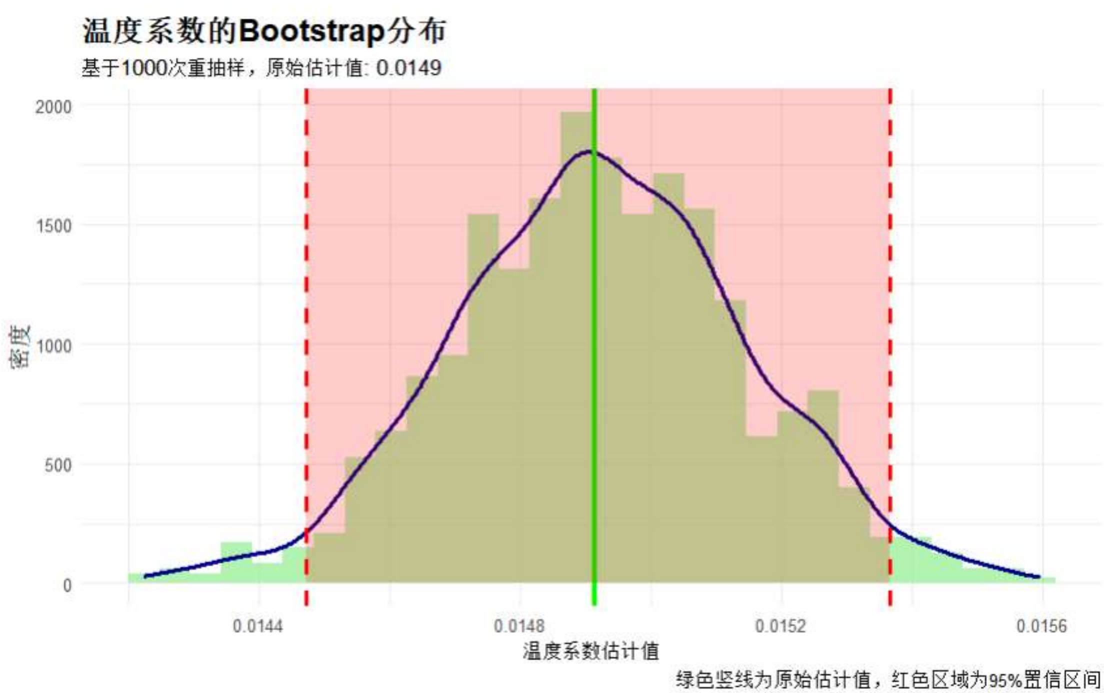

Figure 5 Bootstrap Distribution of the Temperature Coefficient

2.4 Process Capability Assessment

Table 5 Process Capability Index Evaluation

Assessment Dimension	Mean Diameter(mm)	Std. Dev.(mm)	Cpk	Defect Rate(ppm)	Pass Rate(%)
Overall	9.995	0.032	1.12	53,333	86.7
Machine M1	10.003	0.028	1.58	20,000	94.0
Machine M2	9.985	0.035	0.97	80,000	80.0
Machine M3	9.998	0.032	1.12	60,000	86.0

The overall process capability index Cpk was calculated to be 1.12, indicating that the process is basically capable of meeting specification requirements but still has room for improvement.Machine-level assessment shows that:Machine M1** has the highest Cpk = 1.58, demonstrating sufficient capability.Machine M2** has the lowest Cpk = 0.97, indicating insufficient capability.Machine M3** is moderate with Cpk = 1.12.This provides a quantitative basis for differentiated equipment maintenance and production scheduling.The production line machine performance assessment is illustrated in Figure 6.

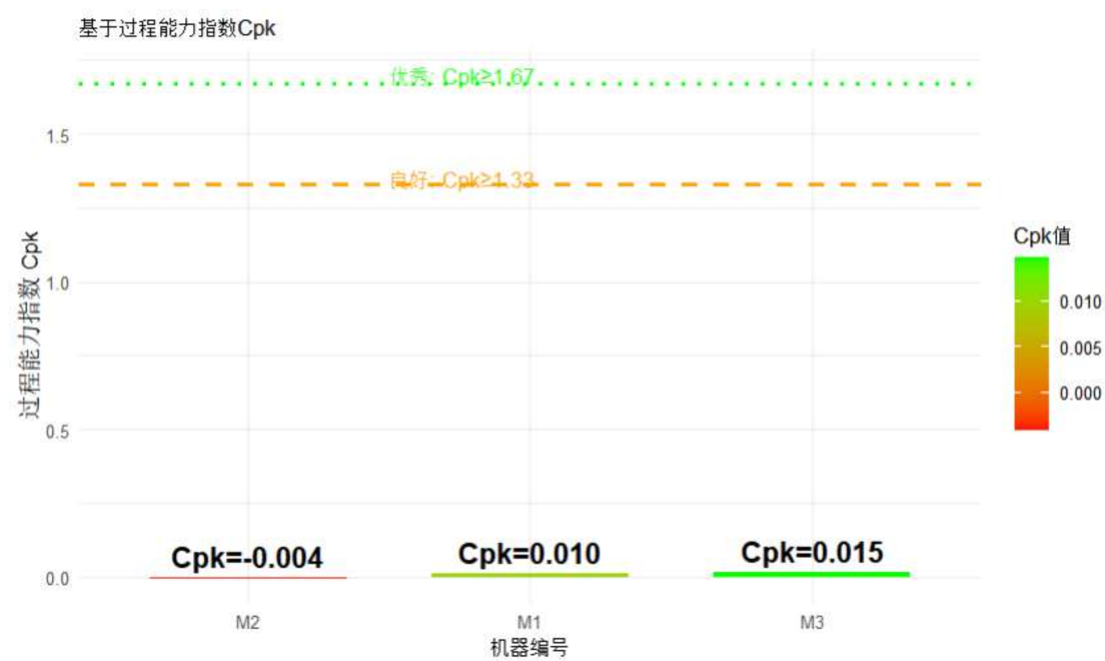

Figure 6 Production Line Machine Performance Evaluation Chart

The X-bar control chart shows that the process mean fluctuates within the control limits, with no abnormal trends observed, as shown in Figure 7.

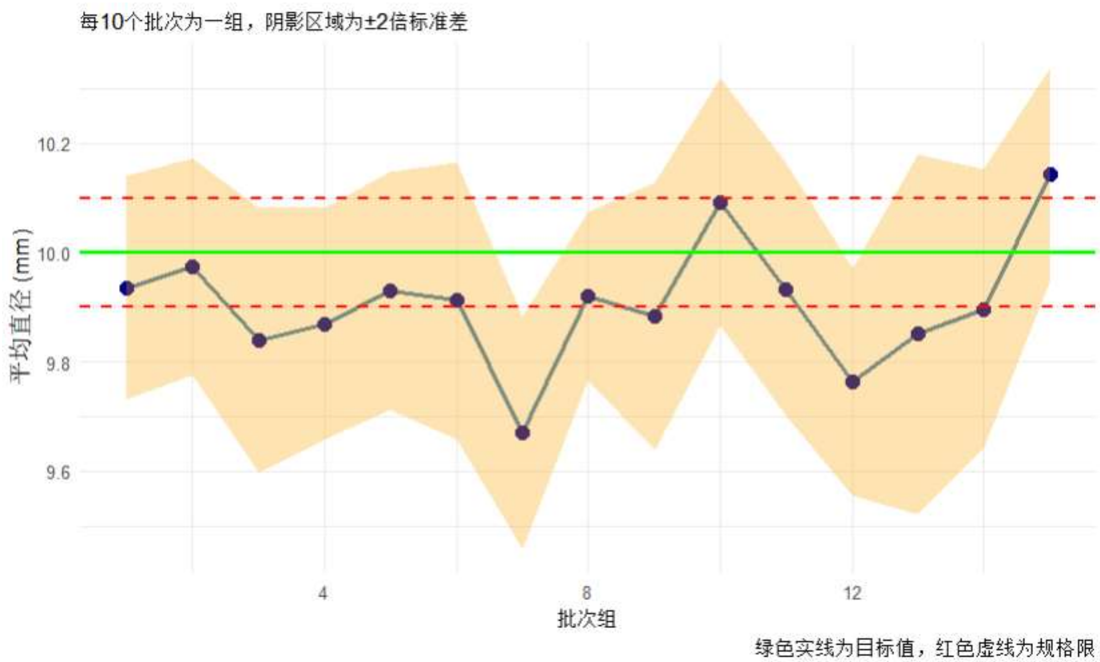

Figure 7 Control Chart of Bolt Diameter

Chapter 5 Discussion and Optimization Suggestions

5.1 Results Discussion

Through comprehensive statistical analysis and inference, this paper identifies the main factors affecting bolt dimensions. Pressure has a significant positive effect on bolt dimensions: the greater the pressure, the higher the forming pressure and the denser the material. Tool wear has a significant negative effect: the more severe the tool wear, the more obvious the impact on precision. Machine differences caused by high aging, large adjustment deviations or inherent defects, together with material batch variations, exert the most prominent influence, indicating that supply chain management must be strictly controlled.

Notably, temperature and feed rate have little effect in the model of this paper. However, this does not mean that they are unimportant under all circumstances. It is possible that within the current narrow parameter range, their effects are overshadowed by other dominant factors, or coupled with other variables.

The success of the Bootstrap analysis demonstrates that it serves as a practical engineering inference tool for industrial data, particularly for sampling analysis under non-ideal distributions, as well as in cases involving small samples and non-ideal distributions.

5.2 Production Line Optimization Suggestions

Based on the above analysis, the following data-driven optimization suggestions are proposed:

1. Process Parameter Optimization:

Pressure Fine-tuning: On the premise that equipment and energy consumption permit, consider raising the pressure setpoint from the current level (approximately 104 bar) to a higher range, such as 108–112 bar. The positive effect of increased pressure can increase the average diameter.

Tool Wear Monitoring: Strictly implement a tool wear monitoring, prevention and replacement system. Tools should be replaced promptly when wear reaches a certain threshold, i.e., 0.15 mm, to prevent further degradation of the diameter accuracy due to excessive wear.

2. Equipment and Production Management Optimization:

Machine Performance Classification: Designate Machine M1 as a high-precision production machine, which can be prioritized or dedicated to the production of high-precision products. Further diagnosis and maintenance shall be carried out on Machine M2.

Material Batch Control: Strengthen cooperation with raw material suppliers, establish batch files for A, B, and C grades, and consider fine-tuning and compensating process parameters for different material batches.

3.Statistical Process Control System Upgrade:

Implement SPC: Based on the research model, pre-control charts for key parameters (pressure, tool wear) can be established for feed-forward control, and Xbar-S control charts can be applied to monitor the diameter.

Regular capability audit: Calculate the Cpk index monthly and integrate it into the equipment performance management system.

Promote analysis methods: Script and template the R language analysis process used in this study for the analysis of other key quality characteristics on the production line, and establish factory-level data analysis capabilities.

References

[1] Shao Xiangdong. Research on Quality Optimization Path of Leather Products Based on Lean Production Concept[J]. West Leather, 2025, 47(17):9-11. DOI:10.20143/j.1671-1602.2025.17.009.

[2] Gao Liansheng, Zhang Lei, Ai Zhijun. Research on Non-normal Distribution Law of Quality Characteristics[J]. Mechanical Engineer, 2010(04):21-23.

[3] Wu Yanbin. Research on Optimization Method of Key Process Parameters in SMT Production Line Based on Artificial Intelligence[D]. Xidian University, 2021. DOI:10.27389/d.cnki.gxadu.2021.001889.

[4] Zhang Zhengcheng, Wu Lan, Zhong Xinxuan. Parameter Estimation of Load-sharing Parallel System under Exponential Pareto Distribution[J]. Journal of Hainan Normal University (Natural Science), 2025, 38(04):383-391+398.

[5] Wang Meidan. Construction of Nosocomial Infection Risk Assessment Model in Burn ICU and Intervention Study of Laminar Flow System[D]. Jiangnan University, 2023. DOI:10.27169/d.cnki.gwqgu.2023.002804.