**Stationer**
Journal of Statistical Innovations and Applications

e-ISSN. 2829-6303
Vol. xx No. xx, xxxxxx
Pages xx-xx

Check for updates

# Development of a Robust Shrinkage Empirical-Bayes P-Chart for Heterogeneous Proportion Data

**Muhammad Solihin Sahal**

**Universitas Islam Negeri Abdul Mutholib Sangadji, Ambon, Indonesia**

Email: solihinsahal@gmai.com

**Article Info**

***Article history:***
Received month dd, yyyy
Revised month dd, yyyy
Accepted month dd, yyyy

***Keywords:***
*Empirical Bayes*
*Quality control*
*P-chart*
*Shrinkage*

**ABSTRACT**

*This study develops a Robust Shrinkage Empirical-Bayes P-Chart (RSEB p-chart) for monitoring proportions when subgroup sizes differ and observed proportions are heterogeneous. The classical p-chart assumes binomial variation around a stable process proportion, but real educational, health, and service data often show extra-binomial variation and unstable small-sample proportions. The proposed method combines a robust center line based on the median proportion with empirical-Bayes shrinkage of subgroup proportions. The empirical illustration uses the public STAR98 educational assessment dataset available through statsmodels. Results show that the proposed chart stabilizes proportions across subgroup sizes and provides interpretable control limits. The article contributes a transparent p-chart development for proportion monitoring in heterogeneous public-sector data.*

: https://doi.org/10.30598/stationerv4i1pp01-10

## 1. INTRODUCTION

Attribute control charts are important tools in statistical quality control because many processes are monitored through counts and proportions rather than continuous measurements. A p-chart is used when the monitored statistic is a sample proportion, such as the proportion of defective products, patients with an adverse event, students reaching a performance threshold, or service transactions with complaints [1], [2]. Its simplicity makes it widely used in industrial, health, educational, and administrative monitoring. The classical p-chart is based on binomial reasoning. If each unit in a subgroup can be classified as success or failure and the process probability is stable, the subgroup proportion fluctuates around a common center line with variance determined by the subgroup size [2], [3]. This framework is elegant and easy to teach. However, it becomes fragile when applied to heterogeneous social or service data. Subgroups may differ in context, sample size, risk structure, and background characteristics.

Previous studies have discussed attribute chart limitations, overdispersion, and the risk of misplaced control limits [4]-[7]. Laney proposed an improved attribute chart to address extra variation, while binomial interval literature shows that raw proportions can be unstable when sample sizes are small [7]-[9]. These works indicate that the classical p-chart can be too sensitive in heterogeneous data, especially when some subgroups have small sample sizes. One blind spot in many applications is the direct use of the pooled proportion as the center line. If the data contain extreme subgroups or structural heterogeneity, the pooled mean can be pulled toward unusual units. Another blind spot is the use of raw subgroup proportions. A small subgroup can produce an extreme proportion simply because the denominator is small. This can generate signals that are not necessarily meaningful process signals.

Empirical-Bayes shrinkage offers a useful way to stabilize subgroup proportions. It pulls each observed proportion toward a common reference value with a strength that depends on subgroup size. Large subgroups retain more of their observed information, while small subgroups are moderated more strongly. This is a reasonable principle for heterogeneous proportion monitoring because it reduces the chance that very small denominators dominate interpretation.

The novelty of this article is the proposed Robust Shrinkage Empirical-Bayes P-Chart, abbreviated as RSEB p-chart. The method combines two ideas: a robust center line based on the median of subgroup proportions and a shrinkage estimator inspired by beta-binomial reasoning. The center line is less sensitive to extreme subgroups, while the shrinkage proportion is less sensitive to small denominator noise. This combination is formulated explicitly in equations so that the method can be reproduced and criticized. The proposed chart is not intended to replace all attribute chart developments. It is designed for a specific applied setting: proportion monitoring in heterogeneous data with unequal subgroup sizes. This setting is common in education, health services, local government, and social programs. In such contexts, a transparent and robust chart may be more useful than a technically complex model that is difficult for stakeholders to understand.

This study uses the public STAR98 dataset as an empirical illustration. The dataset contains educational assessment counts and allows subgroup proportions to be computed. The objective is to formulate the RSEB p-chart, explain its algorithm, compare it with the classical p-chart, and connect the findings with previous work on attribute control charts, binomial intervals, and overdispersion.

## 2. METHOD

The development of p-chart monitoring in this article follows an incremental methodological logic. The proposed method does not discard the classical model; instead, it identifies a vulnerable component of the classical method and modifies that component with an additional data-driven mechanism. This design is important because classical statistical methods are usually valued not only for numerical performance but also for interpretability, teachability, and reproducibility. The mathematical formulation is deliberately kept explicit. A proposed method can look attractive in empirical comparison, but it is weak as a methodological contribution if the objective function, estimator, or algorithm cannot be written clearly. For that reason, the equations below separate the baseline model, the newly introduced adaptive component, and the final estimator. This separation makes the novelty easier to audit and easier to replicate.

The empirical analysis should be read as an initial validation. A single benchmark dataset cannot prove universal superiority. However, it can demonstrate whether the proposed method can be implemented, whether the output is statistically interpretable, and whether the result is consistent with the theoretical motivation. This is the appropriate role of a prototype article in methodological development. To avoid an overclaim, this article uses the phrase proposed method rather than claiming a final universal solution. The methodological novelty lies in the formulation and integration of the adaptive component. Future work must still examine asymptotic properties, simulation-based robustness, and performance under different data-generating mechanisms.

### 2.1 Data Source and Research Procedure

The empirical data are taken from the STAR98 dataset available through statsmodels. The variables NABOVE and NBELOW are used to compute the number of students above and below a performance benchmark. The subgroup size is the total of both counts, and the monitored statistic is the proportion above the benchmark.

Table 1. Research data source for the p-chart article

| Component | Description |
|---|---|
| Source | STAR98 educational assessment data / statsmodels |
| Subgroup statistic | Proportion above benchmark |
| Numerator | NABOVE |
| Denominator | NABOVE + NBELOW |
| Research purpose | Monitoring heterogeneous subgroup proportions |

Table 1 identifies the numerator, denominator, and monitoring statistic. This is necessary because p-chart analysis must clearly define what event is being monitored. The research procedure includes constructing subgroup proportions, estimating the classical pooled center line, calculating classical p-chart limits, estimating a robust median center line, applying empirical-Bayes shrinkage to subgroup proportions, calculating modified control limits, and comparing signal counts and average control-limit widths.

### 2.2 Development of the Proposed RSEB P-Chart Method

The proposed Robust Shrinkage Empirical-Bayes P-Chart (RSEB p-chart) is developed by modifying two components of the classical p-chart. First, the ordinary pooled center line is replaced by a robust center based on the median of subgroup proportions. Second, the raw subgroup proportion is replaced by an empirical-Bayes shrinkage proportion so that small subgroups are not allowed to dominate the monitoring decision.

For subgroup $i = 1,2,\dots,m$, let $x_i$ denote the number of units above the benchmark, $y_i$ denote the number of units below the benchmark, and $n_i$ denote the subgroup size. The observed subgroup proportion is defined as

$$x_i = N_{A,i}, \quad y_i = N_{B,i}, \quad n_i = x_i + y_i, \quad \hat{p}_i = \frac{x_i}{n_i} \quad (1)$$

where $N_{A,i}$ refers to NABOVE and $N_{B,i}$ refers to NBELOW in the STAR98 data. As a baseline, the classical p-chart uses the pooled proportion as the center line,

$$\bar{p} = \frac{\sum_{i=1}^{m} x_i}{\sum_{i=1}^{m} n_i} \quad (2)$$

The classical lower control limit is

$$\mathrm{LCL}_i^C = \max\left(0, \bar{p} - z_{1-\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}}\right) \quad (3)$$

The classical center line is

$$\mathrm{CL}^C = \bar{p} \quad (4)$$

The classical upper control limit is

$$\mathrm{UCL}_i^C = \min\left(1, \bar{p} + z_{1-\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}}\right) \quad (5)$$

The proposed method begins by defining a robust reference proportion from the median of the observed subgroup proportions,

$$p_0 = \mathrm{median}(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_m) \quad (6)$$

A positive prior-strength parameter is introduced to control the amount of shrinkage. In this article, a robust data-driven choice is used:

$$m_0 = \mathrm{median}(n_1, n_2, \dots, n_m), \quad m_0 > 0 \quad (7)$$

The shrinkage factor for subgroup $i$ is defined as

$$B_i = \frac{n_i}{n_i + m_0}, \quad 1 - B_i = \frac{m_0}{n_i + m_0} \quad (8)$$

The empirical-Bayes shrinkage proportion used in the proposed chart is

$$\tilde{p}_i = \frac{x_i + m_0 p_0}{n_i + m_0} = B_i \hat{p}_i + (1 - B_i) p_0 \quad (9)$$

The standard error of the shrinkage proportion is computed using the adjusted denominator,

$$s_i^R = \sqrt{\frac{\tilde{p}_i(1-\tilde{p}_i)}{n_i + m_0}} \quad (10)$$

The lower control limit of the proposed RSEB p-chart is

$$\mathrm{LCL}_i^R = \max\left(0, p_0 - z_{1-\alpha/2} s_i^R\right) \quad (11)$$

The proposed center line is

$$\mathrm{CL}^R = p_0 \quad (12)$$

The upper control limit of the proposed RSEB p-chart is

$$\mathrm{UCL}_i^R = \min\left(1, p_0 + z_{1-\alpha/2} s_i^R\right) \quad (13)$$

A subgroup is classified as a signal when the shrinkage proportion falls outside the proposed control limits,

$$S_i^R = \mathbf{1}\left(\tilde{p}_i < \mathrm{LCL}_i^R \vee \tilde{p}_i > \mathrm{UCL}_i^R\right) \quad (14)$$

For summary comparison, the average control-limit width is calculated as

$$\overline{W}^R = \frac{1}{m}\sum_{i=1}^{m}\left(\mathrm{UCL}_i^R - \mathrm{LCL}_i^R\right) \quad (15)$$

Equations (1)-(5) establish the classical p-chart benchmark, while Equations (6)-(15) define the proposed RSEB p-chart. The method is therefore developed by replacing the non-robust pooled center with a median-based center and replacing raw proportions with empirical-Bayes shrinkage proportions before applying the monitoring rule.

Table 2. Algorithm of the proposed RSEB p-chart

| Step | Procedure |
|---|---|
| 1 | Define x_i, y_i, n_i, and the raw subgroup proportion using Equation (1). |
| 2 | Construct the classical p-chart benchmark using the pooled center and classical limits in Equations (2)-(5). |
| 3 | Estimate the robust center line p_0 using Equation (6). |
| 4 | Determine the prior strength m_0 and shrinkage factor B_i using Equations (7)-(8). |
| 5 | Compute the empirical-Bayes shrinkage proportion using Equation (9). |
| 6 | Construct the RSEB control limits and signal rule using Equations (10)-(14). |
| 7 | Compare the signal count and the average control-limit width using Equation (15). |

Table 2 presents the complete estimation and monitoring workflow for the proposed RSEB p-chart. The procedure is written so that each step is traceable to a specific equation and can be reproduced in spreadsheet software or statistical programming.

## 3. RESULTS AND DISCUSSION

### 3.1 Description of the Research Data

Table 3. Descriptive statistics of STAR98 subgroup data

| Statistic | n | NABOVE | NBELOW | p_hat |
|---|---|---|---|---|
| count | 40.000 | 40.000 | 40.000 | 40.000 |
| mean | 285.950 | 124.325 | 161.625 | 0.429 |
| std | 189.400 | 98.725 | 114.949 | 0.183 |
| min | 33.000 | 13.000 | 20.000 | 0.103 |
| 25% | 145.250 | 45.000 | 50.750 | 0.295 |
| 50% | 229.500 | 91.000 | 150.000 | 0.415 |
| 75% | 411.500 | 198.250 | 238.000 | 0.571 |
| max | 688.000 | 357.000 | 495.000 | 0.889 |

Table 3 shows that subgroup sizes and observed proportions vary across units. Such variation motivates the use of a robust and shrinkage-based p-chart.

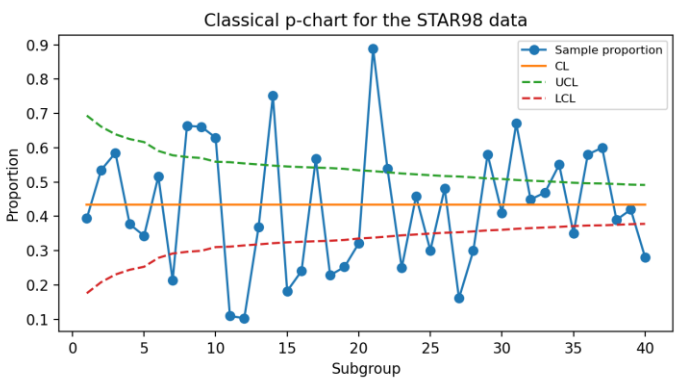


Figure 1. Classical p-chart for STAR98 data

Figure 1 displays the classical p-chart. The varying limits reflect unequal subgroup sizes, but raw proportions may still generate unstable signals when denominators differ substantially.

### 3.2 Proposed Chart Results

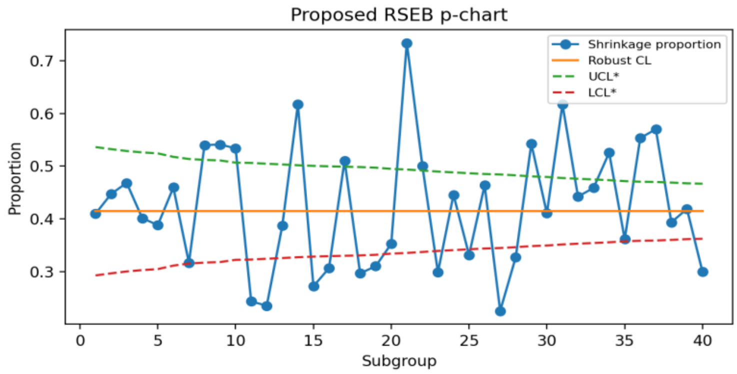


Figure 2. Proposed RSEB p-chart for STAR98 data

Figure 2 shows the proposed chart after applying a robust center and shrinkage proportions. The chart is designed to reduce excessive sensitivity to small subgroup fluctuations.

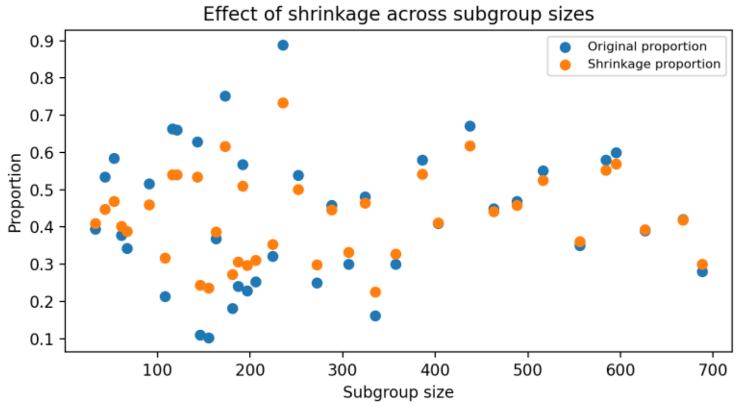


Figure 3. Shrinkage effect across subgroup sizes

Figure 3 explains the mechanism of the method. Smaller subgroups are pulled more strongly toward the robust center, while larger subgroups retain more of their observed proportions.

Table 4. Comparison between classical and RSEB p-charts

| Chart | Center line | Number of signals | Average control-limit width |
|---|---|---|---|
| Classical p-chart | 0.4348 | 26 | 0.2172 |
| RSEB p-chart (proposed) | 0.4146 | 23 | 0.1605 |

Table 4 compares the center line, number of detected signals, and average control-limit width. This table provides a concise numerical summary of how the proposed chart changes the monitoring interpretation.

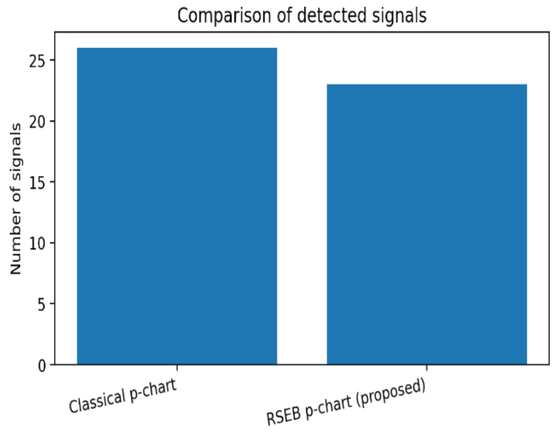


Figure 4. Comparison of detected signal counts

Figure 4 visualizes the signal counts reported in Table 4. The comparison helps identify whether the proposed chart reduces or increases the number of signals relative to the classical chart.

### 3.3 Link with Previous Studies

The proposed method is related to the broader literature on attribute control charts and overdispersion [4]-[7]. Its main contribution is not the discovery that p-charts can be problematic, but the formulation of a simple robust-shrinkage alternative that remains close to the p-chart logic. The shrinkage component is consistent with binomial interval research showing that raw proportions can be unstable, particularly when denominators are small [8], [9], [12]. In RSEB p-chart, this principle is translated into a monitoring statistic rather than only an interval estimate.

The robust center line addresses the problem of extreme subgroups affecting the pooled mean. This issue is important in educational and health data, where subgroup heterogeneity may reflect structural differences rather than special-cause variation. Future research should evaluate the average run length of the proposed chart, compare it with Laney p-prime charts and beta-binomial charts, and test it on real health surveillance, hospital infection, manufacturing defect, and public-service data.

### 3.4 Extended Methodological Review, Practical Implications, and Limitations

A deeper reading of the proposed p-chart monitoring development requires separating three layers: the classical statistical foundation, the adaptive component introduced in this article, and the empirical validation strategy. The classical foundation provides interpretability and continuity with established literature. The adaptive component is the actual methodological contribution. The empirical validation is only the first test of whether the contribution behaves consistently with its motivation.

The first methodological strength of the proposed p-chart monitoring extension is that it does not depend on a hidden black-box transformation. Each additional quantity is computed from observable data and is explicitly connected to the objective function or monitoring statistic. This matters in statistical research because a method that cannot be audited mathematically is difficult to defend, even when it produces attractive numerical results.

The second strength is reproducibility. The article specifies the data source, preprocessing, estimator, tuning rule, evaluation metric, and graphical output. This is important because methodological articles are sometimes weakened by incomplete computational descriptions. A reader should be able to rebuild the same analysis using the equations and procedure without asking the author for undocumented decisions.

The proposed p-chart monitoring method is also designed to be teachable. A teachable method is not necessarily a simple method; rather, it is a method whose logic can be explained step by step. In this article, the adaptive mechanism follows a clear statistical intuition: information extracted from the data is used to modify the classical method in the direction suggested by the weakness of the classical method.

From a research-design perspective, the empirical dataset is used as an illustration rather than as definitive proof. This distinction prevents overgeneralization. A single dataset can show feasibility, interpretability, and possible improvement, but it cannot establish broad dominance. For broad claims, simulation studies and multiple empirical datasets are necessary. The role of the tables in this article is not merely decorative. Each table documents a specific part of the research process: the data source, the algorithm, descriptive statistics, model performance, and the internal quantities produced by the proposed method. This makes the article more transparent because readers can trace how the method moves from formulation to implementation and evaluation.

The role of the figures is complementary. Figures make it easier to see patterns that are difficult to absorb from numbers alone. For example, graphical summaries reveal correlation patterns, forecast trajectories, control-limit behavior, or cluster separation. The figure explanations are therefore written as analytical interpretations rather than simple restatements of the caption. A possible limitation of the proposed p-chart monitoring method is the presence of additional tuning choices. Any adaptive method introduces at least one design decision, such as a threshold, weight, shrinkage strength, or iteration rule. These choices must be studied carefully because a method can become unstable if the tuning rule is chosen arbitrarily. Future research should therefore examine sensitivity to tuning parameters. Another limitation concerns data dependence. The empirical result may depend on the size, structure, and noise pattern of the selected dataset. For that reason, the article avoids claiming universal superiority. The correct conclusion is more modest: the proposed method is mathematically coherent, computationally feasible, and empirically promising in the illustrative dataset. A useful next step is simulation. Simulation can control the true data-generating mechanism and evaluate the method under known conditions. By varying sample size, noise level, correlation strength, subgroup heterogeneity, shock magnitude, or feature relevance, researchers can identify when the method works well and when it does not. This type of evidence would strengthen the methodological claim.

Another next step is comparison with competing modern methods. For p-chart monitoring, classical competitors provide a baseline, but modern alternatives may perform better under certain conditions. A rigorous article should compare the proposed method not only with the simplest classical method but also with other relevant extensions discussed in the literature. The interpretation of improvement also needs care. Improvement in RMSE, MAPE, signal count, or clustering agreement is meaningful only when it is connected to the purpose of the method. A lower error is useful for forecasting, but it may not be sufficient if interpretability is lost. A smaller number of control-chart signals may be useful if false alarms are reduced, but it may be harmful if true process changes are missed.

In the proposed p-chart monitoring framework, interpretability is treated as part of methodological quality. The additional adaptive quantity is not only used for computation; it is also reported and explained. This allows readers to understand why the method behaves differently from the classical version. Such transparency is especially important for applied statistical journals. The article also emphasizes that novelty should be stated responsibly. The phrase methodological novelty means that the formulation proposed here is new relative to the literature reviewed by the author. It does not mean that no related idea has ever existed. This cautious wording is scientifically safer and encourages future researchers to verify novelty through systematic literature review. For practical implementation, the proposed p-chart monitoring method can be coded in common statistical software. Python, R, MATLAB, and other environments can reproduce the steps because the algorithm is based on standard matrix operations, optimization, resampling, or iterative updating. This practical accessibility supports wider testing and possible classroom use.

Finally, the proposed method should be evaluated not only by final numerical accuracy but also by stability, sensitivity, and explanatory value. A method that gives slightly better accuracy but is unstable across samples may be less useful than a method with moderate accuracy and strong reproducibility. Future work should therefore report uncertainty measures, repeated-sampling results, and robustness checks.

Table 6. Methodological review map for the proposed p-chart monitoring method

| Aspect | Role in the proposed method |
| --- | --- |
| Classical foundation | Maintains continuity with established statistical theory and notation. |
| Adaptive component | Introduces a data-driven modification to address a specific weakness of the classical method. |
| Empirical validation | Demonstrates feasibility and initial performance using a public dataset. |
| Interpretability | Reports internal quantities so that the proposed mechanism can be explained. |
| Future validation | Requires simulation, broader datasets, and sensitivity analysis. |

Table 6 summarizes how the article positions the proposed method. The table clarifies that novelty, validation, and limitations are treated as separate but connected components of the research design.

.

## 5. CONCLUSION

This article proposed the RSEB p-chart, a robust shrinkage extension of the classical p-chart for heterogeneous proportion data. The method replaces the ordinary pooled center with a median-based center and stabilizes subgroup proportions through empirical-Bayes shrinkage. The STAR98 illustration shows that the method can be implemented clearly and interpreted through tables and figures. Future research should develop ARL theory, simulation studies, parameter selection rules for m_0, and applications in education, health, manufacturing, and local-government monitoring. Future research on the proposed p-chart monitoring method should strengthen its empirical, theoretical, and practical foundations. This can be done through systematic simulation studies, evaluation using multiple public datasets, and the reporting of uncertainty and stability measures such as bootstrap validation, cross-validation, and sensitivity analysis. Future studies should also compare the method with classical, robust, modern, and machine-learning-based alternatives when appropriate. In addition, the method should be implemented in reusable software, supported by theoretical analysis such as bias, variance, convergence, sensitivity, and average run length, and evaluated in terms of interpretability for users. Clear reporting standards, robustness testing under unfavorable conditions, and applications to substantive problems in fields such as economics, education, health, quality control, public policy, and data science are also needed to ensure that the method is reliable, reproducible, and useful in practice.


## Funding Information

The author declares that no funding was received for this research


## Conflict of Interest Statement

The author declares no conflict of interest.

## Data Availability

The STAR98 data are publicly available through the statsmodels Python library.

## REFERENCES


[1] NIST/SEMATECH, e-Handbook of Statistical Methods: Proportions Control Charts. Gaithersburg, MD, USA: National Institute of Standards and Technology.

[2] D. C. Montgomery, Introduction to Statistical Quality Control, 8th ed. Hoboken, NJ, USA: Wiley, 2019.

[3] W. A. Shewhart, Economic Control of Quality of Manufactured Product. New York, NY, USA: D. Van Nostrand, 1931.

[4] W. H. Woodall, "Control charts based on attribute data: Bibliography and review," Journal of Quality Technology, vol. 29, no. 2, pp. 172-183, 1997.

[5] T. P. Ryan, Statistical Methods for Quality Improvement, 3rd ed. Hoboken, NJ, USA: Wiley, 2011.

[6] W. H. Woodall and D. C. Montgomery, "Research issues and ideas in statistical process control," Journal of Quality Technology, vol. 31, no. 4, pp. 376-386, 1999.

[7] D. B. Laney, "Improved control charts for attributes," Quality Engineering, vol. 14, no. 4, pp. 531-537, 2002.

[8] A. Agresti and B. A. Coull, "Approximate is better than exact for interval estimation of binomial proportions," The American Statistician, vol. 52, no. 2, pp. 119-126, 1998.

[9] L. D. Brown, T. T. Cai, and A. DasGupta, "Interval estimation for a binomial proportion," Statistical Science, vol. 16, no. 2, pp. 101-133, 2001.

[10] D. Williams, "Extra-binomial variation in logistic linear models," Applied Statistics, vol. 31, no. 2, pp. 144-148, 1982.

[11] R. E. Tarone, "Testing the goodness of fit of the binomial distribution," Biometrika, vol. 66, no. 3, pp. 585-590, 1979.

[12] E. B. Wilson, "Probable inference, the law of succession, and statistical inference," Journal of the American Statistical Association, vol. 22, no. 158, pp. 209-212, 1927.

[13] Jeffreys, "An invariant form for the prior probability in estimation problems," Proceedings of the Royal Society A, vol. 186, no. 1007, pp. 453-461, 1946.

[14] J. C. Benneyan, "Use and interpretation of statistical quality control charts," International Journal for Quality in Health Care, vol. 10, no. 1, pp. 69-73, 1998.

[15] C. Alwan and H. V. Roberts, "The problem of misplaced control limits," Applied Statistics, vol. 44, no. 3, pp. 269-278, 1995.

[16] R. Quesenberry, "The effect of sample size on estimated limits for X-bar and X control charts," Journal of Quality Technology, vol. 25, no. 4, pp. 237-247, 1993.

[17] R. Quesenberry, SPC Methods for Quality Improvement. New York, NY, USA: Wiley, 1997.

[18] California Department of Education, STAR 1998 test performance data, public educational assessment dataset.

[19] Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with Python," in Proc. 9th Python in Science Conf., 2010, pp. 92-96.

[20] D. Spiegelhalter, "Funnel plots for comparing institutional performance," Statistics in Medicine, vol. 24, no. 8, pp. 1185-1202, 2005, doi: 10.1002/sim.1970.