

통계적 품질관리를 위한 왜도의 활용

김훈태* · 임성욱*†

* 대진대학교 산업경영공학과

Utilization of Skewness for Statistical Quality Control

Kim, Hoontae* · Lim, Sunguk*†

* Dept. of Industrial and Management Engineering, Daejin University

ABSTRACT

Purpose: Skewness is an indicator used to measure the asymmetry of data distribution. In the past, product quality was judged only by mean and variance, but in modern management and manufacturing environments, various factors and volatility must be considered. Therefore, skewness helps accurately understand the shape of data distribution and identify outliers or problems, and skewness can be utilized from this new perspective. Therefore, we would like to propose a statistical quality control method using skewness.

Methods: In order to generate data with the same mean and variance but different skewness, data was generated using normal distribution and gamma distribution. Using Minitab 18, we created 20 sets of 1,000 random data of normal distribution and gamma distribution. Using this data, it was proven that the process state can be sensitively identified by using skewness.

Results: As a result of the analysis of this study, if the skewness is within ± 0.2 , there is no difference in judgment from management based on the probability of errors that can be made in the management state as discussed in quality control. However, if the skewness exceeds ± 0.2 , the control chart considering only the standard deviation determines that it is in control, but it can be seen that the data is out of control.

Conclusion: By using skewness in process management, the ability to evaluate data quality is improved and the ability to detect abnormal signals is excellent. By using this, process improvement and process non-substitutability issues can be quickly identified and improved.

Key Words: Control Chart, Process Capability, Skewness, SQC

● Received 12 November 2023, 1st revised 13 November 2023, accepted 27 November 2023

† Corresponding Author(sulim@daejin.ac.kr)

© 2023, Korean Society for Quality Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-Commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

품질관리 분야에서 평균과 분산은 오랫동안 핵심적인 품질 지표로 사용되어 왔다. 평균은 데이터의 중심 위치를 나타내고, 분산은 데이터의 변동성을 평가하여 제품 또는 프로세스의 안정성을 확인하는 데 유용했다. 그러나 현대 품질관리에서는 이러한 전통적인 지표 외에도 "왜도(Skewness)"라는 통계적 개념이 점차 더 중요해지고 있다. 왜도는 데이터 분포의 비대칭성을 측정하는 데 사용되는 지표이다. 과거에는 제품의 품질을 평균과 분산만으로 판단했지만, 현대 경영과 제조 환경에서는 다양한 요인과 변동성을 고려해야 한다. 따라서 왜도는 데이터 분포의 형태를 정확하게 파악하고, 이상치나 문제점을 식별하는 데 도움이 되며, 이러한 새로운 관점에서 왜도는 활용될 수 있다.

왜도가 필요한 이유 중 하나는 데이터의 비대칭 분포에서 유용한 정보를 제공한다는 점이다. 대부분의 실제 데이터는 정규 분포가 아니며, 데이터가 한쪽으로 치우칠 수 있다. 왜도를 통해 데이터 분포의 치우침 정도를 측정할 수 있으며, 이것은 예측 및 문제해결에 중요한 정보를 제공한다. 뿐만 아니라, 왜도를 활용하면 데이터 분포의 변화를 감지하고 제품 또는 프로세스의 문제를 조기에 발견할 수 있다. 예를 들어, 왜도가 급격하게 변할 때, 이는 제품 또는 프로세스에서 중요한 변화가 일어났음을 나타낼 수 있다.

따라서, 이번 글에서는 왜도의 품질관리에서의 역할과 중요성에 대해 자세히 다루고, 왜도가 제품 또는 프로세스의 품질 향상에 어떻게 기여하는지에 대해 논의할 것이다. 왜도의 변화가 실제 어떻게 활용될 수 있는지 시뮬레이션을 통하여 증명하였다.

2. 분산을 이용한 품질관리

2.1 관리도

통계적 공정관리는 제조 및 생산 프로세스에서 품질을 유지하고 개선하기 위한 중요한 방법 중 하나이다. 이 접근법에서 주요한 주제 중 하나는 품질 특성치, 즉 제품 또는 프로세스의 관심 대상인 품질 특성이 정규분포 또는 정규 분포와 유사한 분포를 따른다고 가정하고 이를 관리하는 방법이다. 이를 위해 평균과 분산을 일정한 범위 내에서 관리하려는 방법들이 사용된다. 이러한 방법으로 가장 널리 쓰이는 것이 바로 관리도이다.

관리도는 제조 및 공정에서 제품 또는 프로세스의 품질 특성치를 모니터링하고 관리하는 데 사용되는 통계적 도구이다. 주로 정규 분포 또는 정규 분포와 유사한 분포를 따르는 데이터에 적용된다. 관리도는 시간의 경과에 따른 데이터의 변화를 시각적으로 보여주어 이상 신호를 탐지할 수 있도록 도와준다.

주요 방법으로는 평균을 관리하기 위한 " \bar{x} -관리도"와 분산 또는 표준편차를 관리하기 위한 " R -관리도" 또는 " S -관리도"가 있다.

이 중에서 " R -관리도"는 통계적 성질이 우수하고 관련 이론이 잘 정립되어 있어 현장에서 간편하게 적용할 수 있으며, S -관리도에 비해 계산이 덜 복잡한 장점을 갖고 있다. R -관리도의 핵심 개념 중 하나는 "평균런의 길이(average run length, ARL)"라는 값인데, 이 값은 공정이 이상 신호를 검출하는 데 얼마나 빈번하게 사용되는지를 나타내는 지표이다. ARL은 일반적으로 공정이 정상적으로 운영될 때의 ARL(ARL_0)과 공정에 이상이 있을 때의 ARL(ARL_1)로 구분된다. 바람직한 관리도는 ARL_0 는 커야 하고 ARL_1 은 작아야 한다(Lee, 2012).

관리도의 개념은 통계적 가설검정의 한 응용으로 해석되며, “공정이 정상이다(귀무가설)”와 “공정에 이상이 있다(대립가설)”인 가설을 반복적으로 검정하여 품질 특성을 모니터링하고 이상 신호를 탐지하는 것과 관련이 있다. 이러한 접근법을 통해 제조 및 생산 프로세스에서 품질을 효과적으로 관리하고 향상시키는 데 도움이 되었다.

2.2 공정능력지수

제조업 분야에서 제품의 품질은 핵심적인 고려 요소 중 하나이며, 품질 향상은 기업의 경쟁력을 유지하고 개선하는 데 중요하다. 공정능력지수(Cp)는 품질 향상을 위한 중요한 도구 중 하나로, 제조 프로세스의 품질을 정량화하고 관리하는 데 사용된다(이도경, 2009).

Cp는 제품 또는 프로세스의 품질 특성이 목표 품질 요구 사양 내에서 어떤 정도로 퍼져있는지를 측정하는 지표로, 목표 품질 요구 사양은 고객의 기대 품질 수준을 나타낸다. Cp는 다음과 같이 계산된다(조중재, 2010).

$$C_p = (USL - LSL) / 6\sigma$$

여기서,

- Cp: 공정능력지수
- USL: 상한 목표치(상한 제한)
- LSL: 하한 목표치(하한 제한)
- σ : 프로세스의 표준편차를 의미한다.

Cp 값이 클수록 프로세스는 목표 품질 요구 사양을 충족하는 데 더 우수한 능력을 갖는다. 일반적으로, Cp 값이 1보다 크면 프로세스는 목표 품질 요구 사양을 충족할 가능성이 높아지고, 반면에 Cp 값이 1보다 작으면 프로세스가 목표 품질 요구 사양을 충족하기 어려울 수 있으며, 개선이 필요할 수 있다. 일반적으로 기업에서 경쟁력을 갖기 위해서는 Cp가 1.33보다는 커야 한다고 이야기한다. 여기서 Cp는 생산 목표와 프로세스의 평균이 같을 경우를 의미한다. 만일 공정의 치우침이 생길 경우에는 Cp로 생산공정의 변화를 파악하지 못하게 된다.

이를 극복하기 위하여 Cpk를 사용하여 공정능력을 계산하게 된다(Figure 1).

공정능력지수(Cpk)는 공정의 치우침이 발생하였거나 규격이 상한 또는 하한만 존재할 경우에도 사용된다. Cpk 값은 다음과 같이 계산된다.

- $C_{pu} = (USL - \text{Process Mean}) / (3\sigma)$
- $C_{pl} = (\text{Process Mean} - LSL) / (3\sigma)$

여기서,

- USL: 상한 목표치(상한 제한),
- LSL: 하한 목표치(하한 제한),
- Process Mean: 프로세스의 평균값,
- σ : 프로세스의 표준편차를 의미한다.

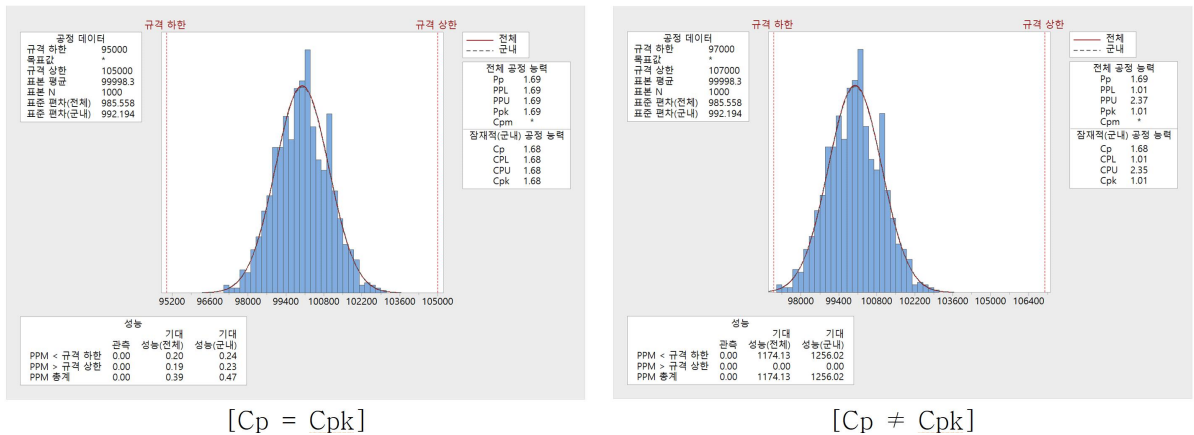


Figure 1. Relationship between Cp & Cpk

공정능력지수(Cpk)는 제조업에서 품질 향상을 위한 중요한 도구로 사용되며, 프로세스의 품질과 안정성을 평가하는 데 도움을 준다. 하지만 이러한 공정능력지수는 공정의 능력을 판단하는데 규격과 표준편차의 값만을 이용하기 때문에 동일한 규격과 표준편차가 같은 경우 동일한 공정의 능력으로 판단할 수 밖에 없다. 하지만 실제로는 같은 규격과 표준편차를 가지고 있음에도 불구하고 다른 분포의 모양을 갖는 경우가 존재하고 있으며, 빅데이터가 수집되는 현장에서는 이제 규격과 표준편차만으로는 공정의 능력을 판단하는 데 한계가 있다.

3. 품질관리를 위한 분산의 한계 및 왜도

공정상태를 파악하는 방법으로 가장 널리 사용되는 것은 기존 연구에서 언급한 것처럼 관리도와 공정능력지수를 이용한 통계적 품질관리가 많이 사용되어 왔다. 관리도는 제조 프로세스에서 제품의 핵심 품질 특성을 추적하고 모니터링하는 데 사용되며 품질 특성의 평균 및 변동성을 추적하며, 목표 품질과의 비교를 통해 공정 상태를 확인할 수 있는 장점이 있다. 공정능력지수 (Cp, Cpk)는 제조 프로세스의 품질 능력을 평가하는 데 사용한다. Cp는 제품 품질의 퍼짐 정도를 나타내고, Cpk는 평균을 중심으로 제품 품질 특성의 편차를 고려한다. 이러한 지수를 사용하여 공정의 능력을 확인할 수 있다. 하지만 이 두 가지 대표적인 지표는 모두 공통적으로 분산(표준편차)를 이용하는 지표로 예를 들어 다음과 같은 상황에서는 공정의 능력의 차이를 구분할 수 없다(Figure 2).

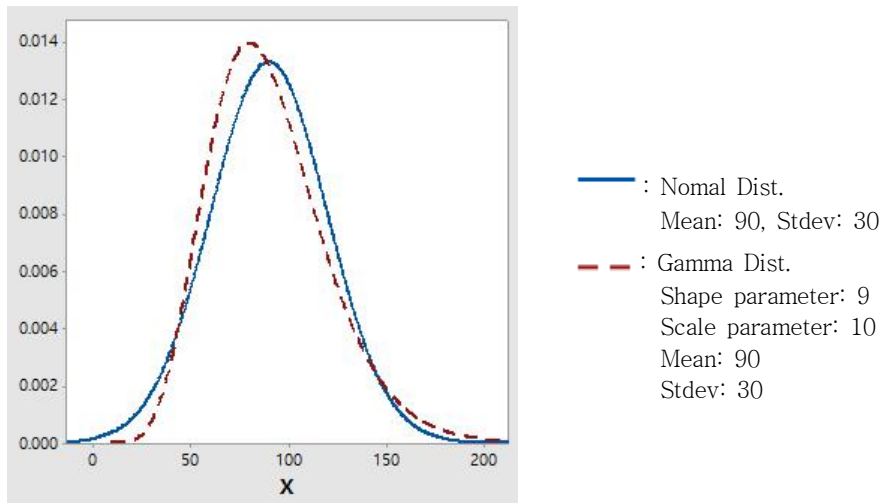


Figure 2. Distributions with the same variance but different skewness

본 논문에서는 감마분포와 정규분포를 이용하여 동일한 평균과 표준편차를 가지고 있는 데이터를 생성하여 왜도의 차이를 분석하였다. 정규분포는 통계적 품질관리에서 가장 널리 사용되는 분포이며, 감마분포는 모수값에 따라 좌우 대칭과 비대칭을 만들 수 있는 분포로 각 특성을 다음과 같다.

정규분포는 다양한 분야에서 확률분포로 가장 많이 사용되는 분포이면서 특히 제조업 분야에 특히 중요한 분포 중 하나이다. 정규분포는 다음과 같은 특징으로 제조업 현장을 중심으로 기업 현장에서 매우 효과적으로 사용되어 왔다.

첫째, 중심극한정리에 따르면, 여러 독립적인 확률 변수의 합은 정규분포에 가까워진다. 이것은 제조 프로세스에서 여러 요소가 결합하여 최종 제품의 품질을 형성할 때 정규분포가 자연스럽게 발생할 수 있다는 근거로 활용된다. 둘째, 정규분포는 평균과 분산으로 분포의 모양이 결정된다. 평균은 분포의 중심을 나타내며, 분산은 데이터의 퍼진 정도를 나타낸다. 이 두 매개변수를 조절함으로써 정규분포의 모양을 다양하게 조절할 수 있고 이는 제조업 현장에서 유용하게 사용되어 왔다. 마지막으로 정규분포는 표준 정규분포(평균이 0이고 분산이 1인 정규분포)로 표준화될 수 있으며, 이는 통계 분석에서 편리하게 사용되었다.

요약하면, 정규분포는 제조업 분야에서 품질 특성을 모델링하고 통계적 품질 관리를 수행하는 데 주로 사용되어 왔고, 그 특성은 중심극한정리, 평균과 분산의 사용 편의성이다. 이러한 특징들은 제조업에서의 품질 향상과 문제 해결에 유용하게 활용되어 왔다. 하지만 정규분포는 대칭성을 가지고 있는 분포이기에 실제 제조 현장에서 발생하는 비대칭적인 상황에 대하여 대처할 수 없는 한계를 가지고 있다.

정규분포의 확률밀도함수, 평균, 분산, 그리고 왜도는 다음과 같다(Nadarajah, 2005).

$$\text{(Probability Density Function, PDF): } f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

여기서:

$f(x; \mu, \sigma)$: 확률밀도함수

μ : 정규분포의 평균, 분포의 중심을 정의

σ : 정규분포의 표준편차, 분포 데이터의 퍼진 정도를 정의

$$\gamma_1 = E\left[\left(\frac{x-\mu}{\sigma}\right)^3\right] = 0$$

확률변수 X의 왜도는 다음과 같은 3차 표준 적률로 정규분포의 왜도는 “0”이다

정규분포의 평균(μ)은 확률밀도함수의 중심을 나타내며, 확률 변수의 평균값을 나타낸다. 정규분포의 평균은 확률 밀도함수의 꼭대기 지점에 해당한다. 정규분포의 분산(σ^2)은 확률 변수의 변동성을 나타낸다. 분산이 클수록 데이터가 평균에서 멀리 퍼져있고, 분산이 작을수록 데이터가 평균 주변에 밀집되어 있음을 나타낸다. 정규분포의 분산은 확률밀도함수의 넓이에 영향을 미친다. 왜도는 확률분포의 비대칭 정도를 나타낸다. 따라서 좌우 대칭성을 가지고 있는 정규분포의 왜도는 0이다. 왜도의 값에서 양수의 왜도는 오른쪽으로 긴 꼬리를 나타내며, 음수의 왜도는 왼쪽으로 긴 꼬리를 나타낸다.

감마 확률분포(Gamma Probability Distribution)는 확률 이론과 통계 분야에서 사용되는 확률분포 중 하나이다. 이 분포는 양수 값을 가지는 연속 확률분포로, 형상모수 값(k)이 증가하면 분포의 모양이 대칭에 가까워지는 특성을 가지고 있다. 주로 양수 범위에서 확률 변수의 확률 분포를 모델링하거나 설명하는 데 사용된다. 감마 분포의 주요 특징과 확률밀도함수는 다음과 같다.

감마 확률분포는 두 개의 매개변수를 가지며, 일반적으로 형상 모수(shape parameter)인 k 와 비율 모수(scale parameter)인 θ 로 나타냅니다. 이러한 매개변수를 조절함으로써 감마 확률분포의 모양을 조절할 수 있다. 감마확률 분포는 양수 값을 가지는 연속 확률분포로, 양수 범위에서 확률 변수의 분포를 모델링하는 데 사용된다. 감마 확률분포는 다양한 응용 분야에서 사용되며, 그 확률밀도함수와 평균, 분산, 왜도는 다음과 같이 계산된다(Dubey, 1970).

감마 확률밀도함수 (Probability Density Function, PDF):

$$f(x; k, \theta) = f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{(\alpha-1)} e^{-\frac{x}{\beta}}, & x > 0 \\ 0, & elsewhere \end{cases} \quad (\text{단, } \alpha > 0, \beta > 0)$$

여기서:

$f(x; k, \theta)$: x에서의 확률밀도함수

k : 형상 모수(shape parameter)로 양수이며, 분포의 모양을 결정

θ : 비율 모수(scale parameter)로 양수이며, 분포의 폭과 위치를 결정

$\Gamma(\alpha)$: 감마 함수(Gamma function), $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$

감마 분포의 평균은 형상 모수(α)와 비율 모수(β)를 사용하여 다음과 같이 계산된다.

$$\mu = \alpha\beta$$

감마 분포의 분산은 형상 모수(α)와 비율 모수(β)를 사용하여 다음과 같이 계산된다.

$$\sigma^2 = \alpha\beta^2$$

감마 분포의 왜도는 정규분포와는 다르게 양수 또는 음수일 수 있다. 왜도는 분포의 비대칭 정도를 나타내며, 아래와 같이 계산된다.

$$\gamma_1 = E\left[\left(\frac{x - \mu}{\sigma}\right)^3\right] = \frac{2}{\sqrt{\alpha}}$$

따라서, α 값이 증가하면 왜도는 0에 가까워지고 분포가 더 대칭에 가까워진다. 즉, 형상 모수 α 가 증가하면 감마 분포는 더 정규분포에 가까워진다. 따라서 본 연구에서는 정규분포와 감마분포를 이용하여 대칭성을 유지할 때와 대칭성을 유지하지 못할 경우(왜도값에 따른 변화)를 파악하였다. Table 1은 20개의 세트를 평균한 데이터의 요약이다. 세트별 랜덤 데이터는 Minitab 18을 이용하여 1,000를 생성하였다.

Table 1. Comparison of gamma distribution and normal distribution

| # of set (# of random numbers) | mean | standard deviation | gamma dist. | | | | normal dist. | |
|---|---------|-----------------------|------------------------------------|-----------------------------------|----------------------------|---|----------------------------|---|
| | | | shape parameter (α) | scale parameter (β) | skewness (γ_1) | > average # of $\pm 3\sigma$ ($\times 10^{-3}$) | skewness (γ_1) | > average # of $\pm 3\sigma$ ($\times 10^{-3}$) |
| 20(1,000) | 100,000 | 1,000 | 10,000 | 10 | 0.0149 | 3.00 | 0.0006 | 2.50 |
| 20(1,000) | 1,000 | 100 | 100 | 10 | 0.2052 | 2.60 | 0.0019 | 3.00 |
| 20(1,000) | 100 | 31.6 | 10 | 10 | 0.6588 | 7.00 | 0.0011 | 2.05 |
| 20(1,000) | 90 | 30 | 9 | 10 | 0.6349 | 7.05 | -0.0424 | 2.75 |
| 20(1,000) | 80 | 28.3 | 8 | 10 | 0.6646 | 7.55 | -0.0018 | 3.00 |
| 20(1,000) | 70 | 26.5 | 7 | 10 | 0.7649 | 8.00 | -0.0037 | 3.20 |
| 20(1,000) | 60 | 24.5 | 6 | 10 | 0.8352 | 8.90 | 0.0028 | 2.65 |
| 20(1,000) | 50 | 22.4 | 5 | 10 | 0.9010 | 9.70 | -0.0114 | 2.90 |
| 20(1,000) | 40 | 20 | 4 | 10 | 0.9616 | 10.10 | 0.0005 | 2.50 |
| 20(1,000) | 30 | 17.3 | 3 | 10 | 1.1701 | 12.65 | 0.0128 | 2.90 |
| 20(1,000) | 20 | 14.1 | 2 | 10 | 1.4135 | 13.40 | -0.0139 | 2.80 |
| 20(1,000) | 10 | 10 | 1 | 10 | 2.0255 | 17.85 | -0.0134 | 2.70 |
| average | | | | | 0.85418 | 8.98333 | -0.00556 | 2.74583 |

table 1에서 나타난 것과 같이 정규분포는 관리도에서 활용하는 관리한계($\pm 3\sigma$)를 넘어설 확률이 2.74583으로 정규분포의 계산된 확률값과 다르지 않음을 확인할 수 있었으나 감마분포는 왜도가 증가됨에 따라 관리한계($\pm 3\sigma$)를 넘어설 확률이 증가됨을 파악할 수 있었다. 이를 바탕으로 회귀분석을 실시한 결과 ANOVA 분석결과 유의확률이 0.000으로 회귀식이 매우 유의하다는 것을 알 수 있다. 또한 기여율(R^2)이 0.98로 설명력도 매우 높은 것으로 나타났다.

Table 2. Regression Model Summary

| Model Summary | | | | |
|---------------|---------|----------|-------------------|----------------------------|
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | .990(a) | .980 | .978 | .63116 |

a. Predictors: (Constant), skewness

Table 3. ANOVA analysis

| ANOVA(a) | | | | | | |
|----------|------------|----------------|----|-------------|---------|---------|
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 196.593 | 1 | 196.593 | 493.505 | .000(b) |
| | Residual | 3.984 | 10 | .398 | | |
| | Total | 200.577 | 11 | | | |

a. Dependent Variable: average # of $\pm 3\sigma$

b. Predictors: (Constant), skewness

Table 4. Coefficients

| Coefficients(a) | | | | | | | | |
|-----------------|------------|-----------------------------|------------|---------------------------|--------|------|---------------------------------|-------------|
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | 95.0% Confidence Interval for B | |
| | | B | Std. Error | Beta | | | Lower Bound | Upper Bound |
| 1 | (Constant) | 2.131 | .358 | | 5.947 | .000 | 1.332270 | 2.928802 |
| | skewness | 8.023 | .361 | .990 | 22.215 | .000 | 7.217970 | 8.827292 |

a. Dependent Variable: average # of $\pm 3\sigma$

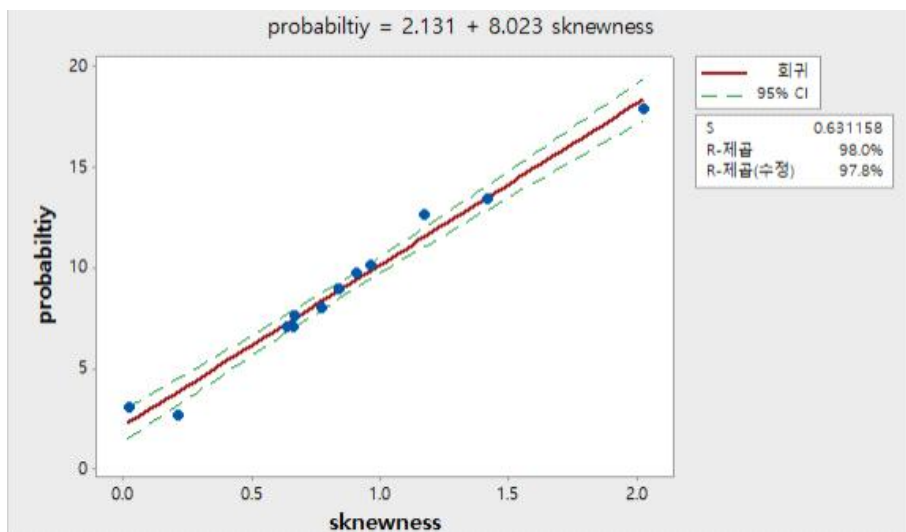


Figure 3. Fitted line plot

회귀분석식의 95% 신뢰구간에 따라 관리한계($\pm 3\sigma$)를 넘어설 확률인 0.9973을 넘어설 왜도는 0.2를 넘어설 경우이다. 따라서 왜도가 0.2 이내일 경우는 95%의 유의확률에서 정규분포(좌우 대칭)를 가정한 관리한계선을 사용하는 것과 같은 불량률 발생 확률이 있지만 0.2보다 큰 값을 가질 경우는 동일한 산포에서도 관리한계를 벗어나는 다른 확률을 가지기 때문에 왜도를 고려해야 한다(table 5).

Table 4. Upper and lower bound limits according to skewness

| skewness | Lower Bound | Upper Bound |
|-------------|-----------------|-----------------|
| 0.1 | 2.054067 | 3.650599 |
| 0.11 | 2.126247 | 3.722779 |
| 0.12 | 2.198426 | 3.794958 |
| 0.13 | 2.270606 | 3.867138 |
| 0.14 | 2.342786 | 3.939318 |
| 0.15 | 2.414966 | 4.011498 |
| 0.16 | 2.487145 | 4.083677 |
| 0.17 | 2.559325 | 4.155857 |
| 0.18 | 2.631505 | 4.228037 |
| 0.19 | 2.703684 | 4.300216 |
| 0.2 | 2.775864 | 4.372396 |
| 0.21 | 2.848044 | 4.444576 |
| 0.22 | 2.920223 | 4.516755 |

4. 왜도를 이용한 관리도

본 연구의 분석결과 왜도가 ± 0.2 이내일 경우 품질관리에서 이야기하는 관리상태에서 범할 수 있는 오류 확률($\mu \pm 3\sigma$)로 관리를 하는 것과 판단의 차이가 없다. 하지만 왜도가 ± 0.2 를 벗어날 경우에는 표준편차만을 고려한 관리도에서는 관리상태라고 판단하고 있지만 데이터는 관리상태를 벗어나는 것을 알 수 있었다. 신뢰구간을 이용하여 95%의 하한값을 확인해보면 0.19가 2.703684×10^{-3} 이고 0.2에서 2.775864×10^{-3} 로 왜도가 0.2를 넘어간다면 실제관리상태의 확률(0.27%)을 벗어난다고 볼 수 있다. 따라서 관리도를 이용한 관리상태를 파악할 때 중심의 변화를 나타내는 \bar{x} -관리도, 분산 또는 표준편차를 관리하기 위한 "R-관리도" 또는 "S-관리도"와 함께 왜도의 변화를 파악하는 γ 1-관리도를 사용해야 하며 관리선은 중심과 산포를 이용하는 $\pm 3\sigma$ 가 아닌 ± 0.2 로 고정되어 사용되어야 공정의 안정상태를 파악할 수 있다.

Table 5는 평균이 형상모수가 100, 척도모수가 10인 감마 분포를 이용하여 각 1,000개씩 20세트의 확률변수값을 생성한 데이터 세트이다. 이를 이용하여 관리도를 그린 것이 Figure 4로 관리상태에 있는 것으로 파악할 수 있다. 하지만 데이터에서는 $\pm 3\sigma$ 를 벗어나는 경우가 총 52건이 발생하였다(table 5). 이것은 기존의 관리도에서는 데이터의 변화를 민감하게 찾아내지 못했다는 것을 의미하는 것이다.

반면에 Figure 4 는 평균이 왜도가 +2 이상이 상태가 발생하는 부분과 왜도의 변화를 통하여 관리상태가 아닌 이상상태임을 제시하고 있다. 따라서 왜도를 이용하여 관리가 이루어진다면 평균과 분산보다 민감한 데이터 관리가 가능하다.

Table 5. Random number of Gamma distribution($k=100, \Theta=10$)

| | mean | standard deviation | skewness | # of $<\mu-3\sigma$ | # of $>\mu+3\sigma$ | sum |
|---------|----------|--------------------|----------|---------------------|---------------------|-----|
| | sum | | | 7 | 45 | 52 |
| 1 | 1003.787 | 102.218 | 0.083783 | 1 | 4 | 5 |
| 2 | 1001.652 | 98.59587 | 0.198356 | 0 | 0 | 0 |
| 3 | 1003.56 | 102.2829 | 0.200215 | 0 | 4 | 4 |
| 4 | 1003.277 | 99.10767 | 0.151418 | 0 | 1 | 1 |
| 5 | 995.2962 | 94.57049 | 0.21924 | 1 | 3 | 4 |
| 6 | 993.405 | 97.19307 | 0.166442 | 0 | 2 | 2 |
| 7 | 1003.65 | 98.3479 | 0.168025 | 0 | 1 | 1 |
| 8 | 999.9735 | 99.4902 | 0.197349 | 0 | 1 | 1 |
| 9 | 998.371 | 103.1957 | 0.251122 | 0 | 3 | 3 |
| 10 | 999.3483 | 92.43896 | 0.18491 | 0 | 1 | 1 |
| 11 | 1001.356 | 98.53658 | 0.215135 | 0 | 1 | 1 |
| 12 | 997.5084 | 99.48092 | 0.183616 | 1 | 4 | 5 |
| 13 | 1001.602 | 97.92063 | 0.115319 | 0 | 1 | 1 |
| 14 | 1000.342 | 100.0848 | 0.32827 | 0 | 2 | 2 |
| 15 | 999.1622 | 99.68311 | 0.207499 | 0 | 1 | 1 |
| 16 | 999.3711 | 101.0565 | 0.271426 | 1 | 2 | 3 |
| 17 | 1003.861 | 101.4296 | 0.210359 | 2 | 1 | 3 |
| 18 | 1000.346 | 100.7263 | 0.295861 | 0 | 5 | 5 |
| 19 | 997.3933 | 95.84669 | 0.228548 | 1 | 5 | 6 |
| 20 | 998.1693 | 101.0289 | 0.228054 | 0 | 3 | 3 |
| average | 1000.072 | 99.16173 | 0.205247 | 0.35 | 2.25 | 2.6 |

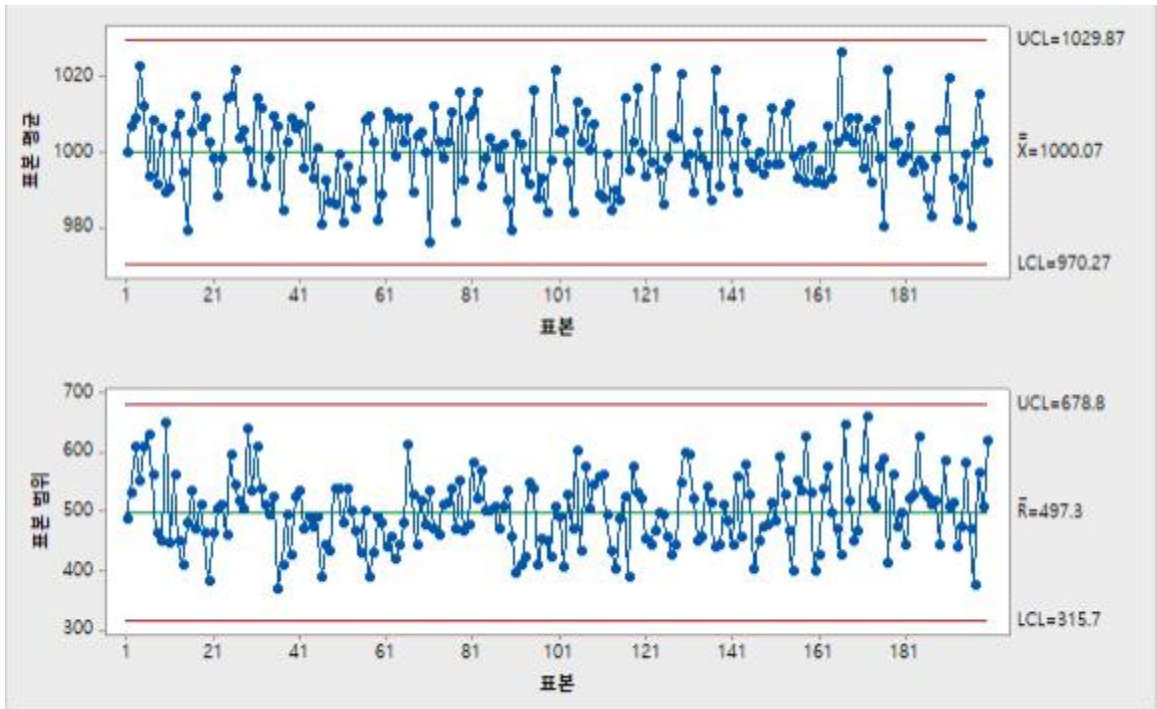


Figure 3. $\bar{x} - R$ control chart

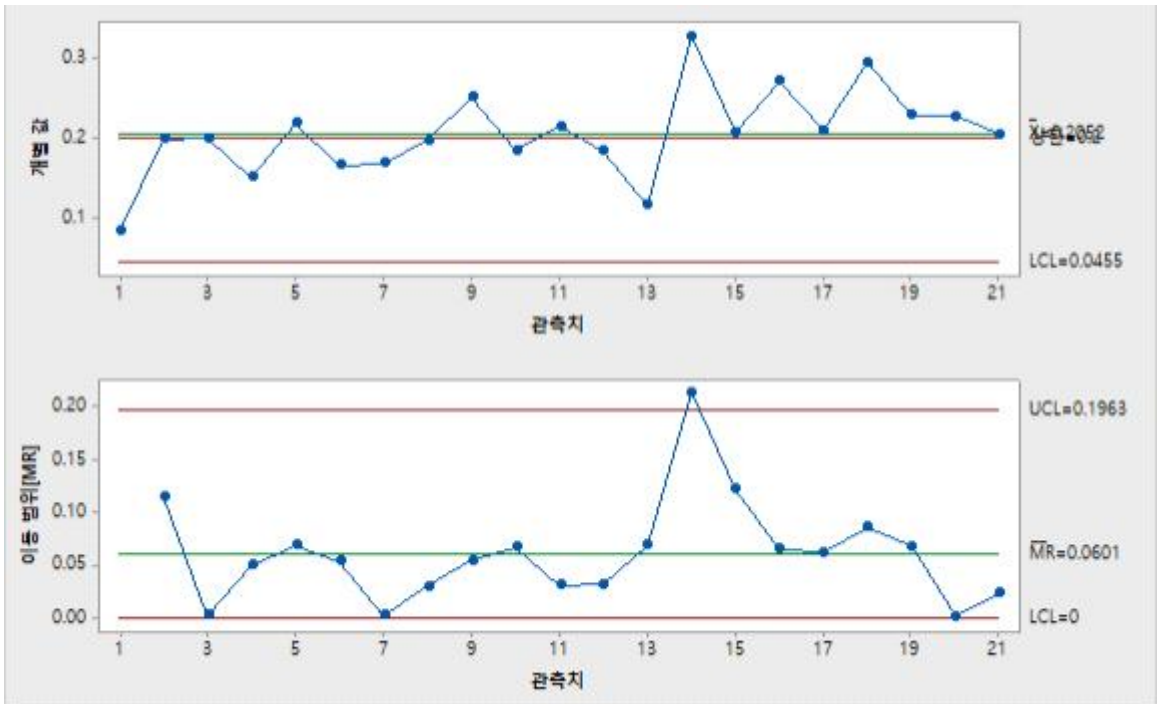


Figure 4. $\nu_1 - R$ control chart

5. 결 론

이번 연구는 왜도를 이용한 공정관리의 필요성을 제시하였다. 왜도(Skewness)는 확률분포의 비대칭 정도를 나타내는 통계적 척도이다. 공정관리에서 왜도를 이용한다면 다음과 같은 장점이 있다. 첫째, 데이터 품질 평가가 우수해진다. 왜도는 데이터 분포의 비대칭 정도를 측정하므로, 이를 통해 데이터의 품질을 평가할 수 있다. 대칭 분포에 가까운 데이터와 비대칭에 해당하는 다른 특성을 빠르게 파악할 수 있다. 둘째, 품질 이상 신호 탐지능력이 탁월해진다. 왜도는 데이터 분포가 대칭이 아닐 때 품질 이상 신호(불량, 결함 등)를 탐지하는 데 우수하다. 데이터 분포의 비대칭성은 문제가 발생한 가능성을 나타낼 수 있으며, 이로써 이상 신호를 조기에 감지할 수 있다. 셋째, 프로세스를 개선하는 데 도움을 준다. 왜도를 모니터링하고 분석함으로써 공정의 비대칭성 문제를 식별하고 개선할 수 있다. 비대칭 분포는 잘못된 조정 또는 기계 상태의 문제를 나타낼 수 있으며, 개선 기회를 제공한다. 마지막으로 품질 예측이 가능해진다. 왜도 정보는 미래 품질을 예측하는 데 사용될 수 있다. 예를 들어, 데이터 분포가 오른쪽 또는 왼쪽으로 긴 꼬리를 가질 때, 더 많은 불량품이 발생할 가능성이 있다. 이를 빠르게 예측할 수 있다는 것이다. 하지만 본 연구는 감마 분포와 정규분포를 이용하여 난수를 발생시켜서 만들어진 데이터를 사용하였다는 한계점을 가지고 있다. 실제 산업현장에서 이러한 분석이 얼마나 정확하게 사용될 수 있는지 파악하는 것이 필요하다. 또한 왜도의 변화 포인트를 ± 2 를 사용한 기준이 기존의 관리를 벗어날 1종 과오의 $\pm 3\sigma$ 확률을 적용하였으나 그 이유의 타당성이 부족하다. 따라서 향후 실제 데이터를 사용하여 관리상태의 유무를 파악하는 연구가 필요하다.

REFERENCES

- Cho, J. J., Yu, H. K., and Han, J. S. 2010. Test of Hypothesis in Assessing Process Capability Index Cpmk. *CSAM (Communications for Statistical Applications and Methods)* 17(3):459–471.
- Dubey, S. D. 1970. Compound gamma, beta and F distributions. *Metrika* 16(1):27–31.
- Kim, J. B., Yun, W. Y., and Seo, S. K. 2022. Development of a New Index to Assess the Process Stability. *Journal of Korean Society for Quality Management* 50(3):473–490.
- Lee, D. K. 2009. A Study on Expression of Process Capability and Sigma Level. *Journal of Society of Korea Industrial and Systems Engineering* 32(1):111–116.
- Lee, S. B., and Park, R. G. 2012. Propose Modified Control Chart For Management In Real Field. *Asia-Pacific Journal of Business Venturing and Entrepreneurship* 7(2):151–156.
- Nadarajah, S. 2005. A generalized normal distribution. *Journal of Applied statistics* 32(7):685–694.
- Seo, H. J., Byun, J. H., and Kim, D. H. 2021. Quality 4.0: Concept, Elements, Level Evaluation and Deployment Direction. *Journal of Korean Society for Quality Management* 49(4):447–466.
- Song, H. J., Kim, M. G., and Shin, W. S. 2021. A Study on the Relative Importance of Quality Scorecard(QSC) for Open Quality. *Journal of Korean Society for Quality Management* 49(1):61–79.

저자소개

김훈태 서울대학교 대학원에서 산업공학 박사학위를 취득하였으며, 현재 대진대학교 산업경영공학과 교수로 재직 중이다. 주요 관심분야는 생산 프로세스 혁신, 스마트공장 등이다.

임성욱 성균관대학교 대학원에서 산업공학 박사학위를 취득하였으며, 현재 대진대학교 산업경영공학과 교수로 재직 중이다. 주요 관심분야는 6시그마 경영혁신, 품질경영, 품질지표 등이다.