비정규 공정에서 비모수 통계의 적용
- Implementation of Nonparametric Statistics in the Non-Normal Process-

최 성 운*

Abstract

Based on latest research, the parametric quality statistics cannot be used in non-normal process with demand pattern of many-variety and small-volume, since it involves extremely small sample size. The research proposes nonparametric quality statistics according to the number of lot or batch in the non-normal process. Additionally, the nonparametric Process Capability Index (PCI) is used with 14 identified non-normal distributions.

Keywords: Demand Pattern, Many-Variety and Small Volume, Non-Normal Process, Nonparametric, Number of Lot or Batch, Nonparametric PCI, Non-Normal Distributions

1. 서 론

품질개선시 제품 또는 데이터를 집단화(Grouping)하거나 유형화(Classification)하여 문제 차원이나 모형을 간소화하는 방법을 많이 활용한다. 집단화하는 방식으로 층별(Stratification)과 취락(Clustering)을 사용하는데 각각 집단내를 일정하게 하고 집단간 차이를 나타내는 방식, 집단내를 차이나게 하고 집단간을 일정하게 하는 방식이다. 특히 층별은 생산과정에서 작업내용을 4M(Man, Machine, Material, Method)별로 작업일지에 구분, 정리하여 로트(Lot) 및 배취(Batch)를 추적하여 불량에 대한 원인을 개선하려는 경우 사용되는 중요 품질도구이다.

* 가천대학교 산업공학과

수치인 데이터를 그림으로 알기 쉽게 표현한 것이 분포(Distribution)이다. 확률은 스펙과 데이터에 해당하는 샘플의 크기가 서로 다른 경우 1.0점 만점으로 비교하기 위한 상대척도의 개념이다. 따라서 확률분포는 스펙별 샘플의 크기가 다른 로트들의 수치를 그림으로 비교하기 위해 사용하는 품질도구로 이론적으로 가정한 모양에 실제 산포가 적합한가를 가설검정(Hypothesis Test)을 통해 확인해야만 한다. 이런 분포 중 좌우대칭인 종모양이고(Symmetric Bell-Shaped), μ, σ 두 모수를 모두 사전에 축적하여 알아야 하는 부지런한 분포인 정규분포(Normal Distribution)가 가장 많이 사용된다. 정규분포는 취락된 모양으로 평균에 가장 많고 데이터가 극단적으로 크거나 작은 확률이 적게 나타나는 사람의 키를 나타내는 경우 별의별 사람이 다 모인 정상적인 사회 구성원의 비율을 나타내는 분포이다. 따라서 SQC(Statistical Quality Control), DOE(Design of Experiment)에서는 샘플의 크기를 크게 하는 중심극한정리(Central Limit Theorem)를 이용하여 정규성 근사(Nomal Approximation) 방식의 품질통계 기법 등을 활용한다.

그러나 최근 수요패턴이 소품종 대량수요에서 다품종 소량수요 유형으로 바뀜에 따라 샘플의 크기가 적어지거나[1,2], 샘플의 크기가 클 경우라도 공정의 특성, 스펙의 종류에 따라 정규분포를 만족하지 못하는 경우 비모수, 비정규 근사방법[3,4]을 활용해야 한다.

따라서 본 연구에서는 모수 가절검정[5]에서 정규성을 만족하지 못하는 경우 모집단수 유형별 비모수검정(Nonparametric Test)의 적용방안을 제시한다. 또한 공정능력지수를 구하는 경우 스펙의 유형에 따른 개별분포방식과 정규변수변환방법을 제시하고 비정규 근사 모부적합품률 검정방식에 대해 논의한다.

본 연구의 기여도는 실무와의 이해를 돕기 위해 정규성 가정을 만족하지 못할 경우 모수검정별 비모수 검정 방식을 유형별로 제시하고, 한쪽규격인 경우 공정능력지수 적용방안을 제시한 데 있다.

2. 비모수 검정

분포(Distribution)를 가정하는 모수(Parametric) 검정에는 평균에 의해 축위치(정확도) 검정과 산포에 의한 폭넓이(정밀도) 검정이 있으며 분포집단간 비교시 등분산에 의한 폭넓이를 일정하다고 확인된 상태에서 축위치인 평균의 비교를 하게 된다.

그러나 비모수(Nonparametric) 검정은 정규분포의 가정을 만족하지 못하나 샘플의 크기가 작은 경우 순위(Rank)에 의한 통계량으로 검정을 실시한다. <표 1>에서 Sign Test는 Z검정, t검정 모수검정의 비모수방법으로 사용되며 Median을 가설로 중앙값을 초과하는 데이터수 n^+로 검정통계량을 생성하여 판정한다. Wilcoxon Signed Rank Test는 Z검정, t검정, Paired t검정, 모수검정의 비모수방법으로 자신을 포함한 Walsh Average $\frac{(x_i+x_j)}{2}$가 Median을 초과하는 개수로 검정통계량값을 생성한다. <표 1>에서 Mann-Whitney Test는 합동분산인 Pooled t차 모수 검정의 비모수 방법으로 두 개의 모집단을 합하여 순위를 구한 후 첫 번째 모집단의 순위의 합으로 검정통계량값을 구한다.

<표 1> 모집단별 (Lot별, Batch별) 비모수 축의 검정

1개 모집단(Lot, Batch)	2개 모집단(Lot, Batch)	3개 이상의 모집단 (Lot, Batch)
· Sign Test · Wilcoxon Signed Rank Test	· Mann-Whitney Test	· Kruskall-Wallis Test · Mood's Median Test · Friedman Test

<표 1>에서 3개이상의 모집단에 대한 비모수 방법은 모수검정에서 실험계획법의 ANOVA에 해당되며 Kruskall-Wallis Test는 3개이상의 모집단 전체를 대상으로 한 순위로 검정통계량을 생성하며 Mood's Median Test는 전체 모집단을 대상으로 한 메디안(Overall Median)으로 2집단으로 수준 m개로 분할된 2×m표로 검정을 실시한다.

난괴법의 모수검정에 해당하는 Friedman Test는 각 블록(Block)마다 순위로 검정통계량을 생성한다.

다품종 소량생산에서 샘플의 크기 n이 작거나 소품종 대량생산에서 n이 크더라도 정규분포를 만족하지 않을 경우 모수검정에 대응하는 <표 1>과 같은 비모수 방법을 사용할 것을 권장한다.

3. 비정규, 비모수 공정능력지수

3.1 비정규 공정능력지수

SPC에서는 공정의 Monitoring 개선에 사용되는 관리도(Controal Chart)와 스펙(Specification)과의 비교를 통해 정규성(Normality)과 공정능력(Process Capability)을 시각적으로 파악하는 히스토그램등의 통계적 기법이 사용된다. SPC 관리도는 계량연속형, 계수이산형 모두 중심극한정리(Central Limit Theorem) 또는 정규근사(Normal Approximation)의 정규분포를 가정하고 있다.

따라서 양쪽스펙을 갖는 제품이 정규공정인 경우 공정능력지수 산출전에 반드시 NPP(Normal Propability Plot 검정, Anderson-Darling 검정, Ryan-Joiner 검정, Kolmogorov-Smirnov 검정 등의 정규성 검정을 실시해야 한다, P-Value가 α보다 작은 경우 C_{pk}, C_{pm}등의 PCI(Process Capability Index)를 사용하면 된다.

그런데 LSL(Lower Specification Limit), USL(Upper Specification Limit)의 한쪽스펙인 경우 히스토그램의 모양은 반쪽이 잘려진 절벽형이므로 MINITAB에서 제공하는 양쪽스펙의 정규성검정을 사용할 수 없다. 이런 경우의 정규성 검정방법에는 변수변한법과 비변환법등 두 가지가 존재하는데 변수변환법에는 데이터가 양수인 경우 λ를 통해 데이터를 정규분포로 변환하는 Box-Cox Transform과 SB(경계있는 분포), SL(로그정규분포), SU(경계없는 분포)등의 3종류 최적함수를 선택하는 Johnson Transform

이 있다. 그러나 이 방법은 데이터가 원래 정규분포이나 Outlier 또는 일부 오염된 데이터로 인해 정규성을 상실한 경우 변수변환을 통해 정규성을 회복시켜주는 방법이다. 한쪽스펙처럼 애초부터 정규분포가 아닌 경우 MINITAB의 품질도구에서 제공하는 14개의 개별분포식별의 적합도 검정을 통해 P-Value가 α보다 큰 분포를 선택해서 3.2절의 비모수 공정능력지수 공식을 이용한다. 정규분포군에는 정규분포, 대수정규분포, 3-모수 대수정규분포, 지수분포군에는 지수분포, 2-모수 지수분포, Weibull 분포군에는 Weibull분포, 3-모수 Weibull분포, 극단값 분포군에는 최대극단값분포, 최소극단값분포, Gamma분포군에는 Gamma분포, 3-모수 Gamma분포, Logistic분포군에는 Logistic분포, 대수 Logistic분포, 3-모수 대수 Logistic 분포가 있다.

품질분임조 QC Story 15단계에서 현상파악에서 비정규분포인 공정능력지수를 구할 경우 대책실시에서도 가설검정, 구간추정, 실험계획법의 모수적 방법이 아닌 2절에서와 같은 비모수 방법을 사용해야 한다. 따라서 현상파악, 결과분석, 사후관리에서도 정규분포를 가정하는 SPC 관리도가 한쪽스펙의 비정규공정에서는 사실상 엄격하게 가정을 적용할 경우 사용이 불가능하다.

3.2 비모수 공정능력지수

3.1절과 같이 14개의 비정규 개별분포로 식별된 경우 공정능력지수 해당분포의 백분위수(Percentile) t_p를 사용한다. $t_{0.00135}$, $t_{0.99865}$, $t_{0.5}$는 각각 해당분포의 0.135%, 99.865%, 50% 백분위수이다. 비모수 공정능력지수에서 모수 공정 능력지수의 $6\hat{a}$는 99.73%의 면적이므로 $6\hat{a}=t_{0.99865}-t_{0.00135}$이고, $3\hat{a}$는 $6\hat{a}$의 절반이므로 LSL, USL인 경우 $3\hat{a}=t_{0.5}-t_{0.00135}=t_{0.99865}-t_{0.5}$이다. 따라서 양쪽스펙인 경우 C_p에 해당하는 비모수 공정능력지수 C_p'=(USL-LSL)/($t_{0.99865}-t_{0.0135}$), 상한스펙인 경우 C_{pu}에 해당하는 비모수 공정능력지수 C_{pu}'=(USL-$t_{0.5}$)/($t_{0.99865}-t_{0.5}$), 하한규격인 경우 C_{pu}에 해당하는 비모수 공정능력지수 C_{pu}'=($t_{0.5}$-LSL)/($t_{0.5}-t_{0.0135}$)이다. 치우침이 있는 C_{pk}에 해당하는 비모수 공정능력지수 C_{pk}'=min { C_{pu}', C_{pc}' } 이며 치우침이 있는 C_{pm}에 해당하는 비모수 공정능력지수 C_{pm}'=(USL-LSL) / $6(1/6(t_{0.99865}-t_{0.00135})+((\mathrm{USL}+\mathrm{LSL})/2-t_{0.5}))^{1/2}$이다.

4. 비정규근사 모부적합품률 추정

계수이산형 데이터는 개개의 스펙을 벗어나는 부적합, 결점(Nonconformities)과 Unit로 벗어나는 부적합품, 불량(Nonconforming Unit)이 있으며 결점이 모여 불량이 된다. 부적합품 분포는 계산의 간략화 관점에서 가장 정확한 초기하 분포, 가장 계산이 간단한 포아송 분포, 중간에 해당하는 이항분포가 있으나, np≥5인 경우는 정규근사에 수렴하는 $P=p\pm Z_{1-\alpha/2}(p(1-p)/n)^{1/2}$의 구간추정식을 사용한다. 그러나 np<5인 경우는 $\phi_2/(\phi_1 F(\phi_1,\phi_2 : \alpha/2)+\phi_2)\le p\le \phi_3 F(\phi_3,\phi_4 : \alpha/2)/(\phi_4+\phi_3 F(\phi_3,\phi_4 : \alpha/2))$, where $\phi_1 = 2(n-r+1)$, $\phi_2 = 2r$, $\phi_3 = 2(r+1)$, $\phi_4 = 2(n-r)$의 비정규근사구간 추정식을 적용한다.

5. 결 론

본 연구에서는 소품종대량 수요패턴의 정규공정에서 사용되는 모수통계의 품질개선기법과 비교하면서 다품종소량 수요패턴의 비정규공정에서 적용 가능한 품질개선기법을 제시하였다. 특히 모집단의 개수에 따른 비모수 메디안 축 위치(정확도)의 검정 방법을 모수 통계 방법과 비교 제시하였고 한쪽스펙인 경우와 같이 애초부터 비정규 분포인 경우 개별분포사용방법과 비모수 공정능력지수의 산출방안을 논의하였다.

6. 참 고 문 헌

[1] 최성운외, “부적합품률의 이항 신뢰구간 추정 및 응용”, 대한안전경영과학회지, 9(4)(2007):143-147.

[2] 최성운, “정량적, 정성적 회귀분석의 모적용과 이해”, 대한안전경영과학회 추계학술대회 발표문집, (2011):213-217.

[3] Gibbons J.D., Chakraborti S., Nonparametric Statistical Inference, 5 Edition, Chapman and Hall/CRC, 2010.

[4] Sprent P., Smeeton N.C., Applied Nonparametric Statistical Methods, 4 Edition, Chapman and Hall/CRC, 2007.

[5] http://www.minitab.com.