

순서형 변수를 위한 관리도

장대홍

부경대학교 수리과학부 통계학전공

Control Charts for Ordinal Variables

Jang, Dae-Heung

Division of Mathematical Sciences, Pukyong National University

Key Words: Ordinal Variables, Control Charts

Abstract

Many practical problems of quality control in service management are derived from the use of ordinal variables. Ordered linguistic variables differ from measurement variables. This paper presents a new control chart of a production process based on ordinal variables.

1. 서론

서비스산업에서는 서비스품질의 평가가 주로 순서형 변수 형태로 주어진다. 예로 호텔 서비스의 품질을 평가하기 위하여 투숙고객들에게 앙케이트 조사를 할 때 서비스의 품질을 평가하기 위한 항목에서 5 가지 범주('매우 나쁘다', '나쁘다', '보통이다', '좋다', '매우 좋다')로 주어져 평가된다. 이렇게 순서형 변수인 경우 우리는 주로 '매우 나쁘다'=1, '나쁘다'=2, '보통이다'=3, '좋다'=4, '매우 좋다'=5 등으로 등간격으로 점수화(codification)하여 계량형 관리도인 $\bar{x} - R$ 관리도를 사용한다. 그런데 이러한 등간격 점수화는 문제를 일으킬 수 있다. 예로 <표 1>과 같이 어느 호텔 투숙객 60명에게 호텔서비스의 품질을 평가하기 위한 특정 항목에서 5 가지 범주('매우 나쁘다', '나쁘다', '보통이다', '좋다', '매우 좋다')로 나누어 평가했다고 하자.

도수	3	12	17	13	15
----	---	----	----	----	----

이러한 범주에 등간격으로 점수화하여 평균을 구하면 $\bar{x} = 3.42$ 가 되어 '보통이다'와 '좋다' 사이 값이 되나 '보통이다'에 가깝게 된다. 그러나 등간격으로 점수화하지 않고 '매우 나쁘다'=1, '나쁘다'=2, '보통이다'=4, '좋다'=8, '매우 좋다'=16 등으로 이간격(등비 $r = 2$ 가 되게)으로 점수화하면 7.32가 되어 '보통이다'와 '좋다' 사이 값이 되나 '좋다'에 가깝게 된다. 즉 같은 자료에 대하여 점수화하는 방법에 따라 중심위치에 해당하는 범주가 다르게 나타나게 된다. 그래서 순서형 변수에 기존의 계량형 관리도인 $\bar{x} - R$ 관리도를 사용하는 것은 공정 이상의 판단에 문제를 일으킬 수 있다.

위와 같은 문제점을 개선하기 위하여 Wang 과 Raz(1990), Franceschini와 Romano(1999), Franceschini, Galetto와 Varetto(2005) 등은 순서형 변수를 위한 관리도들을 제안하였다. 그러나 이러한 관리도들은 계산이 복잡하거나 이론상 문제점이 있다. 그래서 순서형 변수에서

<표 1> 호텔서비스의 품질 평가 결과

범주	매우나쁘다	나쁘다	보통이다	좋다	매우좋다
----	-------	-----	------	----	------

중심위치와 산포도를 간단히 계산할 수 있는 방법을 이용하여 새로운 관리도를 제안하고자 한다.

2. 순서형 변수를 위한 관리도

순서형 변수에서 중심위치를 나타내는 척도로서는 우리가 잘 알고 있는 중앙값 \tilde{x} 를 이용할 수 있다. 순서형 변수에서는 순서통계량을 이용한 중심위치 척도로서 중앙값이 좋은 척도가 된다. 중앙값은 일반적으로 범주 중의 하나가 된다. Yager와 Filev(1994), Franceschini와 Romano(1999), Franceschini, Galetto와 Varetto(2004) 등이 제안한 OWA(Ordered weighted average) emulator나 Franceschini, Galetto와 Varetto(2005)이 제안한 dominance criterion은 계산이 복잡한 반면 중앙값보다 큰 장점이 없다.

순서형 변수에서 산포도를 나타내는 척도로서는 Pielou(1966)가 제안한 다양성지수(diversity index)를 이용할 수 있다. 우리는 균등지수(evenness index)를 다음과 같이 정의할 수 있다.

$$J = \frac{H}{H_{\max}}$$

여기서, $H = \frac{n \log n - \sum_{i=1}^q f_i \log f_i}{n}$, $H_{\max} = \log q$,

q 는 범주의 개수이고, f_i 는 로트의 크기가 n 인 임의의 로트에서 범주 i 에 속하는 빈도수이다. 이 균등지수는 로트의 크기가 다르더라도 각 범주의 비율이 같으면 불변하다(invariant)라는 좋은 성질을 갖고 있다. 다음 <표 2>는 이러한 사실을 나타낸다. 균등지수는 $0 \leq J \leq 1$ 을 만족한다. 중앙값도 이러한 불변의 성질을 갖고 있다.

<표 2> 균등지수의 계산

순서형 변수에서 중심위치를 나타내는 척도로서는 중앙값 \tilde{x} 을 쓰고 산포도를 나타내는

척도로서는 균등지수 J 를 쓰는 관리도를 앞으로 $\tilde{x} - J$ 관리도라 부르자. 로트의 개수가 k 라고 하고 $i(i=1, 2, \dots, k)$ 번째 중앙값과 균등지수를 각각 \tilde{x}_i 와 J_i 라 하자. \tilde{x} 관리도에서 중심선 CL은 $med(\tilde{x}_i, i=1, 2, \dots, k)$ 또는 k 개

표본의 크기	범주/도수				J
	1	2	3	4	
20	5	5	5	5	1
40	10	10	10	10	1
20	1	1	1	17	0.42
40	2	2	2	34	0.42
20	0	0	0	20	0
40	0	0	0	40	0

로트 전체의 중앙값을 쓸 수 있다. 순서형 변수에서는 중앙값의 분포를 모르므로 관리한계선(UCL, LCL)은 k 개 로트 전체의 분포를 이용하여 구한다. 더 정확한 관리한계선을 구하려면 붓스트랩 방법을 이용할 수 있다. J 관리

도에서 중심선 CL은 $\bar{J} = \frac{\sum_{i=1}^k J_i}{k}$ 또는 k 개 로트 전체의 균등지수를 쓸 수 있다. 균등지수 J 는 $0 \leq J \leq 1$ 을 만족하므로 관리한계선을 따로 두지 않는다. 관리한계선을 구하려면 붓스트랩 방법을 이용할 수 있다.

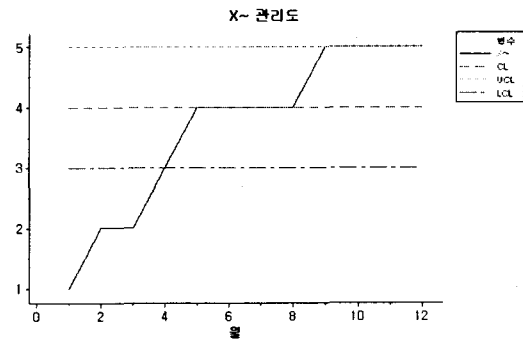
3. 수치 예

<표 3>과 같이 어느 호텔에서 서비스품질개선을 위한 식스시그마운동을 전개하며 이 운동이 효과가 있는 지를 확인하기 위하여 투숙객 100명에게 호텔서비스의 품질을 평가하기 위한 특정 항목에서 5 가지 범주('매우 나쁘다', '나쁘다', '보통이다', '좋다', '매우 좋다')로 나누어 매 월 단위로 1년 동안 평가했다고 가정하자.

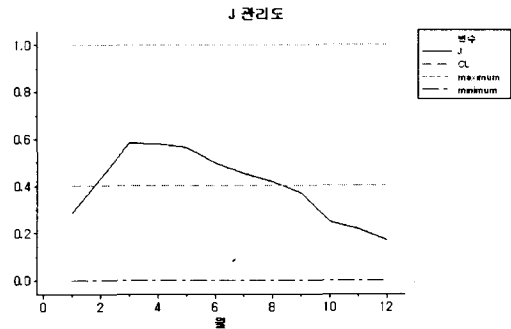
<표 3> 호텔서비스의 품질 평가 결과

월	범주/도수				
	매우나쁘다	나쁘다	보통이다	좋다	매우좋다
1	63	37	0	0	0

2	41	45	14	0	0
3	27	31	29	13	0
4	15	19	36	30	0
5	5	7	35	43	10
6	0	2	21	39	38
7	0	0	18	42	40
8	0	0	11	40	49
9	0	0	7	32	61
10	0	0	0	26	74
11	0	0	0	20	80
12	0	0	0	13	87



\tilde{x} 관리도에서 중심선 CL은 $med(\tilde{x}_i, i=1, 2, \dots, 12)$ 을 이용하여 구하였고 관리한계선(UCL, LCL)은 12개월 전체의 분포를 이용하여 구하였다. 중심선 CL은 '좋다'라는 범주가 된다. 12개월 전체의 표본개수는 1,200명이고, 이 표본들 중 '보통이다', '좋다', '매우 좋다' 범주에 속하는 빈도수는 각각 171명, 298명, 439명이어서 이 세 개의 범주들에 속하는 비율이 75.7%이므로 관리한계선 LCL을 '보통이다'라는 범주로, 관리한계선 UCL을 '매우 좋다'라는 범주로 정할 수 있다.



<그림 1> $\tilde{x} - J$ 관리도

J 관리도에서 중심선 CL은 $\bar{J} = \frac{\sum_{i=1}^{12} J_i}{12}$ 를 이용하여 구하였다. 중심선 CL은 0.4009이 된다. 이들을 이용하여 $\tilde{x} - J$ 관리도를 그리면 다음 <그림 1>과 같았다. \tilde{x} 관리도를 통하여 각 월의 중앙값이 '매우 나쁘다'(1월) → '나쁘다'(2, 3월) → '보통이다'(4월) → '좋다'(5-8월) → '매우 좋다'(9-12월)로 계속 상승하고 있음을 알 수 있다. J 관리도를 통하여 산포도가 1월에서 3월까지 급격히 커지다가 4월부터 12월까지 점차적으로 작아짐을 알 수 있다. 특히 9월부터는 CL 이하로 작아짐을 알 수 있다.

4. 결론

서비스산업에서는 서비스품질의 평가가 주로 순서형 변수 형태로 주어진다. 이러한 순서형 변수에 기존의 계량형 관리도인 $\tilde{x} - R$ 관리도를 사용하는 것은 공정 이상의 판단에 문제를 일으킬 수 있다. 그래서 우리는 순서형 변수에 적합한 $\tilde{x} - J$ 관리도를 제안할 수 있다.

(1:매우 나쁘다, 2:나쁘다, 3:보통이다, 4:좋다, 5:매우 좋다)

참고문헌

- [1] Franceschini, F., Galetto, M. and Varetto, M.(2004). "Qualitative Ordinal Scales: The Concept of Ordinal Range", *Quality Engineering*, Vol. 16, 515-524.
- [2] Franceschini, F., Galetto, M. and Varetto, M.(2005). "Ordered Samples Control Charts for Ordinal Variables", *Quality*

and Reliability Engineering International, Vol. 21, 177-195.

- [3] Franceschini, F. and Romano, D.(1999). "Control Charts for Linguistic Variables: A Method Based on the Use of Linguistic Quantifiers", *International Journal of Production Research*, Vol. 37, 3791-3801.
- [4] Pielou, E. C.(1966), "The Measurement of Diversity in Different Types of Biological Collections", *Journal of Theoretical Biology*, Vol. 13, 131-144.
- [5] Wang, J-H. and Raz, T.(1990), "On the Construction of Control Charts Using Linguistic Variables", *International Journal of Production Research*, Vol. 28, 477-487.
- [6] Yager, R. R. and Filiv, D. P.(1994), *Essentials of Fuzzy Modeling and Control*, John Wiley & Sons, Inc., New York.