

☒ 연구논문

측정시스템 평가에 관한 연구⁺

백재욱

한국방송대학교 응용통계학과

A Study for the Measurement System Assessment

Jai Wook Baik

Korea National Open University

Abstract

It is essential to have a correct measurement system assessment to help improve process or quality. In this paper, we emphasize that a confidence interval approach for the parameters of measurement error study and GR&R is crucial and a simple approach for constructing them is presented. We also discuss on the issue of linearity of the gauge. Next, we try to reveal the relationship between measurement error and the process capability index. Finally some aspects of conducting the measuring experiment concerning the number of parts, number of measurements per part, and recommendations on how the experiment is to be performed will be considered.

⁺ 이 논문은 1997년 한국학술진흥재단의 공모과제 연구비에 의하여 연구되었음.

1. 서론

통계적 공정관리 추진시 데이터의 신뢰성 확보는 무엇보다 중요하며, 이를 위해 여러 가지 통계적 방법을 이용할 수 있다. 그 중에서 백재욱과 조진남(1996)에서 설명하는 Measurement Error Study, GR&R(gauge repeatability and reproducibility), 실험 계획법적 접근방법은 현장에서 많이 쓰이는 방법에 속한다.

Floyd and Laurent(1995)가 설명하는 첫 번째 방법은 아주 간단하지만 통상적으로 측정결과를 점추정치 하나로만 요약한다는 단점이 있다. 따라서 본 연구에서는 중심 극한정리를 이용하여 해당 모수에 대한 간단한 구간추정 방법을 소개한다.

다음으로 GR&R 분석에서도 통상적으로 $\sigma_{repeatability}$ 와 $\sigma_{reproducibility}$ 에 대한 점추정치만을 구하는데, 본 연구에서는 구간추정의 필요성과 간단한 구간추정방법에 대해서 먼저 살펴본다. 아울러 GR&R 분석에서는 분포의 위치(location)와 관련하여 안정성(stability), 정확성(accuracy)[이를 어떤 책에서는 편의(bias)라고도 함], 직선성(linearity)에 대해 논하고, 분포의 산포(spread)와 관련하여 정밀도(repeatability)[이를 어떤 책에서는 반복성이라고도 함] 및 재현성(reproducibility)에 대해 논한다. 그런데 민철희와 백재욱(1998)에서 안정성, 정확성, 정밀도, 재현성에 대해서 이미 논했으므로 본 연구에서는 직선성에 대해서만 간단히 고찰하고자 한다.

한편, Deming(1982)은 제품의 품질을 향상시키기 위해서는 공정의 변동(process variability)에 대한 이해가 필요하다고 하는데, 이 변동은 크게 측정과정 중에서 생기는 변동(measurement variability)과 제품 자체 변동(product variability)의 합으로 구성된다고 한다. 그러므로 측정 결과 얼마나 오류를 범하느냐에 따라 공정의 능력이 어떻게 나타나는지 결정된다. 따라서 본 연구에서는 측정과정 중 생기는 데이터의 변동이 우리가 달성하고자 하는 공정능력지수에 어떤 영향을 미치는지 알아보하고자 한다. 마지막으로 측정시스템을 보다 현실적으로 평가하기 위해서 어떤 일반적 과정을 밟아야 하는지에 대해서도 알아본다. 한 마디로, 본 연구에서는 측정시스템 평가와 관련하여 Measurement Error Study와 GR&R 분석에 대해 백재욱과 조진남(1996), 민철희와 백재욱(1998)에서 논의한 내용중 미진한 부분에 대해 보완하여 설명하고자 한다.

2. Measurement Error Study

이 연구방법은 측정시스템 분석에 이용되는 방법중에서 가장 단순한 기법에 속한다. 구체적으로 Floyd and Laurent (1995)에 나오는 것처럼 한 명의 작업자가 하나의 계측기를 이용하여 평균이 μ 이고 표준편차가 σ (이것은 사실 후술하는 $\sigma_{repeatability}$ 라고 할 수 있음)인 품질특성치를 여러 번 잰 결과 x_1, x_2, \dots, x_n 이라고 하는 경우 MER(Measurement Error Ratio)

$$MER = 6\sigma / (USL - LSL) \quad (2.1)$$

을 구하여, 그 값이 0.1보다 작다면 측정과정에 문제가 없다고 본다. 여기서 USL (upper specification limit) 및 LSL(lower specification limit)은 각각 규격상한 및 규격하한이다.

그런데 통상적으로 표준편차 σ 는 잘 알려져 있지 않으므로 추정해야 한다. σ 를 추정하는 데 많이 이용되는 통계량은 표본표준편차 S 이므로 MER의 추정치

$$\widehat{MER} = 6S / (USL - LSL) \quad (2.2)$$

를 구하여, 이 값이 0.1보다 작도록 하고자 한다. 그런데 백재욱과 조진남(1996)에서 지적한 바와 같이 앞에서와 같은 일률적인 해석에는 문제가 있다고 본다. 왜냐하면 앞의 (2.2)는 점추정치에 불과하므로 항상 불확실성이 내재하기 때문이다. 따라서 MER에 대한 구간추정을 해야될 필요가 있다. MER에 대한 구간추정은 주어진 데이터가 정규분포를 따르는 경우 표본분산 S^2 을 이용한 다음의 통계량 $(n-1)S^2/\sigma^2$ 이 $\chi^2_{(n-1)}$ 분포를 따른다는 것을 이용한다. 따라서

$$P\left(\chi^2_{n-1, 1-\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1, \alpha/2}\right) = 1 - \alpha$$

를 이용하면 MER에 대한 $100(1-\alpha)\%$ 신뢰구간은

$$\left(\frac{6}{USL - LSL} \sqrt{\frac{(n-1)S^2}{\chi^2_{n-1, \alpha/2}}}, \frac{6}{USL - LSL} \sqrt{\frac{(n-1)S^2}{\chi^2_{n-1, 1-\alpha/2}}} \right) \quad (2.3)$$

과 같이 나온다. 그러므로 이 신뢰구간의 상한이 0.1보다 작은지 확인하여, 그렇다면 정말로 측정과정에 문제가 없다고 보아야 할 것이다.

앞의 (2.2)에서는 σ 대신 표본표준편차 S를 사용했는데 품질관리 현장에서는 사용의 편의상 범위 R을 이용한 통계량 R/d_2 를 공정의 표준편차 σ 에 대한 추정치로 많이 사용한다. 여기서 d_2 는 부분군의 크기 n의 함수로서 추정량 R/d_2 을 σ 의 불편추정량으로 만드는 수정계수이다. 따라서 이 때에는 MER에 대한 추정량으로

$$\widehat{MER} = 6(R/d_2) / (USL - LSL) \quad (2.4)$$

를 사용한다. 그런데 앞에서와 같이 식 (2.4)는 하나의 추정치에 불과하므로 불확실성

이 내재한다. 따라서 식 (2.4)를 쓰는 경우에도 MER에 대한 구간추정을 해야될 필요가 있다. 일반적으로 부분군이 합리적으로 설정되며, m 개의 부분군이 있는 경우 \bar{R} 를 m 개의 부분군에서 나온 범위들의 평균이라고 하면 (2.4)의 R대신 \bar{R} 를 이용한

$$\widehat{MER} = 6(\bar{R}/d_2)/(USL - LSL) \quad (2.5)$$

으로 MER을 추정한다. 여기서 d_2 는 앞의 설명에서와 같이 부분군의 크기에 의해 결정되는 상수이다. 그런데 Patnaik (1950)에 의하면 $\nu \bar{R}^2/(c^2 \sigma^2)$ 는 대략적으로 자유도가 ν 인 χ^2 분포를 따른다. 여기서 ν 와 c 는 부분군의 크기 및 부분군의 수에 의해 결정되는 값들이다. 한 예로, 부분군의 수 $m=1$ 이고 부분군의 크기 $n=10$ 인 경우 $\nu=7.689$ 이고 $c=3.1789$ 이다. 따라서 식 (2.5)에서와 같이 부분군이 있어서 \bar{R} 을 구할 수 있는 경우 식 (2.1)의 MER에 대한 $100(1-\alpha)\%$ 신뢰구간은 다음과 같이 대략적으로 나온다.

$$\left(\frac{6}{USL-LSL} \sqrt{\frac{\nu \bar{R}^2}{c^2 \chi_{\nu, \alpha/2}^2}}, \frac{6}{USL-LSL} \sqrt{\frac{\nu \bar{R}^2}{c^2 \chi_{\nu, 1-\alpha/2}^2}} \right) \quad (2.6)$$

그런데 식 (2.3)과 (2.6)은 주어진 데이터가 정규분포를 따른다고 하는 가정하에 의미가 있는 식이다. 따라서 일반적으로 데이터가 특정한 분포를 따르지 않는 경우에도 MER에 대한 점추정 및 구간추정을 잘 할 수 있도록 ‘적절한 통계량’을 사용해야 한다. 그래서 데이터가 정규분포를 따른다고 하는 확신이 없는 경우에는 EDA에서와 같이 저항성이 강한 추정치를 사용하여 MER을 추정할 수도 있을 것이다. 따라서 이런 경우 해당 통계량에 대한 추가적인 분석은 어떻게 실시할 수 있는지 살펴보는 것도 의미가 있을 것이다.

3. GR & R 분석

3.1 구간추정

GR&R 분석에 자주 인용되는 예는 Montgomery(1991)에서와 같이 3명의 작업자가 어느 공정으로부터 랜덤하게 뽑은 20개의 제품에 대해 품질특성치를 각각 2번씩 반복 측정하는 것이다. 이때 계측기와 관련된 분산 σ_{gauge}^2 은 동일 제품을 여러 번 측정해도 똑같이 나오지 않기 때문에 생기는 데이터의 변동 $\sigma_{repeatability}^2$ 와 작업자 간의 차이 때문에 생기는 데이터의 변동 $\sigma_{reproducibility}^2$

$$\sigma_{gauge}^2 = \sigma_{repeatability}^2 + \sigma_{reproducibility}^2 \quad (3.1)$$

으로 크게 구분된다. 그러면 GR&R 분석에서는 계측기의 정밀도(repeatability)와 재현성(reproducibility)을

$$\hat{\sigma}_{repeatability} = \overline{R}/d_2 \quad (3.2)$$

$$\hat{\sigma}_{reproducibility} = R \overline{x}_i \quad (3.3)$$

로 추정한다(백재욱과 조진남(1996) 참조). 이와 같이 GR&R 분석에서는 보통 어느 요인에 의한 자료의 변동을 하나의 점추정치로만 나타낸다. 하지만 계측결과가 항상 같을 수만은 없다. 따라서 각 요인에 의한 변동의 크기를 구간추정할 필요가 있다.

구체적인 예로 정밀도를 측정하는 $\hat{\sigma}_{repeatability} = \overline{R}/d_2$ 의 경우 데이터가 정규분포를 따른다면 기대값과 분산은 각각 $\sigma_{repeatability}$ 와 $d_3^2 \sigma_{repeatability}^2 / (op d_2^2)$ 으로 나타난다. 여기서 o와 p는 각각 작업자의 수와 제품의 수를 나타내고 d_2 와 d_3 는 Montgomery (1991)에 나와있는 바와 같이 반복추정수와 관련된 계수이다.

한편, $\sigma_{repeatability}$ 의 추정치로 표본범위 대신 표본표준편차를 이용할 수 있다. 구체적으로 i 번째 부분군의 표준편차를 S_i 라고 하고, 이들 부분군의 평균을 \overline{S} 라고 하는 경우 식 (3.2) 대신

$$\hat{\sigma}_{repeatability} = \overline{S}/c_4 \quad (3.4)$$

를 사용할 수 있을 것이다. 이 경우 추정치 (3.4)의 기대값과 분산은 각각 $\sigma_{repeatability}$ 와 $(1-c_4^2)\sigma_{repeatability}^2 / (op c_4^2)$ 으로 나타난다. 여기서 c_4 는 Montgomery(1991)에 나와 있는 바와 같이 반복추정수와 관련된 계수이다. 일반적으로 표본범위보다는 표본표준편차가 더 효율적이다. 그래서, 예를 들어, 반복추정수 $n=2$ 인 경우 $d_3^2/d_2^2=0.5718 > (1-c_4^2)/c_4^2=0.5707$ 이 되어, 범위에 의한 $\sigma_{repeatability}$ 의 추정치가 표본표준편차에 의한 $\sigma_{repeatability}$ 의 추정치 보다 더 분산이 크다. 따라서 가능하다면 $\sigma_{repeatability}$ 를 추정하는 데 식 (3.2)보다는 (3.4)를 이용하는 것이 더 좋을 것이다.

특히, 부분군의 수가 많은 경우에는 중심극한정리에 의해 다음 식

$$P\left(-Z_{\alpha/2} < \frac{\overline{S}/c_4 - \sigma_{repeatability}}{\sqrt{\frac{(1-c_4^2)\sigma_{repeatability}^2}{op c_4^2}}} < Z_{\alpha/2}\right) = 1 - \alpha \quad (3.5)$$

에 의거하여 $\sigma_{repeatability}$ 에 대한 $100(1-\alpha)\%$ 신뢰구간을 대략적으로 다음

$$\left(\frac{\bar{S}/c_4}{1 + Z_{\alpha/2} \sqrt{\frac{(1-c_4^2)}{op c_4^2}}}, \frac{\bar{S}/c_4}{1 - Z_{\alpha/2} \sqrt{\frac{(1-c_4^2)}{op c_4^2}}} \right) \quad (3.6)$$

과 같이 구할 수 있다.

다음으로 $\sigma_{reproducibility}$ 에 대해서 생각해 보자. m명의 작업자에 의한 측정치를 각각 $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_m$ 이라고 하면 [Montgomery(1991)에서는 작업자가 세 명임] 이들 측정치에 대한 범위 $R_{\bar{X}}$ 의 기대값과 표준편차는 각각

$$E(R_{\bar{X}}) = d_2 \sigma_{reproducibility} \quad (3.7)$$

$$SD(R_{\bar{X}}) = d_3 \sigma_{reproducibility} \quad (3.8)$$

라고 할 수 있다. 여기서 d_2 및 d_3 는 작업자의 수에 따른 상수이다. 그러면 범위관리도에서 관리한계선을 잡듯이 $\sigma_{reproducibility}$ 에 대한 관리한계선은 다음

$$\left(\frac{R_{\bar{X}}}{d_2} - 3 \frac{d_3 R_{\bar{X}}}{d_2}, \frac{R_{\bar{X}}}{d_2} + 3 \frac{d_3 R_{\bar{X}}}{d_2} \right) \quad (3.9)$$

과 같이 구할 수 있다. 물론 \bar{X} 들로부터 나온 범위의 분포는 한쪽으로 치우쳐져 있다. 따라서 $\sigma_{reproducibility}$ 에 대해 비교적 정확한 신뢰구간을 추정하기 위해서는 Duncan(1974)에 있는 OC 곡선을 참조해야 할 것이다.

3.2 계측기의 직선성

측정시스템의 변동을 기술하는 데 사용되는 분포는 위치와 산포의 두 가지 특성으로 묘사될 수 있다[AIAG(1995) 참조].

위치(location)

안정성(stability)

정확성(accuracy)

직선성(linearity)

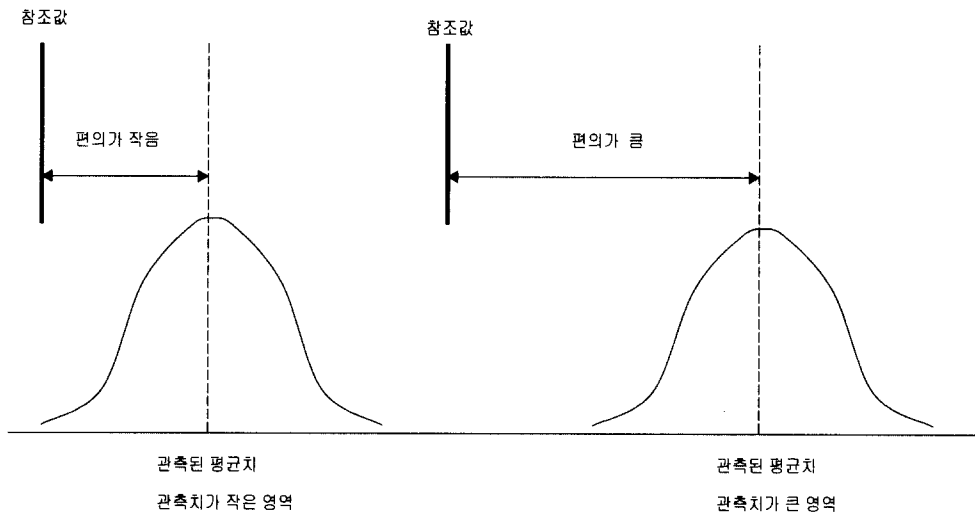
산포(spread)

정밀도(repeatability)

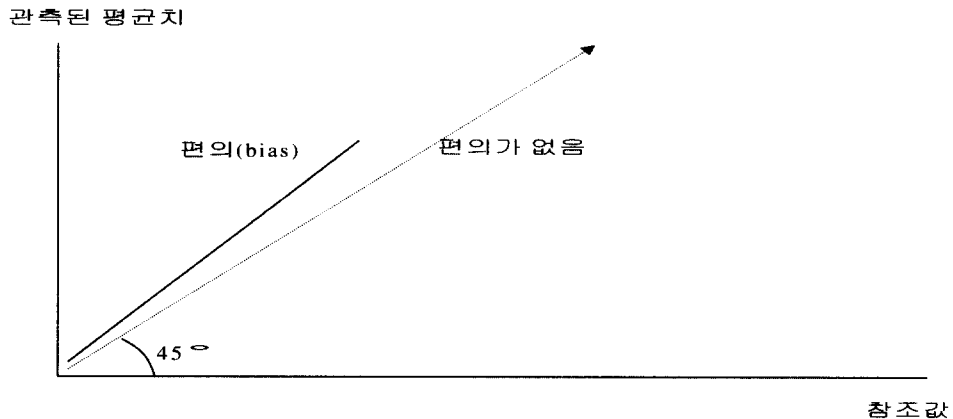
재현성(reproducibility)

위의 특성들 중에서 안정성, 정확성, 정밀도 및 재현성에 대해서는 백재욱과 조진남 (1996)에서 자세히 설명했으므로 본 절에서는 직선성에 대해서만 살펴보기로 한다.

직선성이란 <그림 1>에서 보는 바와 같이 계측기의 기대 운용범위에 걸쳐 나타나는 편의(bias)의 차이를 말한다. 여기서 편의란 정확성을 달리 표현한 것에 불과하다. 편의가 참조값(이를 참값 또는 기준값이라고도 부름)이 증가하면서 선형적으로 증가하는 경우 <그림 2>와 같은 형태를 띤다.



< 그림 1 > 직선성



< 그림 2 > 참값이 커지면서 편의가 커지는 경우의 직선성

직선성은 측정도구의 운영범위 전반에 걸쳐 사용되는 부품을 선택함으로써 결정된다. 선택된 각 부품의 편의는 부품평균으로부터 참조값을 뺀으로써 구해지며(<표 1> 참조), 부품의 참조값은 참값이 알려진 경우에는 그 값으로 하고, 그렇지 않은 경우 정밀계측기에 의하여 얻어진다. 이때 각 부품의 참조값과 편의간의 관계를 2차원상에 나타내어, 적합시킨 회귀직선의 결정계수 R^2 (이를 여기에서는 적합도라고 함)은 어느 정도인지 파악하고, 해당 직선의 경사는 어느 정도인지 파악함으로써 두 변수간 직선성이 얼마나 강한지 알 수 있다.

다음은 직선성을 살펴볼 수 있는 예제이다[AIAG(1995)]. 공장의 작업자가 측정시스템의 직선성을 확인하고자 한다. 공정의 변동에 기초하여 측정시스템의 운영범위 전반에 걸쳐 사용되는 부품이 5개 선정되었고, 각 부품을 정밀계측기로 측정한 결과 <표 1>의 상단에서와 같은 결과를 얻었으며(이것이 참조값임), 이들을 한 명의 측정자가 각각 12번씩 반복 측정한 결과 부품평균과 편의가 <표 1>의 하단에서와 같이 계산되었다. 그러면 각 부품에 대한 참조값과 편의간의 관계는 <그림 3>에서 *로 나타낼 수 있으며, 이 점들을 가장 잘 대표할 수 있는 선형회귀선은 실선으로 나타낼 수 있다.

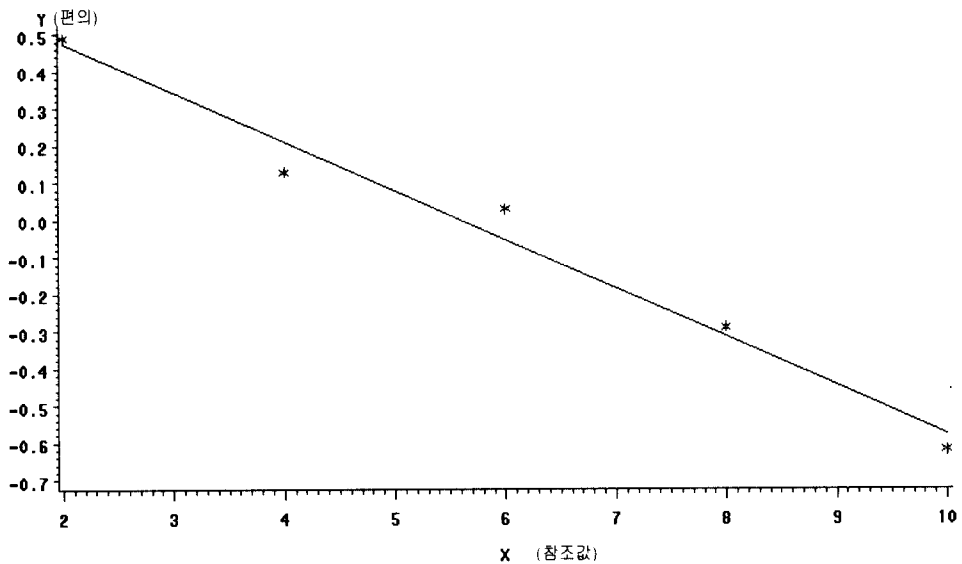
< 표 1 > 계측기 데이터

	부품 참조값	1 2.00	2 4.00	3 6.00	4 8.00	5 10.00
시 행	1	2.70	5.10	5.80	7.60	9.10
	2	2.50	3.90	5.70	7.70	9.30
	3	2.40	4.20	5.90	7.80	9.50
	4	2.50	5.00	5.90	7.70	9.30
	5	2.70	3.80	6.00	7.80	9.40
	6	2.30	3.90	6.10	7.80	9.50
	7	2.50	3.90	6.00	7.80	9.50
	8	2.40	3.90	6.10	7.70	9.50
	9	2.40	3.90	6.40	7.80	9.60
	10	2.40	4.00	6.30	7.50	9.20
	11	2.60	4.10	6.00	7.60	9.30
	12	2.40	3.80	6.10	7.70	9.40
	부품평균	2.49	4.13	6.03	7.71	9.38
	참조값(x)	2.00	4.00	6.00	8.00	10.00
	편의(y)	+0.49	+0.13	+0.03	-0.29	-0.62
	범위	0.4	1.3	0.7	0.3	0.5

이때 참조값 x 와 편의 y 간에 선형적인 관계가 있다면 두 변수간에는 다음

$$y = b + ax$$

와 같은 아주 간단한 관계가 성립한다. 여기서 기울기 a 와 절편 b 는 <표 1>의 데이터의 경우에는 최소자승법에 의해 각각 다음과 같이 나온다.



< 그림 3 > 두 변수간의 직선성

$$a = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = -0.1317 \quad (3.10)$$

$$b = \frac{\sum y}{n} - a \times \frac{\sum x}{n} = 0.7367 \quad (3.11)$$

이때 결정계수 R^2 은 다음과 같이 구해진다.

$$R^2 = \frac{[\sum xy - \frac{\sum x \sum y}{n}]^2}{[\sum x^2 - \frac{(\sum x)^2}{n}] \times [\sum y^2 - \frac{(\sum y)^2}{n}]} = 0.98 \quad (3.12)$$

따라서 주어진 데이터는 다음과 같이 요약할 수 있다.

$$\begin{aligned} \text{편의} &= 0.7367 - 0.1317 \times \text{참조값} \\ \text{적합도}(R^2) &= 0.98 \end{aligned}$$

적합도가 1에 가까우므로 두 변수간 관계를 회귀직선을 이용해도 별 무리가 없다는 것을 알 수 있고, 그래서 직선성이나 % 직선성을 다음과 같이 구한다.

$$\begin{aligned} \text{직선성} &= |\text{기울기}| \times \text{공정의 변동} \\ &= 0.1317 \times 6.00 = 0.79 \\ \% \text{직선성} &= 100 \times (\text{직선성} / \text{공정의 변동}) \\ &= 100 \times (0.79 / 6.00) \\ &= 13.17\% \end{aligned}$$

여기서 공정의 변동은 부품의 표준편차가 σ 인 경우 6σ 를 말하며, 본 예제의 경우 6.00으로 주어졌다고 가정한다. 하지만 공정의 변동 대신 때로는 공차(tolerance)를 사용하기도 한다. 그러면 일반적으로 기울기의 절대치가 작으면 작을수록 계측기의 직선성은 좋다고 하며, 역으로 기울기의 절대치가 크면 클수록 계측기의 직선성은 나쁘다고 한다. % 직선성은 직선성이 공정의 변동에서 차지하는 비율을 나타내며, 이것 또한 작으면 작을수록 좋다고 한다.

그런데 지금까지 살펴본 직선성 분석에 문제가 있을 수 있다. 첫째, 참조값으로 시료의 선정범위가 2에서 10인데도 불구하고 공정의 변동은 6으로 일정하다고 본다는 점이다. 이는 측정값이 커지면서 편의 또한 커질 수 있다는 것을 배제하므로 현실성이 떨어질 수 있다. 둘째, 참조값이 커지더라도 편의가 일정한 경우 (<그림 2>에서 45°선과 평행인 경우) 분명 계측에 문제가 있는데도 불구하고 직선성은 0이 되어 계측에 문제가 없는 것으로 판명된다는 점이다.

한편, 회귀직선이 적합한 상태가 아니라면 편의와 참조값 간에는 비선형 관계를 가질 가능성이 많고, 그러면 우리는 측정시스템이 만족스런 상태인지 추가적인 분석을 해야 할 것이다. 이 경우에는 왜 비직선성이 생기는지 다음과 같은 여러 가지 가능한 이유들 중에서 찾아야 할 것이다.

- 계측기가 운용범위 중 양 끝 부분(관측치가 작은 부분과 큰 부분)에 제대로 교정이 안되었는가?
- 최저 또는 최고 참조값에 오차가 있는가?

- 계측기가 낡지 않았는가?
 - 내부적인 계측기 설계 특성상 그런가?
- 마지막으로 계측기에 직선성이 있는지는 다음과 같은 과정을 밟아 점검한다.
- 1) 계측기의 운용범위 전반에 걸쳐 측정할 수 있도록 5개 정도의 부품을 선택한다. 이때 측정치가 어느 한쪽으로 치우치지 않도록 부품을 랜덤하게 선택해야 한다.
 - 2) 각 부품의 참조값을 구할 수 있도록 정밀계측실에서 측정하도록 하고, 이 값이 해당 계측기의 운용범위에 포함된다는 것을 확인한다.
 - 3) 계측기를 보통 사용하는 사람이 각 부품을 해당 계측기로 10번 정도 측정한다.
 - 4) 앞의 <표 1>에서와 같이 각 부품에 대해 부품평균과 부품편의를 계산한다.
 - 5) 앞의 <그림 3>에서와 같은 2차원의 그림을 그린다.
 - 6) 점들을 통과하는 적합한 회귀직선 및 적합도 R^2 을 계산하며, 직선성과 %직선성 또한 구한다.

3.3 측정오차와 관련된 공정능력지수의 불확실성

Deming(1982)에 의하면 공정의 변동(process variability)은 크게 측정과정 중에 생기는 변동(measurement variability)과 제품 자체의 변동(product variability)의 합으로 구성된다고 한다. 이를 수식으로 표현하면 다음과 같다.

$$\sigma^2 = \sigma_{gauge}^2 + \sigma_{product}^2 \quad (3-13)$$

전자에 해당하는 측정과정 중 생기는 데이터의 변동에는 그 이유가 여러 가지 있을 수 있는데, 대표적인 것으로는 측정자가 다르므로 생기는 데이터의 변동 $\sigma_{reproducibility}^2$ 과 한 측정자가 측정하더라도 생기는 데이터의 변동 $\sigma_{repeatability}^2$ 의 합으로 나타낼 수 있다. 즉, $\sigma_{gauge}^2 = \sigma_{reproducibility}^2 + \sigma_{repeatability}^2$ 이다. 그런데 요즘과 같이 복잡한 측정환경 하에서 측정과정중에 생기는 데이터의 변동은 무시할 수 없을 정도로 크다. 한 예로 참고문헌 [13]에 의하면 화학공정의 경우 측정에 의한 데이터의 변동이 공정 전체변동의 50%까지 되는 경우도 있다고 한다.

측정에 의한 데이터의 변동이 크면 우리가 목표로 하는 공정능력지수를 달성할 수 없는 경우가 있을 수 있다. 구체적인 예로 규격상한과 규격하한이 $\pm 10\text{cm}$ 이고 공정의 분포가 규격한계의 중간에 위치해 있으며 계측기에 의한 오차가 $\sigma_{gauge} = 2.5\text{cm}$, 우리가 목표로 하는 $C_p = (\text{USL} - \text{LSL}) / 6\sigma$ 는 1.5라고 하자. 이 경우에는 불행히도 $\sigma_{product}$ 가 0이라고 해도 C_p 는 기껏해야 1.33 ($= 20 / [6 \times 2.5]$) 밖에 되지 못해, 우리가 원하는 목표 $C_p = 1.5$ 는 측정과정이 향상되지 않는 한 결코 달성되지 못한다. 따라서 이 경우에는 측정능력에 관한 연구로 측정과정중 문제가 되는 것이 무엇인지 먼저 가려내어 측정관리에 만전을 기하고 난 후 공정의 능력을 개선하도록 힘써야 한다.

Floyd and Laurent(1995)에 의하면 일반 산업표준체계에서는 계측자료에 대한 관리가 제대로 이루어지고 있는지 확인하는 방법으로

측정오차율(measurement error ratio, MER)

$$MER = 6 \sigma_{gauge} / (USL - LSL) \quad (3-14)$$

을 구하여 그 값이 0.1보다 작으면 측정관리가 제대로 이루어졌다고 본다.

이제 MER을 0.1 이하라고 하고(예를 들어 0.1 및 0.05), 이들과 C_p 간의 관계를 이용하여 공정 전체의 표준편차 σ 중에서 측정과정 중에서 생기는 데이터의 표준편차 σ_{gauge} 가 얼마나 되는지 살펴본다. 일반적으로 C_p 의 값은 크면 클수록 좋다. 특히 C_p 의 값이 1.33 이상인 경우 공정의 능력은 좋다고 말한다. 그러므로 구체적인 예로 $C_p=1.33$ 인 경우(100PPM인 경우)와 $C_p=2$ 인 경우(소위 6시그마 품질수준인 경우) σ 와 σ_{gauge} 간의 비율을 살펴보면 다음 <표 2>와 같이 나온다. 이로부터 MER이 0.1 이하인 경우 C_p 가 1.33이 되기 위해서는 전체 공정의 변동을 나타내는 표준편차 중에서 측정과 관련된 표준편차의 비율은 13% 정도까지는 차지해도 되고, C_p 가 2.0이 되기 위해서는 20% 정도까지는 차지해도 된다.

< 표 2 > σ_{gauge}/σ 의 값

		MER의 값			
		0.5	0.3	0.1	0.05
C_p	0.5	0.25	0.15	0.05	0.025
	1 (2700 PPM)	0.5	0.3	0.1	0.05
	1.33 (100 PPM)	0.665	0.399	0.133	0.0665
	2 (6 시그마)	1	0.6	0.2	0.1

한편, Floyd and Laurent(1995)에 의하면 MER의 값이 0.3 보다 크면 측정관리에 문제가 있다고 보는데, 이제 MER을 0.3 이상이라고 하고(예를 들어 0.3 및 0.5), 이들과 C_p 간의 관계를 이용하여 전체 공정의 표준편차 σ 중에서 측정과정 중에서 생기는 데이터의 표준편차 σ_{gauge} 가 얼마나 되는지 살펴본다. 일반적으로 C_p 의 값은 작으면 작을수록 나쁘다. 특히 C_p 의 값이 1 이하인 경우 공정의 능력은 좋지 않다고 말한다. 그러므로 구체적인 예로 $C_p=1$ 인 경우(2700PPM인 경우)와 $C_p=0.5$ 인 경우만을 고려하여 σ 와 σ_{gauge} 간의 비율을 살펴보면 앞의 <표 2>에서와 같이 나타난다. 이

로부터 MER이 0.3 이상인 경우 $C_p=1$ 이면 전체 공정의 변동을 나타내는 표준편차 중에서 측정과 관련된 표준편차의 비율은 적어도 30% 정도까지는 차지하며, $C_p=0.5$ 이면 15% 정도까지 차지한다는 것을 알 수 있다.

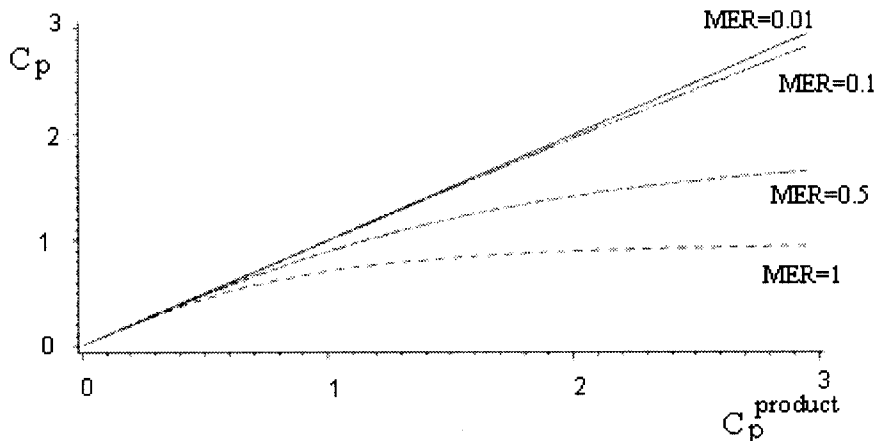
제품 자체의 변동에 대한 규격한계의 비율을 제품 자체의 능력지수

$$C_p^{product} = \frac{USL - LSL}{6\sigma_{product}} \quad (3.15)$$

라고 표시하면 C_p , $C_p^{product}$ 및 MER간의 관계는 다음의 식 (3.16)

$$C_p = \frac{1}{\sqrt{\left(\frac{1}{C_p^{product}}\right)^2 + MER^2}} \quad (3.16)$$

과 같이 나타나며, 여러 가지 MER 값에 대해 C_p 와 $C_p^{product}$ 간의 관계는 <그림 4>와 같이 나타난다. <그림 4>로부터 일정한 MER의 값에 대응하는 C_p 와 $C_p^{product}$ 간의 관계는 비선형이라는 것을 알 수 있으며, 일반적으로 공정의 능력을 나타내는 지수 C_p 는 1.33 이상이 되기를 원하는데 만일 제품 고유의 능력을 나타내는 $C_p^{product}$ 가 1.8 이하라고 하면 이 때에는 측정에 따른 데이터 변동을 나타내는 MER이 0.5 이하이어야 한다는 것을 알 수 있다.



< 그림 4 > C_p , $C_p^{product}$ 및 MER간의 관계

3.4 측정시스템 평가계획

3.4.1 측정되는 부품의 선택과 부품의 수 및 부품별 반복측정치의 수

우선, 측정에 선택되는 부품들은 규격에 꼭 맞지 않더라도 실제로 제조되어 다음 공정으로 넘어간다면 이들도 측정 대상에 포함해야 한다. 즉, 제조과정에서 생길 수 있는 모든 부품은 측정의 대상이 되어야 한다. 물론 많은 경우 현재 생산중인 부품들로부터 몇 개의 부품만을 뽑는다면 이들의 품질특성치는 대부분 규격한계선에 들어갈 것이다. 특히 공정관리가 잘 된 경우 Montgomery(1991)에서와 같이 20개의 부품들을 뽑는다면 이들의 품질특성치는 대부분 규격한계선 안에 들어갈 것이다.

Montgomery(1991) 및 Montgomery and Runger(1993a)에서는 20개의 제품을 랜덤하게 뽑고 각 부품별로 세 종류의 계측기(또는 세명의 관측자)를 이용하여 2번씩 반복측정한다. 그러나 전체 실험회수만 예산이나 시간적인 이유로 인해 정해진 경우 앞에서와는 달리 부품의 수나 계측기의 수는 적게 하더라도 반복측정치의 수는 많게 한 다거나, 반대로 부품의 수나 계측기의 수는 많게 하더라도 반복측정치의 수를 적게 할 수 있다. 일반적으로 부품의 수가 많고 반복측정치의 수나 계측기의 수가 적다면 실제 제조공정의 상황이 시간에 따라 변화하는 경우 이를 잘 파악할 수 있다. 특히 장시간에 걸쳐 제품이 추출되는 경우 시간의 변화에 따른 공정의 변화를 잘 읽을 수 있다. 그러나 공정에 변화가 없고 계측기 관리에 문제가 있다고 판단되면 측정되는 부품의 수를 가능한 한 적게 하며, 반대로 계측기의 수를 늘리고 반복측정치의 수를 증가시켜야 할 것이다. 즉, 전체 공정에 대한 이상을 잘 파악하기 위해서는 많은 부품을 검사해야 하나, 계측기와 관련된 신뢰성을 높이기 위해서는 일정한 수의 부품에 대해 보다 많은 계측기로 반복측정을 많이 해야 한다. 특히 계측기에 따라 측정치의 변동이 심하면 가능한 한 공정에 이용되는 계측기를 모두 조사해야 하며, 어떤 특정한 계측기의 정밀도에 문제가 있다고 판명되면 그 계측기를 이용하여 가능한 한 많이 반복측정해야 한다.

한편, 품질특성치의 측정시 오차는 항상 생기는데, 그 오차(분산이 하나의 척도가 될 수 있음)는 재고자 하는 특성치의 수준에 따라 다를 수 있다. 예를 들어 구멍을 뚫는 공정의 경우 구멍의 직경이 크면 클수록 그 측정오차는 커질 수 있는데, 이 경우 보다 많은 제품에 대해 검사를 한다면 측정치의 분산이 임의의 고정된 상수가 아니라는 것을 발견할 수 있을 것이다. 이런 경우에는 주어진 자료를 R 관리도로 나타내면 그 현상을 쉽게 파악할 수 있다. 이때 품질특성치의 수준이 같은 것들은 똑같은 그룹으로 묶어 R 관리도를 작성하면 더욱 효율적으로 그 현상을 파악할 수 있다.

마지막으로 측정시스템 평가를 하다보면 항상 측정치의 산포에 영향을 끼치는 요인에 대한 추정치가 너무 낙관적일 수 있다. 구체적인 예로, 어느 부품의 제조과정중에 그 부품을 어느 위치에 고정시켜 작업하는 공정이 있다고 하자. 이때 고정된 몇 개의 부품에 대해서만 그 두께를 여러 번 잰다면 측정오차에 부품을 고정시키는 과정은 제대로 포함되지 않아, 결국 측정오차는 과소평가되게 된다. 그렇다면 측정결과에 영향을 미치는 중요한 요인이 빠질 수 있고 이로 인해 측정시 산포에 대한 추정치는 너무 낙관적일 수 있다.

측정되는 부품의 수, 계측기의 수 및 부품별 반복 관측치의 수를 구체적으로 결정하는 판단기준으로 모수의 신뢰구간 길이를 고려할 수 있다. 특히 Montgomery and Runger(1993b), Burdick and Larsen(1997) 및 Borror, Montgomery and Runger(1997)에서는 실험계획법적인 접근방법을 통하여 각 변동요인에 대한 신뢰구간을 구하는데, 이를 이용하면 부품의 수나 반복측정 횟수를 적절히 정할 수 있을 것이다.

3.4.2 완전한 반복측정이 어려운 이유와 그러한 경우의 대책

측정시스템 평가에서 가장 중요한 것은 동일 부품 또는 제품에 대해 어떻게 반복측정을 할 것인지 결정하는 것이다. 왜냐하면 계획이 잘못된 상태에서 반복측정이 이루어진다면 측정오차는 대부분 과소평가되어지기 때문이다. 그러면 우선 다음의 몇 가지 예를 들어 세부적인 반복측정 계획이 얼마나 중요한지 알아본다.

첫 번째 예로 표면의 평평함을 측정하는 경우를 생각해본다. 어느 공정에서 흘러나오는 부품이 어느 특정한 실험대에 부착되어 똑같은 부위에 대해 평평함이 측정된다고 하자. 이러한 측정은 보통 아주 단기간에 이루어지므로 그 사이에 주변의 환경변화라든가 전압이 오르락 내리락 하는 일은 거의 없을 것이다. 따라서 이와 같이 주변 환경이 똑같은 경우에 관측치들은 거의 반복적이라고 할 수 있다. 그러나 현재의 측정과정을 장기적인 면에서 전체 공정의 일부분으로 볼 때 앞에서와 같은 이상적인 상황은 벌어지지 않을 것이다. 아마도 작업자마다 계측기의 사용방법이 조금씩 다를 수 있고, 심지어 어떤 작업자는 계측기의 사용방법에 익숙치 않은 경우도 있을 수 있으며, 또는 여러 대의 계측기중 교정이 제대로 되어 있지 않은 것이 있어서 계측기마다 조금씩 성능이 다를 수도 있고, 어떤 계측기는 좀 낡아서 제대로 작동하지 않을 수도 있다. 또는 부품의 고정되는 부분이 매번 달라서 영땡한 곳의 평평함이 측정되어질 수도 있다. 이와 같이 측정과정을 전체 공정의 일부분으로 볼 때에는, 너무나 많은 요인으로 인해 이상적인 반복측정상황은 벌어지기는 힘들 것이다.

완전한 반복측정이 이루어지기 어려운 또다른 예로 화학공정을 생각해보자. 화학공정에서 새로운 표본을 준비하기 위해서는 물과 화학약품 등을 일정한 비율로 섞어야 하며, 적절한 온도를 가해주어야 하고, 필요하다면 잘 저어주어야 하며, 일정한 시간이 지나면 다른 물질과 혼합하는 등 여러 가지 작업을 해야한다. 이러한 작업들은 물론 각기 다른 작업자나 다른 작업장에서 이루어질 수 있다. 따라서 똑같은 배치(batch)에서 추출되지 않는 한 표본과 표본간은 여러 요인으로 인하여 서로 다를 것이다. 사실 똑같은 배치에서 추출된 표본이라고 할지라도 우리가 모르는 어떤 요인에 의해 서로 조금씩은 다를 수 있다.

앞의 두 예로부터 아무리 똑같은 상황 하에서 공정이 이루어졌다고 하더라도 표본과 표본간에는 여러 가지 다른 요인에 의하여 서로 다른 결과를 보일 수 있다. 그런데 계측기는 제품의 품질을 유지시키고 향상시키는데 사용되므로, 계측기에 의한 자료의 변동도 전체 공정의 일부분으로 다루어져야 한다. 그러므로 이상적인 상황이 아닌 현실적인 상황 하에서 측정시스템을 평가해야 한다. 즉, 측정결과에 영향을 미칠 수 있는 여러 요인들을 측정시스템 평가의 대상에 포함시켜야 한다.

측정시스템 평가시 장기간의 변동을 분석에 포함시키기가 어려운 경우에는 단기간의 변동만을 고려한 분석에 만족할 수 밖에 없을 것이다. 그러나 장기적인 요인이 측정시스템 평가에 중요하다면 연구목적상 장기적인 요인을 가속화시킬 필요가 있다. 예를 들어 단기적으로는 한 종류의 계측기가 한 장소에서만 사용되나, 장기적으로는 여러 개의 서로 다른 계측기가 여러 장소에서 사용된다면 이들 각각의 상황에서 단기간에 실험을 실시하여 장기적인 효과를 파악해볼 수 있다. 마찬가지로 화학공정에서는 여러 가지 다른 배치나 시약에서 나온 결과를 이용하여 실제로 장기간의 제조과정이나 측정과정에서 접하게 되는 자료를 얻을 수 있다.

3.4.3 측정시스템 평가의 일반적인 과정 및 주의점

그러면 이제 측정시스템 평가를 일반적으로 어떻게 진행시키는지 알아보자. 측정시스템 평가에서 가장 먼저 해야할 일은 측정결과에 영향을 미칠 수 있는 여러 요인들에 대해 가능한 한 많이 개괄적으로 분석하는 것이다. 이때 장기적인 요인이 측정결과에 어떤 영향을 미치는지 알아보기 위해서는 앞에서 언급한 가속화 방법을 사용할 수 있다. 즉, 어느 한 부품에 대해 여러 날에 걸쳐 여러 곳에 있는 서로 다른 설비를 이용하여 품질특성치를 측정한다. 또한 다른 부품에 대해서도 여러 날에 걸쳐서 품질특성치를 측정하게 되는데, 이때 사용되는 계측설비는 첫 번째 부품을 측정한 계측설비와 다를 수 있다. 이로부터 나온 자료를 이제 분석하게 되는데, 이 단계에서 실시되는 분석은 측정결과를 처음으로 분석하는 것이므로 사전분석에 해당된다고 할 수 있다. 따라서 특정한 분석방법에 엄매일 필요가 없다.

혹자는 앞에서와 같이 계측기를 여러 날에 걸쳐서 여러 곳에 있는 제품에 대해 마구 사용하는 것은 너무 무계획적이라고 생각할 수 있다. 그러나 이러한 것이 실제 공정관리 현장에서 많이 발생할 수 있는 상황일 것이다. 따라서 이 방법을 이용하여 측정과정에서 변동을 일으킬 수 있는 모든 요인들에 대해 사전정보를 충분히 얻고, 또한 최악의 경우에 측정결과가 어떻게 나올 것인지 알아본다. 물론 이로부터 나온 결과가 만족스럽다면 측정과정이 만족스럽다는 것을 의미한다.

하지만 앞에서와 같은 분석 결과가 만족스럽지 못할 수 있다. 이 때에는 사전실험 자체가 너무 무계획적이었기 때문에 작업자, 계측기 및 주위 환경 등 어느 특정한 요인이 측정결과에 어떻게 영향을 미쳤는지 알 수 없다. 그러므로 이 때에는 보다 철저한 실험계획법적인 접근방법으로 측정결과에 변동을 일으키는 중요한 요인이 무엇인지 찾아내야 한다. 따라서 두 번째 단계로 측정시스템 평가를 위해 취할 수 있는 방법은 측정시스템에 영향을 주는 모든 요인들을 측정시스템 평가에 체계적으로 포함시키는 것이다. 특히 이 단계에서는 실험계획법에서 이용되는 아주 기본적인 원리들이 매우 유용하게 이용될 수 있다. 한 예로 요인배치법을 이용하면 여러 가지 요인들의 효과를 동시에 파악할 수 있다. 따라서 측정시스템을 평가하기 위해서는 분산분석 기법이 절실하게 필요하다.

어느 실험이나 마찬가지로이지만 측정시스템 평가에서도 비용이 많이 든다. 그렇다면 전체 실험에 드는 비용을 어떻게 사용해야 되는지가 문제일 것이다. Montgomery and

Runger(1993a)에 의하면 일반적으로 사전실험에서는 많아야 전체 예산의 25% 정도를 할당하는 것이 좋다고 한다. 사전실험에 해당하므로 측정결과에 영향을 미치는 요인은 가능하면 많이 살펴보고, 각 요인당 2 내지 3 개 정도의 수준만을 고려하도록 한다. 예를 들어, 두명의 작업자가 두 개의 서로 다른 계측기를 이용하여 품질특성치를 두 번 정도 측정한다. 이렇게 실험하고 그 결과를 분석하면 측정결과에 영향을 미치는 중요한 요인들이 어떤 것들인지 쉽게 파악할 수 있으며, 그 다음 단계로는 중요하다고 판단되는 요인들에 대해서만 실험을 다시 실시하는 것이다. 그러나 이 때에는 해당 요인에 대해 보다 많은 수준에서 실험을 실시해야 할 것이며, 그로부터 해당 요인이 측정 결과에 구체적으로 어떤 영향을 미치는지 파악해야 한다. 측정시스템 평가에서 고려해야 될 요인이 아주 많은 경우에는 부분요인실험을 통하여 측정시 산포에 영향을 주는 중요한 요인들을 먼저 걸러낼 수 있다. 그런데 이와 같은 전략은 실험계획법에서 보통 많이 쓰이는 기법이다.

다음으로 측정과정시 주의해야 할 점에 대해서 얘기하기로 한다. 측정을 하다보면 종종 측정값이 어떤 기준치와 상이하다는 것을 발견하게 된다. 이때 그 원인이 확실히 밝혀지는 경우에는 그 요인을 제거하여 측정값이 해당공정을 제대로 반영하도록 해야 할 것이다. 그러나 종종 그 원인이 확실히 밝혀지지 않았는데도 불구하고, 그때의 편차를 구하여 다음부터 나오는 측정값에 대해서는 앞의 편차를 계속적으로 수정해서 관측값을 읽는 경우가 있다. 이때 만일 규명되지 않은 이상요인이 계속적으로 지속된다면 측정관리에 별 문제가 없겠으나, 그렇지 않은 경우 이 방법은 측정결과에 불안정을 초래한다. 이러한 것은 과도한 측정관리의 한 예라고 할 수 있으며, 측정장비가 정확하게 교정이 안된다거나 또는 교정이 쉽지 않은데도 불구하고 정기적으로 자주 교정을 하는 경우 똑같은 문제가 발생하게 된다. 이러한 경우에는 측정공정을 대표할 만한 부품을 뽑아 품질특성치를 재어, 이 값들을 관리도(중심선은 표준에 해당하는 값임)에 그려 넣어 관측값들이 적절히 설정된 관리한계선내에 있을 때에는 교정을 하지 않고, 이를 벗어날 때에만 교정을 하도록 해야 할 것이다. 이와 같은 교정방법은 교정하는데 시간이 많이 걸린다거나 또는 비용이 많이 드는 경우 특히 경제적이며, 이렇게 할 때에 측정과정 또한 원활히 관리될 수 있다.

4. 결론

본 연구에서는 데이터의 신뢰성 확보를 위해 Measurement Error Study 및 GR&R 분석에 대해 백재욱과 조진남(1996) 및 민철희와 백재욱(1998)에서 논의한 내용중 미진한 부분에 대해 보완 설명했다. Measurement Error Study에서는 MER에 대한 구간추정의 필요성과 해당 모수에 대한 간단한 구간 추정방법에 대해 살펴보고, GR&R 분석에서 $\sigma_{repeatability}$ 와 $\sigma_{reproducibility}$ 에 대한 대략적인 구간추정방법에 대해서도 살펴보았다. 계측기의 직선성이 무엇이며, 어떤 방법으로 직선성을 평가하는지 살

펴보았으며, 측정오차와 공정능력지수간의 관계에 대해서도 고찰하였다. 마지막으로 측정시스템을 보다 현실적으로 평가하기 위해 어떤 일반적인 과정을 밟아야 하는지에 대해서도 살펴보았다.

측정시스템 평가는 요즘과 같이 계측이 초정밀, 초소형화 되면서 그 중요성이 더해가고 있다. 특히, 미국의 3대 자동차 회사들이 부품 공급업체들로 하여금 계측 및 검사시스템에 내재한 변동을 분석하기 위해 QS 9000에서 GR&R 분석을 요구하면서 그 필요성이 국외는 물론 국내의 자동차 관련 부품업체에도 점점더 확산되고 있는 실정이다. 이에 국내에서도 계측시스템의 올바른 평가를 할 수 있도록 적극 노력해야 할 것이다.

참고문헌

- [1] 민철희, 백재욱(1998), "통계적 공정관리 추진시 측정시스템 평가의 실시방법에 관한 연구," 응용통계연구 제 11권 1호, pp. 13-28.
- [2] 백재욱, 조진남(1996), "계측기 능력분석과 실험계획법," 품질경영학회지, 제 24권, 제 3호, pp. 145-159.
- [3] AIAG(1995), *Measurement System Analysis Reference Manual*.
- [4] Borrer, C.M., D.C. Montgomery and Runger, G.G.(1997), "Confidence intervals for variance components from gauge capability studies," *Quality and Reliability Engineering International*, Vol. 13, pp. 361-369.
- [5] Burdick, R.K. and Larsen, G.A.(1997), "Confidence Intervals on Measures of Variability in R&R Studies," *Journal of Quality Technology*, Vol. 29, No. 3, pp. 261-273.
- [6] Deming, W.E.(1982), *Quality, Productivity and Competitive Position*, MIT Center for Advanced Engineering Study, Cambridge, MA.
- [7] Duncan, A.J.(1974), *Quality Control and Industrial Statistics*, 4th ed., Homewood, Ill., Richard Irwin, Inc.
- [8] Floyd, D.A. and Laurent, C.J.(1995), "Gauging: An underestimated consideration in the application of statistical process control," *Quality Engineering*, Vol. 8, No. 1, pp. 13-29.
- [9] Montgomery, D.C.(1991), *Introduction to Statistical Quality Control*, 2nd ed., John Wiley & Sons, New York.
- [10] Montgomery, D.C. and Runger, G.C.(1993a), "Gauge Capability Analysis and Designed Experiments. Part I," *Quality Engineering*, Vol. 6, No. 1, pp. 115-135.
- [11] Montgomery, D.C. and Runger, G.C.(1993b), "Gauge Capability Analysis and

- Designed Experiments. Part II," *Quality Engineering*, Vol. 6, No. 2, pp. 289-305.
- [12] Patnaik, P.B.(1950), "The Use of Mean Range as an Estimator of Variance in Statistical Tests," *Biometrika*, Vol. 37, pp. 78-87.
- [13] *Quality Assurance for the Chemical Process Industries*, American Society for Quality Control, 1987.