공업통계분야에서 동등성 검정 및 그 응용

백재욱†

한국방송통신대학교 정보통계학과

Equivalence testing and its applications in industry

Jaiwook Baik†

Department of Information Statistics, Korea National Open University

Key Words : industrial statistics; equivalence testing; hypothesis testing; TOST.

Abstract

As more and more data are collected one may ask whether the data collected within a short period of time are same. In this case traditional hypothesis testing of $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$ is used to determine whether the data are same when there is no knowledge about equivalence testing. However, this type of hypothesis testing has the undesirable property of penalizing higher precision. So TOST is to be performed in the event of equivalence testing. In this study equivalence testing is introduced where one can find the applications in industry. Traditional two sample t testing is to be compared with the equivalent testing and the procedure to perform the equivalence testing is to be presented along with an example. Finally equivalence testing in terms of the other parameters such as variance, proportion or failure rate is to be sought.

1. 서 론

데이터에 의한 통계적 품질관리가 자동화되면서 많은 양의 데이터가 수집된다. 이때 특히 작은 시차를 두고 수집되는 데이터에 대해서는 이들이 서로 동질적이지 않을까 하는 의문을 가지게 된다. 두 모수간 동질성 또는 동등성(equivalence)을 결정하는 것은 제품과 프로세스의 품질을 평가하는 공업통계분야에서 특히 관심이 있다. 한 예로 접착제를 생산하는 회사에서 접착력이 접착제 생산 후 1시간이 지나도 처음 생산 때와 같은지 알고 싶을 수 있다. 만일 두 시점에서의 접착력이 동일하다는 증거가 있다면 1시간 후에 측정을 하지 않아도 되므로 그 만큼 자원을 절약할 수 있다.

동등성 검정(equivalence testing)은 엔지니어에게 가르치는 공업통계 교과과정에는 통상 포함되어 있지 않으며, 공학 관련 문헌에서도 이 방법은 사실 언급조차 되고 있지 않은 실정이다. 그 결과 품질관리분야의 실무자들은 통계학개론에서 배운 두 표본 t검정을 임시변통적으로 사용한다. 예를 들어 평균에 대한 동등성 검정인데도 불구하고 통상적인 가설 H_0: $\mu_1 = \mu_2$, H_1: $\mu_1 \neq \mu_2$을 세워, H_0가 기각되지 않으면 두 모평균 간에는 유의한 차가 있다고 할 만한 충분한 증거가 없다고 해석한다.

동등성 검정은 의약분야에서 값비싼 원래의 약을 대체할 수 있는 값싼 대체약을 개발하여 이들 약이 서로 동등한지 검정하는 데 많이 이용되었다(Metzler, 1974; Westlake, 1976; Schuirmann, 1981; Westlake, 1981; Schuirmann, 1987; Hauck and Anderson, 1992). 하지만 이런 동등성 검정은 Stein and Doganaksoy (1999-2000)와 Lamprecht (2005)에서와 같이 공업통계분야는 물론 심리학(Rogers et al. (1993)) 및 화학(Roy (1997))분야에서도 적용되고 있다.

본 논문에서는 공업통계분야에서 발생할 수 있는 동

† 교신저자 jbaik@knou.ac.kr

※ 이 논문은 2007년 한국방송통신대학교 학술연구비지원을 받아 작성된 것임

등성 검정을 소개하고, 이의 올바른 검정절차를 살펴본다. 구체적으로 2절에서는 동등성 검정을 적용해야 하는데도 통상적인 가설검정을 적용하는 경우 어떤 현상이 나타나는지 살펴보고, 3절에서는 동등성 검정을 적절히 실행하는 방법에 대해 알아본다. 다음으로 4절에서는 구체적인 예를 들어 통상적인 가설검정에 비해 동등성 검정이 왜 필요한지 설명한다. 마지막으로 5절에서는 공업통계분야에서 흔히 발생할 수 있는 평균이외에 다른 모수에 대한 동등성 검정은 어떻게 실시할 수 있는지 알아본다.

2. 통상적인 가설검정과 동등성 검정

품질관리에서 품질특성치에 영향을 미치는 인자의 효과를 파악하는 것은 중요하다. 예를 들어, 온도의 수준간 차이가 재료의 강도에 영향을 미치는지 보고자 하는 경우 검정하려는 가설은 다음과 같다.

H_0: $\mu_1 = \mu_2$ (혹은 $\mu_1 - \mu_2 = 0$)

H_1: $\mu_1 \neq \mu_2$ (혹은 $\mu_1 - \mu_2 \neq 0$)

그리고 이 때에는 두 표본 t검정을 적용하므로 검정통계량은 다음과 같다.

$$T = \frac{\overline{Y_1} - \overline{Y_2}}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t(n_1 + n_2 - 2),$$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$

구체적인 예를 들어 30°C와 20°C에서 실험한 결과가 <표 1>과 같다고 하자.

<표 1> 가상적인 실험결과

	표본1(30°C)	표본2(20°C)
샘플수	n_1=100	n_2=100
표본평균	$\overline{Y_1}$=51.2	$\overline{Y_2}$=51.3
표본표준편차	S_1=0.47	S_2=0.59

이 때에는 α=0.05인 경우 $|t|$=1.326 < $t_{0.975}(198)$ =1.972이므로 H_1을 채택할 수 없다. 이런 가설검정 결과는 100(1-α)% 신뢰구간

$$(\overline{Y_1} - \overline{Y_2}) \pm t_{1-\alpha/2}(n_1 + n_2 - 2)\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

을 구해서도 확인할 수 있다. 앞에서 α=0.05인 경우 $(\mu_1 - \mu_2)$에 대한 95% 신뢰구간은 (-0.249, 0.049)로 0을 포함하므로 대립가설을 채택할 수 없게 된다.

하지만 지금까지 공정조건인 온도를 30°C 적용하여 공정이 이루어졌는데, 재료공학자의 의견에 의하면 온도를 20°C 적용해도(비용이 그만큼 적게 드는 것임) 재료의 강도는 똑같다고 주장한다고 하자. 이때 동등성 검정방법을 모르는 상태에서는 다음 절에서 기술하는 동등성 검정 대신 앞의 통상적인 가설검정을 적용할 것이다. 하지만 앞의 통상적인 가설검정방법은 두 온도간 차이가 없음을 주장하는 동등성 검정에는 적절하지 않다. 왜냐하면 통상적인 가설검정방법은 높은 정밀도에 벌점을 부과하는 비합리적인 속성을 지니고 있기 때문이다. 다시 말해, 일정한 표본의 크기에 대해 동등하다고 선언(통상적 가설검정에서 대립가설 채택)된 표본평균의 차이가 표본의 크기가 큰 경우에는 비동등성이 선언(통상적인 가설검정에서 대립가설 채택)되기 때문이다. 구체적으로 앞의 예에서 두 표본의 크기가 이전의 10배로 각각 1000개씩인 경우 앞에서는 동등성이 선언된 것이 비동등성이 선언된다(<표 2> 참조).

하지만 이런 현상에 대해 의아해할 필요가 없다. 왜냐하면 통상적인 가설검정은 높은 정밀도를 가지면(예를 들어 n_1과 n_2가 커지든 또는 S_p^2이 작아지면) 대립가설을 채택하도록 만들어져있기 때문이다. 따라서 통상적인 가설검정은 비동등성(difference)을 보고자 할 때 적절하며, 동등성(equivalence)을 보고자 할 때에는 다음의 동등성 검정을 실시해야 한다.

<표 2> n_1과 n_2의 크기에 따른 가설검정과 신뢰구간

n_1 & n_2	t	신뢰하한	신뢰상한	결론
100 & 100	-1.326	-0.249	0.049	H_1 채택 불가
1000 & 1000	-4.192	-0.147	-0.053	H_1 채택

3. TOST

앞의 예와 같은 동등성 검정에서는 가설이 다음과 같아야 한다.

H_0: $\mu_1 \neq \mu_2$ vs H_1: $\mu_1 = \mu_2$

하지만 대립가설은 실용적인 차원에서 μ_1과 μ_2가 아주 가깝다(아주 작은 θ>0에 대해 $|\mu_1 - \mu_2| < \theta$)는 것으로 대체할 수 있으므로 앞의 가설은 다음과 같이 쓸 수 있다.

$H_0 : |\mu_1 - \mu_2| \geq \theta$ vs $H_1 : |\mu_1 - \mu_2| < \theta \Leftrightarrow$

$H_0 : \mu_1 - \mu_2 \leq -\theta$ 또는 $\mu_1 - \mu_2 \geq \theta$ vs

$H_1 : -\theta < \mu_1 - \mu_2 < \theta \Leftrightarrow$

$H_{01} : \mu_1 - \mu_2 \leq -\theta$ vs $H_{11} : \mu_1 - \mu_2 > -\theta$ &

$H_{02} : \mu_1 - \mu_2 \geq \theta$ vs $H_{12} : \mu_1 - \mu_2 < \theta$

이와 같은 성질로 인해 동등성 검정은 두 개의 한쪽검정(TOST: two one-sided tests)이라고 한다.

두 표본이 공통분산을 가진 분포로부터 나온 경우 앞의 두 가설에 대한 검정 과정은 다음과 같다. 우선 H_{01}은

$$\frac{(\overline{Y_1} - \overline{Y_2}) + \theta}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} > t_{1-\alpha}(n_1 + n_2 - 2),$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \text{ 또는}$$

$$(\overline{Y_1} - \overline{Y_2}) - t_{1-\alpha}(n_1 + n_2 - 2)\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}$$

$> -\theta$ 인 경우 기각한다.

다음으로 H_{02}는

$$\frac{(\overline{Y_1} - \overline{Y_2}) - \theta}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} < -t_{1-\alpha}(n_1 + n_2 - 2) \text{ 또는}$$

$$(\overline{Y_1} - \overline{Y_2}) + t_{1-\alpha}(n_1 + n_2 - 2)\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}$$

$< \theta$ 인 경우 기각한다. 따라서 두 가설검정에서 귀무가설이 모두 기각될 때 두 모수간 동등성이 선언된다. Schuirmann (1987)에 의하면, 앞의 두 가설이 각각 유의수준 α에서 검정될 때 동등성 검정에서 귀무가설이 잘못 기각될 확률은 α를 넘지 않는다.

한편, Westlake(1981)에 의하면 $\mu_1 - \mu_2$에 대한 통상적인 100(1-2α)% 신뢰구간이 동등성 구간인 (-θ, θ)에 포함되기만 하면 동등성을 선언할 수 있다. 따라서 H_{01}과 H_{02}는 만일 100(1-2α)% 신뢰구간인

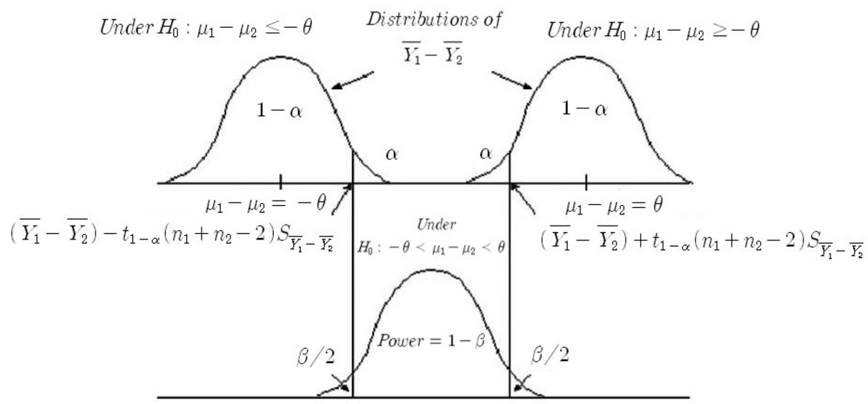
$$(\overline{Y_1} - \overline{Y_2}) \pm t_{1-\alpha}(n_1 + n_2 - 2)\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}$$

이 -θ 와 θ 사이에 완전히 포함되면 유의수준 α에서

-θ $< \mu_1 - \mu_2$이고 $\mu_1 - \mu_2 < \theta$, 곧

-θ $< \mu_1 - \mu_2 < \theta$ 이므로 동등성을 선언할 수 있다.

TOST는 검정결과 평균이 동등한지 보는 것이므로 가설검정의 원리와 일치한다. <그림 1>은 1종의 오류를 범할 확률(동등하지 않은데 동등하다고 잘못 판단할 확률) α와 검정력은 1-β을 보여준다(Rogers et al. (1993) 참조).



<그림 1> TOST를 이용한 동등성 검정의 경우 1종과 2종의 오류

한편, <그림 2>에서 보는 바와 같이 통상적인 가설검정 결과와 동등성 가설검정 결과는 같은 결론을 낼 수도 있지만 서로 다른 결론을 낼 수도 있다(Allen and Seaman(2006) 참조). 구체적으로 <그림 2>에서 첫 번째 칸(Different And Not Equivalent)과 두 번째 칸(Not Different And Equivalent)의 경우는 통상적 가설검정이나 동등성 가설검정 결과 의미상 똑같은 결론에 다다르지만 세 번째 칸(Different And Equivalent)과 네 번째 칸(Not Different And Not Equivalent)의 경우는 서로 상반된 결론에 다다르게 된다. 세 번째 칸의 경우 통상적인 가설검정에서는 두 개의 신뢰구간이 모두 0을 포함하지 않아 두 모평균에 유의한 차가 있다(Different)고 판단하겠지만 동등성검정에서는 그 신뢰구간이 정해진 수준(θ) 내에 있으므로 두 평균간 동등하다(Equivalent)고 판단한다.

4. 동등성 검정 예

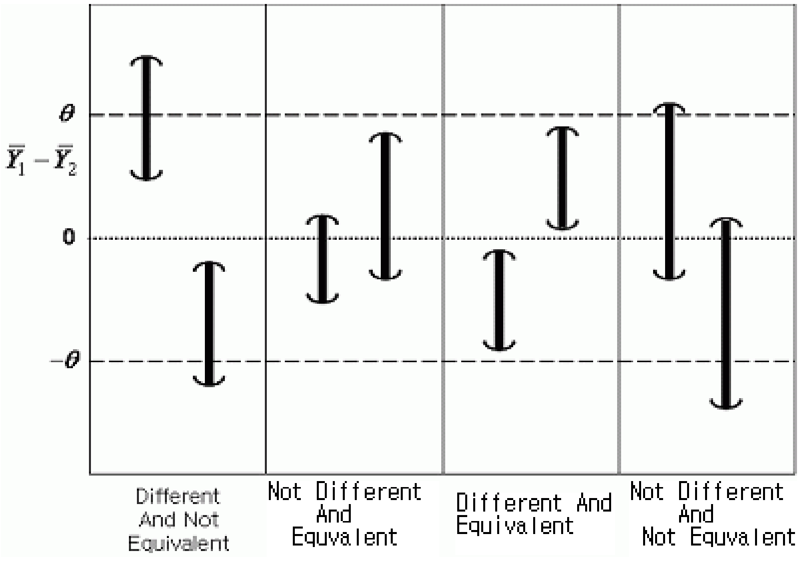
앞의 두 온도에서 처리된 재료의 강도가(또는 서론의 예제에서 두 시점에서 측정한 접착제의 접착력이) 동등한지 알고 싶다고 하자(동등성 검정이 적용될 수 있는 또 다른 예로 Stein and Doganaksoy(1999-2000)의 Case Study를 참조하라). 구체적으로 평균두께가 0.25mil 만큼 차이가 나지 않는다면(즉, $-0.25 < \mu_1 - \mu_2 < 0.25$) 동등하다고 여기고 싶은 경우 가설은 다음과 같다.

H_0: $\mu_1 - \mu_2 \leq -0.25$ 또는 $\mu_1 - \mu_2 \geq 0.25$ vs $H_1: -0.25 < \mu_1 - \mu_2 < 0.25$

이제 신뢰구간을 이용해 동등성 검정을 하고자 하는 경우 $\mu_1 - \mu_2$에 대한 90% 신뢰구간인 (-0.225, 0.025)는 동등성 검정에서의 대립가설을 만족하므로 (즉, 이 구간이 -0.25와 0.25 사이에 있으므로) 귀무가설은 5% 유의수준에서 기각되고 두 평균은 동등하다고 판단한다. 따라서 주어진 표본에 근거하여 낮은 온도에서 가공해도 재료의 강도에는 큰 차이가 없다고 결론을 내린다.

그러나 이와 같은 결론은 통상적인 가설검정을 이용해도 나온다. 즉, α=0.05인 경우 $(\mu_1 - \mu_2)$에 대한 통상적인 95% 신뢰구간은 (-0.249, 0.049)로 0을 포함하므로 통상적인 가설검정에서는 대립가설(H_1: $\mu_1 \neq \mu_2$)을 채택할 수 없다.

이제 동등성 검정의 우위성을 보이기 위해 표본의 크기가 아주 커졌는데도(각각 100개가 아닌 1,000개씩인 경우) 똑같은 표본통계량($\overline{Y_1}$=51.2, S_1=0.47, $\overline{Y_2}$=51.3, S_2=0.59)이 구해졌다고 하자. 이때 동등성 검정에서 90% 신뢰구간은 -0.139$\leq \mu_1 - \mu_2 \leq$-0.061이 되어 -0.25와 0.25 사이에 포함될 뿐만 아니라 앞의 신뢰구간보다 그 폭이 더욱 좁아져 동등성을 지지하는 확실한 증거가 된다. 그러나 통상적인 가설검정방법에 의하면 구간의 폭이 줄어들어 더 이상 0을 포함하지 않으므로 이제 두 평균은 서로 다르다는 정반대의 결론이 나온다. 이와 같이 모순된 현상으로 인하여 통상적인 가설



<그림 2> 두 표본 t검정과 TOST에 의한 4가지 결과

검정방법은 동등성 확인에 적합하지 않다.

5. 모평균이외의 모수에 대한 동등성 검정

품질관리 현장에서는 모평균에 대한 동등성 검정이외에 모표준편차나 모분산, 모비율, 모결점수 및 고장률 등에 대한 동등성 검정을 해야 하는 경우가 있다. 이제 이들 각각의 경우 어떻게 동등성 검정을 실시할 수 있는지 간단히 살펴보고자 한다.

(1) 두 모표준편차에 대한 동등성 검정

두 모표준편차에 대한 동등성 검정을 위한 가설은 다음과 같다.

H_0: $|\sigma_1 - \sigma_2| \geq \theta$ vs $H_1 : |\sigma_1 - \sigma_2| < \theta$

이때 n이 큰 경우 표본표준편차 $s \sim N(\sigma, \dfrac{\sigma^2}{2(n-1)})$를 근사적으로 따르므로 구간가설검정을 이용하여 100(1-2α)% 신뢰구간인

$$(s_1 - s_2) \pm Z_\alpha \sqrt{\frac{\widehat{\sigma_1^2}}{2(n_1-1)} + \frac{\widehat{\sigma_2^2}}{2(n_2-1)}}$$

이 동등성 구간인 (-θ, θ)안에 들어가면 두 모표준편차간 동등하다고 결론을 내린다(여기서 $\widehat{\sigma_1^2}$과 $\widehat{\sigma_2^2}$은 표본분산 s_1^2과 s_2^2을 나타냄).

(2) 두 모분산에 대한 동등성 검정

두 모분산에 대한 동등성 검정을 위한 가설은 다음과 같다.

H_0: $|\sigma_1^2 - \sigma_2^2| \geq \theta$ vs H_1: $|\sigma_1^2 - \sigma_2^2| < \theta$

하지만 두 모분산간 동등한지는 다음과 같은 가설로도 점검할 수 있다.

H_0: $\sigma_1^2/\sigma_2^2 \leq \theta_1$ or $\sigma_1^2/\sigma_2^2 \geq \theta_2$ vs $H_1 : \theta_1 < \sigma_1^2/\sigma_2^2 < \theta_2$ $\Leftrightarrow$

H_{01}: $\sigma_1^2/\sigma_2^2 \leq \theta_1$ vs $H_1 : \theta_1 < \sigma_1^2/\sigma_2^2$ & H_{02}: $\sigma_1^2/\sigma_2^2 \geq \theta_2$ vs $H_{12} : \sigma_1^2/\sigma_2^2 < \theta_2$

앞의 두 가설은 V_1과 V_2가 두 표본에서의 표본분산이라고 하면 $\dfrac{V_1/\sigma_1^2}{V_2/\sigma_2^2} \sim F(n_1-1, n_2-1)$를 이용하여 동등성 검정을 실시할 수 있다.

(3) 두 모비율에 대한 동등성 검정

두 모비율에 대한 동등성 검정을 위한 가설은 다음과 같다.

H_0: $|p_1 - p_2| \geq \theta$ vs $H_1 : |p_1 - p_2| < \theta$

이때 구간가설검정을 이용하면 100(1-2α)% 신뢰구간인

$$(\hat{p}_1 - \hat{p}_2) \pm Z_\alpha \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

이 동등성 구간인 (-θ, θ)안에 들어가면 두 모비율간 동등하다고 결론을 내린다.

(4) 두 모결점수에 대한 동등성 검정

두 모결점수에 대한 동등성 검정을 위한 가설은 다음과 같다.

H_0:$|m_1 - m_2| \geq \theta$ vs $H_1 : |m_1 - m_2| < \theta$

이때 구간가설검정을 이용하면 100(1-2α)% 신뢰구간인 $(\widehat{m}_1 - \widehat{m}_2) \pm Z_\alpha \sqrt{\dfrac{\widehat{m}_1}{n_1} + \dfrac{\widehat{m}_2}{n_2}}$ 이 동등성 구간인 (-θ, θ)안에 들어가면 두 모결점수간 동등하다고 결론을 내린다.

(5) 두 고장률에 대한 동등성 검정

근래에 품질관리 현장에서는 신뢰성향상 활동이 활발히 전개되고 있다. 특히 전자부품의 경우 지수분포를 가정하고 신뢰성분석을 많이 실시한다. 따라서 두 고장률이 동등한지 동등성 검정에 대해 관심이 있을 수 있다. 두 고장률에 대한 동등성 검정을 위한 가설은 다음과 같다.

H_0: $|\lambda_1 - \lambda_2| \geq \theta$ vs $H_1 : |\lambda_1 - \lambda_2| < \theta$ $\Leftrightarrow$

H_0: $\lambda_1/\lambda_2 \leq \theta_1$ or $\lambda_1/\lambda_2 \geq \theta_2$ vs

$H_1 : \theta_1 < \lambda_1/\lambda_2 < \theta_2$ $\Leftrightarrow$

$H_{01}: \lambda_1/\lambda_2 \le \theta_1$ vs $H_{11}: \theta_1 < \lambda_1/\lambda_2$

& $H_{02}: \lambda_1/\lambda_2 \ge \theta_2$ vs $H_{12}: \lambda_1/\lambda_2 < \theta_2$

앞의 두 가설은 $r_i(i=1,2)$를 n_i개의 표본 중 고장수라고 하면 $\frac{r_1\lambda_1\hat{\lambda}_2}{r_2\hat{\lambda}_1\lambda_2} \sim F(2r_1, 2r_2)$를 이용하여 동등성 검정을 실시할 수 있다.

6. 결 론

동등성 검정은 의약분야에서 기존의 약과 새로운 약의 동등성을 점검하는데 많이 사용되어왔다. 하지만 공업통계분야에서는 별로 소개된 바가 없어 동등성 검정을 해야 되는데도 불구하고 통상적인 두 표본 t검정을 이용해왔다. 본 논문에서는 그렇게 하는 경우 일정한 크기의 표본에 대해 동질적이라고 선언한 표본평균의 차이가 표본크기기 큰 표본에 대해서는 비동질적이라고 하는 비합리적인 결론에 도달한다는 것을 보여주었다. 아울러 이때에는 TOST와 같은 합리적인 동등성 검정방법을 이용하든 또는 통상적인 100(1-2α)% 신뢰구간이 동등성 구간인 (-θ, θ)에 포함되는지 보고 동등성 검정을 해야 한다는 것을 설명했다.

마지막으로 공업통계분야에서는 두 모평균의 동등성 검정이외에 모분산, 모비율, 모결점수 및 모고장률에 대한 동등성 검정을 해야 될 때가 있다. 본 논문에서는 이때 어떤 방식으로 동등성 검정을 실시할 수 있는지 살펴보았다.

추후의 연구과제로는 모분산, 모비율, 모결점수 및 모고장률의 동등성 검정방법에 대한 통계적인 성질(예를 들어 1종 및 2종의 오류 등)을 살펴보고, 두 모집단간 모평균은 물론 모분산에 대한 동등성 검정은 어떻게 실시할 수 있는지 알아보고, 3 집단 이상의 모수의 동등성 검정에 대해서도 살펴보고자 한다.

참고문헌

[1] Allen I.E. and Seaman, C. A., Different, Equivalent or Both?, *Quality Progress*, July 2006, pp. 77-79.

[2] Lamprecht, J.L., *Applied Analysis for Process Improvement*, 2005, ASQ Quality Press.

[3] Hauck, W. and Anderson, S., Type of Bioequivalence and Related Statistical Considerations. *Journal of Clinical Phamacology, Therapy and Toxicology*, 1992, 30, pp. 181-187.

[4] Metzler, C. M., Bioavailability. A Problem in Equivalence, *Biometrics*, 1974, 30. pp. 309-317.

[5] Rogers, J.L., Howard, K.I. and Vessey, J.T., *Psychological Bulletin*, 1993, 113, pp. 553-565.

[6] Roy, T., Calibrated Nonparametric Confidence Sets, *Journal of Mathematical Chemistry*, 1997, 21, pp. 103-109.

[7] Schuirmann, D.J., On Hypothesis Testing to Determine if the Mean of a Normal Distribution is Contained in a Known Interval, *Biometrics*, 1981, 37, p. 617.

[8] Schuirmann, D.J., A Comparison of the Two One-sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability, *Journal of Pharmacokinetics and Biopharmaceutics*, 1987, 15 pp. 657- 680.

[9] Stein, J. and Doganaksoy, N., Sample Size Consideration for Assessing the Equivalence of Two Process Means, *Quality Engineering*, 1999-2000, 12, pp. 105-110.

[10] Westlake, W.J., Symmetric Confidence Intervals for Bioequivalence Trials, *Biometrics,* 32, 1976, pp. 741-744.

[11] Westlake, W.J., Response to T.B.L. Kirkwood: Bioequivalence Testing-A Need to Rethink, *Biometrics*, 1981, 37, pp. 589-594.