가설검정 유형에 의한 신뢰구간 추정의 종류 및 유도
- Derivation of Confidence Intervals According to the Use Types of Hypothesis Tests -

최 성 운*
Sungwoon Choi*

Abstract

The paper contributes deriving the confidence intervals according to the use types of hypothesis tests. The guidelines of usuage on the types of hypothesis tests and interval estimation are proposed. These formulars can be used to evaluate the effects of quality improvement activity.
Keywords : Confidence Intervals, Hypothesis Tests, Guidelines, Effects, Quality Improvement

1. 서론

기업의 개선활동에서 개선의 효과를 파악하는 통계적 기법으로 가설검정(Hypothesis Tests)과 SPC(Statistical Process Control) 관리도가 있다. 관리도는 로트, 날짜 등의 장기간 대량 데이터의 꺾은선 그래프 형태로 이상원인을 제거, 개선하고 우연원인에 의한 관리한계로 공정을 예측 관리하기 때문에 생산자 위험 $\alpha = 0.27\%$로 아주 작게 설계한다.

이와 반면에 가설검정[2, 7]은 비용과 시간이 많이 소요되지 않는 작업자, 실무자 중심의 단기적이고 작은 개선 활동의 효과를 파악하는 SQC(Statistical Quality Control) 기법이다. 따라서 관리도보다 생산자 위험 $\alpha = 5\%$, 1%로 크게 설계한다. 가설검정의 유의수준과 구간추정의 신뢰수준은 분포의 관점에서 합이 1이다. 관리도의 꺾은선 그래프의 한 점이 검정, 추정[1, 4, 8]에 해당하며 결국 관리도는 연속된 검정, 추정을 관리한계로 가시관리를 이용하여 공정의 개선을 추구하는 방법이다.

* 경원대학교 산업공학과

가설검정은 국가품질경영대회에 출전한 품질분임조 과제, 식스시그마 프로젝트에서 가장 많이 사용되는 기법이지만 아쉽게도 개선 용도별 가설검정과 구간추정 방법이 제대로 적용되고 있지 않고 있다. 또한 KS 3251, 3252[5, 6]가 제개정되면서 정규(U)분포, t분포, χ^2분포, F분포의 면적표가 바뀌면서 품질경영 실무자들은 심한 혼란을 일으키고 있다. 특히 KS 3251의 정규(U)분포표는 우측면적(α)이 주어지면서 표준화 변수 $U_{1-\alpha}$는 좌측면적을 찾는 것으로 잘못 제시되어 있다. 또한 t, χ^2, F 분포표 역시 주어진 표의 면적과 구하려고 하는 양측검정 α, 우측검정 α, 좌측검정 α가 구별되지 않고 분포표 상단에 위치해 있어 사용자들이 큰 혼선을 빚고 있다. 또한 현재 KS는 국제표준 60%, 외국표준 30%, 고유표준 10%로 구성되어 있고 현재도 WTO/TBT에서 권고하는 부합화를 위해 국제 표준의 번안 그대로(IDT) 또는 약간 수정(MOD)하여 제정되고 있다. 그러나 표준간 기호, 용어, 통계표, 사용용도 방법 등의 가이드라인이 없어 실무자뿐만 아니고 전문자 조차도 적용에 있어 많은 혼란과 의사소통의 애로를 겪고 있다.

따라서 본 연구에서는 계량연속형, 계수이산형 공정에서 용도별 검정의 종류를 선정하는 방안과 그에 따른 구간추정의 유도방법을 제시한다 또한 현재 KSA 3251, 3252에서 분포표의 주어진 면적과 구하려고 하는 유의수준의 면적을 명확히 제시[3]하고 이에 근거하여 추론식을 유도하고자 한다.

2. 추론에서 샘플링 오차의 설계 및 용도

가설검정(Hypothesis Test)은 개선활동에서 개선전후의 효과, 시제품 테스트에서 시제품 결과와 설계 목표치와의 차이, 시험분석시 실험값과 이론값의 차이를 파악하는 경우 사용된다. SQC의 가설검정, 관리도, 샘플링 검사에서 설계되는 샘플링 오차는 <표1>과 같다.

<표1> SQC 샘플링 오차

SQC 기법	대상	α	β	$1-\alpha$	$1-\beta$
가설검정	가설(H)	$H_0 \rightarrow H_1$	$H_1 \rightarrow H_0$	$H_0 \rightarrow H_0$	$H_1 \rightarrow H_1$
관리도	공정	정상→이상	이상→정상	정상→정상	이상→이상
샘플링 검사	로트 또는 배취	합격→불합격	불합격→합격	합격→합격	불합격→불합격

<표1>에서 α는 제1종오차, 생산자위험, 위험률, 기각률, 유의수준, β는 제2종오차, 소비자위험, $1-\alpha$는 신뢰수준, 포함인자, $1-\beta$는 검출력이라 한다. 가설검정에서 β가 α보다 리스크가 크므로 β를 작게하기 위해 α=5%, 1%로 크게 설정하고 유의수준(Significance Level)이라 부른다. 그러나 β 리스크가 큰 경우 α=10%로 더 크게 하여 검정을 수행한다. $1-\alpha$는 신뢰수준으로 검정에서 바깥쪽 α는 추정에서의 안쪽 $1-\alpha$와 같으므로 구간추정식을 유도할 경우 양측검정, 우측검정, 좌측검정은 양측추정, 우

측추정, 좌측추정과 대응관계식으로 유도된다. 교정에서 확장불확도를 구할 경우 $1-\alpha$는 포함인자(Coverage Factor)로 불리운다. MINITAB 등 통계 패키지에서는 관측검정 통계량이 갖는 우측, 좌측 면적을 P-Value로 계산하고 유의수준 α보다 작을 경우 유의적인 판정을 한다.

계량연속형 데이터는 계측기를 사용하여 비용과 시간이 많이 소요되므로 중요한 스펙에 적용되며 평균과 분산으로 요약정리된다. 계수형 데이터는 Unit로 세는 부적합품과 개개의 스펙으로 카운트되는 부적합이 있다. 용도별 추론의 종류는 <표2>와 같다.

<표2> 용도별 추론의 종류

데이터	용도	스펙	검정의 종류		검정의 방향	추정
계량연속형	정확도 개선	양쪽스펙	1개 모집단		양측검정	양측추정
		하한스펙	모평균검정		우측검정	우측추정
		상한스펙	σ_0 Known : Z검정 σ_0 Unknown : t검정		좌측검정	좌측추정
			2개 모집단		양측검정	양측추정
	정밀도 개선		모분산검정	1개 모집단	좌측검정	좌측추정
				2개 모집단	양측검정	양측추정
계수이산형	경영관리자 관심사항	한도견본 표본견본	모부적합품률 Z검정	1개 모집단	좌측검정	좌측추정
				2개 모집단	양측검정	양측추정
	작업, 기술자 관심사항		모부적합Z검정	1개 모집단	좌측검정	좌측추정
			모단위당부적합 Z검정	2개 모집단	양측검정	양측추정

3. 계량연속형 추론

3.1 Z 추론

KSA 3251, 3252의 정규분포표를 찾기 위한 표준화 변수는 U(k로도 되어 있음)로 되어 있지만 사용자들이 모평균 μ의 기호와 혼란을 일으켜 Z를 사용하기로 한다.

Z 분포표[3]는 좌측 $1-\alpha$ 면적의 $Z_{1-\alpha}$를 찾는 표로 설계되어 있다. 우측검정 α는 표의 $Z_{1-\alpha}$가 되며 좌측검정 α는 좌우대칭으로 $-Z_{1-\alpha}$ 값이 되고 양측검정 α는 $\pm Z_{1-\alpha/2}$가 된다.

양측추정은 $-Z_{1-\alpha/2} \le (\overline{x}-\mu)/(\sigma/\sqrt{n}) \le +Z_{1-\alpha/2}$에서 $\overline{x}+Z_{1-\alpha/2}\,\sigma/\sqrt{n^2} \ge \mu \ge \overline{x}-Z_{1-\alpha/2}\,\sigma/\sqrt{n}$, 우측추정은 $(\overline{x}-\mu)/(\sigma/\sqrt{n}) \le +Z_{1-\alpha}$에서 $\mu \ge \overline{x}-Z_{1-\alpha}\,\sigma/\sqrt{n}$, 좌측추정은 $-Z_{1-\alpha} \le (\overline{x}-\mu)/(\sigma/\sqrt{n})$에서 $\overline{x}+Z_{1-\alpha}\,\sigma/\sqrt{n} \ge \mu$가 된다.

<표2>와 같이 양쪽스펙에는 양측검추정, 하한스펙에는 우측검추정, 상한스펙에는 좌측검추정을 사용한다.

3.2 t 추론

t분포표[3]는 Z분포표[3]와 마찬가지로 좌측 $1-\alpha$ 면적의 $t_{1-\alpha}(\nu)$를 찾는 표로 설계되어있다. 우측검정 α는 표의 $t_{1-\alpha}(\nu)$가 되며 좌측검정 α는 좌우대칭으로 $-t_{1-\alpha}(\nu)$가 되고 양측검정 α는 $+t_{1-\alpha/2}(\nu)$가 된다. 여기서 자유도 ν는 DF, ϕ로도 쓰이며 불편분산 s^2은 MS, V로 사용하며 KS 가이드라인에서 통일지침을 제공해 주어야 한다.

양측추정은 $-t_{1-\alpha/2}(\nu) \le (\overline{x}-\mu)/(s/\sqrt{n}) \le +t_{1-\alpha/2}(\nu)$에서 $\overline{x}+t_{1-\alpha/2}\,s/\sqrt{n} \ge \mu \ge \overline{x}-t_{1-\alpha/2}\,s/\sqrt{n}$, 우측추정은 $(\overline{x}-\mu)/(s/\sqrt{n}) \le +t_{1-\alpha}(\nu)$에서 $\mu \ge \overline{x}-t_{1-\alpha}(\nu)\,s/\sqrt{n}$, 좌측검정은 $-t_{1-\alpha} \le (\overline{x}-\mu)/(s/\sqrt{n})$에서 $\overline{x}+t_{1-\alpha}(\nu)\,s/\sqrt{n} \ge \mu$가 된다.

<표2>와 같이 양쪽스펙에는 양측검정, 하한스펙에는 우측검추정, 상한스펙에는 하한검추정을 사용한다.

3.3 χ^2 추론

χ^2 분포표[3]는 Z, t 분포표[3]와 같이 좌우대칭이 아니므로 좌측 $1-\alpha$의 면적 $\chi^2_{1-\alpha}(\nu)$과 좌측 α의 면적 $\chi^2_{\alpha}(\nu)$를 찾는 표로 설계되어 있다. 우측검정 α는 표의 $\chi^2_{1-\alpha}(\nu)$가 되며 좌측검정 α는 표의 $\chi^2_{\alpha}(\nu)$가 되고 양측검정 α는 하측은 $\chi^2_{\alpha/2}(\nu)$, 상측은 $\chi^2_{1-\alpha/2}(\nu)$가 된다.

양측 추정은 $\chi^2_{\alpha/2}(\nu) \le (n-1)s^2/\sigma^2 \le \chi^2_{1-\alpha/2}(\nu)$에서 $(n-1)s^2/\chi^2_{\alpha/2}(\nu) \ge \sigma^2 \ge (n-1)s^2/\chi^2_{1-\alpha/2}(\nu)$, 우측 추정은 $(n-1)s^2/\sigma^2 \le \chi^2_{1-\alpha}(\nu)$에서 $\sigma^2 \ge (n-1)s^2/\chi^2_{1-\alpha}(\nu)$, 좌측추정은 $\chi^2_{\alpha}(\nu) \le (n-1)s^2/\sigma^2$에서 $(n-1)s^2/\chi^2_{\alpha}(\nu) \ge \sigma^2$이다.

<표2>와 같이 모분산이 작으면 좋으므로 좌측 검추정을 사용한다.

3.4 F 추론

F분포표[3]는 χ^2분포표[3]와 같이 1상한의 양수값을 가지는 분포로 좌측 $1-\alpha$ 면적

의 $F_{1-\alpha}(\nu_1, \nu_2)$를 찾는 표로 설계되어 있다. ν_1, ν_2, $1-\alpha$의 3차원으로 설계되어 있어 $F_\alpha(\nu_1, \nu_2) = 1/F_{1-\alpha}(\nu_2, \nu_1)$의 관계식을 이용한다.

우측검정 α는 표의 $F_{1-\alpha}(\nu_1, \nu_2)$가 되며 좌측검정 α는 $F_\alpha(\nu_1, \nu_2)$로 표에서 $F_{1-\alpha}(\nu_2, \nu_1)$가 되고 양측검정 α는 하측은 $F_{\alpha/2}(\nu_1, \nu_2)$로 표에서 $F_{1-\alpha/2}(\nu_2, \nu_1)$가 되고 상측은 $F_{1-\alpha/2}(\nu_1, \nu_2)$가 된다.

양측추정은 $F_{\alpha/2}(\nu_1, \nu_2) \le (s_1^2/\sigma_1^2)/(s_2^2/\sigma_2^2) \le F_{1-\alpha/2}(\nu_1, \nu_2)$에서 $\frac{s_1^2}{s_2^2} 1/F_{\alpha/2}(\nu_1, \nu_2) \ge \frac{\sigma_1^2}{\sigma_2^2} \ge \frac{s_1^2}{s_2^2} 1/F_{1-\alpha/2}(\nu_1, \nu_2)$, 우측추정은 $(s_1^2/\sigma_1^2)/(s_2^2/\sigma_2^2) \le F_{1-\alpha}(\nu_1, \nu_2)$에서 $\sigma_1^2/\sigma_2^2 \ge \frac{s_1^2}{s_2^2} 1/F_{1-\alpha}(\nu_1, \nu_2)$, 좌측추정은 $F_\alpha(\nu_1, \nu_2) \le (s_1^2/\sigma_1^2)/(s_2^2/\sigma_2^2)$에서 $\frac{s_1^2}{s_2^2} 1/F_\alpha(\nu_1, \nu_2) \ge \sigma_1^2/\sigma_2^2$이다.

<표2>와 같이 2개의 모집단의 비교는 정보가 없다는 전제하에 양측검추정을 수행하는 것이 합리적이다.

4. 계수이산형 추론

4.1 모부적합품률 추론

1개 이상의 부적합으로 구성되며 Unit로 세는 부적합품은 초기하, 이항, 포아송 분포를 사용하나 $np \ge 5$인 경우 정규근사로 3.1절의 Z분포표[3]를 사용한다.

<표2>와 같이 부적합품률은 좌측검추정을 사용하므로 $-Z_{1-\alpha} \le (p-P)/(p(1-p)/n)^{1/2}$에서 $p + Z_{1-\alpha}(p(1-p)/n)^{1/2} \ge P$가 된다.

4.2 모결점수 추론

개개의 스펙을 세며 일정한 샘플중의 부적합은 포아송 분포를 적용하나 $c \ge 5$인 경우 정규근사로 3.1절의 Z분포표[3]를 사용한다.

<표2>와 같이 모부적합수는 좌측검추정을 사용하므로 $-Z_{1-\alpha} \le (c-C)/\sqrt{c}$ 에서 $c + Z_{1-\alpha}\sqrt{c} \ge C$가 된다.

4.3 모 단위당 부적합 추론

4.2절의 부적합과 다르게 샘플이 일정하지 않은 경우 단위당 부적합을 적용한다. 정규근사로 $u \geq 5$인 경우 3.1절의 Z분포표[3]를 사용하며 <표2>와 같이 모 단위당 부적합은 좌측검추정을 사용하므로 $-Z_{1-\alpha} \leq (u-U)/(u/n)^{1/2}$에서 $u+Z_{1-\alpha}(u/n)^{1/2} \leq U$가 된다.

5. 결론

본 연구에서는 개선효과를 파악하는 가설검정을 스펙, 데이터의 평균, 분산, 부적합품, 부적합, 단위당 부적합에 따른 용도별 검정종류의 적용 방안을 제시하였다. 또한 가설 검정에서 유의적일 경우 검정의 유형에 맞는 구간 추정식을 유도하였다. KSA 3251, 3252에서 혼란스러운 분포표를 주어진 수치표의 면적(Given Area)과 유의수준의 찾는 면적(Find Area)으로 체계적으로 재설계하였다.

6. 참 고 문 헌

[1] 최성운, “신뢰도와 신뢰수준을 고려한 기대수명 공차구간 설정에 관한 연구”, 대한안전경영과학회지, 7 (2) (2005) : 73-83.

[2] 최성운, “소표본인 경우 비모수 순위척도를 이용한 정규성 검정”, 대한안전경영과학회지, 10 (3) (2008) : 237-247.

[3] 최성운, 통계적 품질 및 공정관리 강의 Packet, 경원대학교, 2009.

[4] Casella G., Statistical Inference, 2nd Edition, Duxbury Press, 2001.

[5] KSA 3151-1 : 2008 데이터의 통계적 해석 방법 - 제 1부 : 데이터의 통계적 기술.

[6] KSA 3151-2 : 2001 데이터의 통계적 해석 방법 - 제 2부 : 평균 및 분산에 관한 검정방법 및 추정방법.

[7] Lehmann E.L., Romano J.P., Testing Statistical Hypothesis, 3rd Edition, Springer, 2008.

[8] Young G.A., Smith R.L., Essentials of Statistical Inference, Cambridge University Press, 2005.