확률변수(random variable)
확률 : random or probability
- 표본공간에 있는 각 원소에 값을 대응시켜 주는 규칙 또는 함수
- 표본공간에 있는 값을 숫자로 변경
- 확률변수 = X , 확률변수의 값 = y
예) X= 동전 2개를 던질 때 동전 앞면의 수
표본공간 : S = [HH, HT, TH, YY] ----> 확률변수 : X = {2,1,1,0}

종류
이산확률변수(discrete random variable)
- 특정한 수치만을 가지는 확률변수(정수)
- ex)불량품의 수, 고속도로에서 사고건수, 방문자수 등
연속확률변수(continuous random variable)
- 어떤 범위에서 연속적인 값을 취할 수 있는 확률변수(실수)
- ex)전구의 수명, 몸무게, 체온, 통근시간 등

이항분포 (Binominal Distribution)
- 성공률이 ㅠ인 베르누이 시행을 n번 시행했을 때의 확률분포
- ex)100개의 제품을 불량품과 양호품으로 구분하는 경우
- 1000명의 유권자에게 정부정책에 대한 찬성과 반대를 묻는 경우
X = 성공확률변수 , ㅠ =1회시행시 성공확률 , n=시행횟수

포아송분포(Poisson distribution)
- 랜덤하게 선택한 일정한 단위 시간이나 공간 내에 발생하는 사건의 개수를 설명
- 보통 단위시간당 도착에 대한 모델에 많이 사용되므로 시간이 주로사용
- 반대로 도착에 따른 시간을 측정하기 위해서는 연속분포인 지수분포를 사용함
- 경영학에서는 대기시간 모형에서 많이 사용

표집분포(sampling distribution)
- 모집단에서 추출한 표본크기 n개인 추정치의 확률분포
- 모집단에서 일정한 크기(n)로 표본을 모두 (k개) 뽑아서 각 표본의 평균을 계산하였을 때,
- 그 표본의 평균 X1 ,X2,X3...Xk의 확률분포

평균 : 모집단의 평균과 표본들의 평균은 같음
표준오차 : 통계량의 표준편차
-
표본오차 vs 표준편차 vs 표준오차
-

표집분포는 표준오차의 크기에 의해 오차범위의 크기가 결정됨

추정
표본의 평균과 표준오차를 구해서 모수의 범위를 구하는 것.
신뢰구간 : 일정한 확률범위 내에서 모수의 값이 포함될 가능성이 있는 범위
90% 95% 99% -> 확률값 95%로 추정
종류 : 점추정 , 구간추정

가설검정

모수는 얼마이다라고 정하고 그것이 맞는 지 틀리는지를 검증하는 방법
유의수준(a) : 모수와 통계량의 차이가 커서 확률적으로 가설을 기각할수 있는 값
1% 5% 10% --> 중에서 5%
종류 : 귀무가설 , 연구가설

구간추정

모수의 값이 속할 것으로 기대되는 일정한 범위(신뢰구간)을 이용하여 모수를 추정
표본오차가 존재 -> 추정 시에는 점추정보다는 구간추정을 이용

신뢰구간

일반적으로 신뢰수준은 (1 – a)로 측정
신뢰수준의 확률 : 90% 95% 99% 중에서 95%
신뢰구간 95%의 의미
- 동일한 모집단에 대해서 동일한 방법으로 표본을 다시 뽑아서 신뢰구간을 구하게 되면
  100번중 95번은 모수를 포함(표집분포의 의미)

방법

(가정)모집단의 표준편차 o를 알 경우 : 표준정규분포
(실체) 모집단의 표준편차 o를 모를 경우 : student t 분포

가설검정

모집단 모수의 값을 설정하고 (가설설정), 표본 통계치를 통해 확률적으로 진위를 판정하는 과정

가설

귀무가설(Null Hypothesis)
- 기존에 알려져 있는 사실(통계적 검정대상)
대립가설, 연구가설(Alternative Hypothesis)
- 새로운 사실, 현재 믿음에 변화가 있는 사실, 뚜렷한 증거로 입증하려고 하는 주장
 ex) b아스크림회사에서 판매하는 아스크림중 파인트의 용량은 320g이어야 한다.
 b아스크림 회사가 진짜 350g을 판매하고 있는가를 검증
귀무가설 – 320g
대립가설 – 350g

T-검정

- 두 집단 간의 평균을 비교하는 검정방법
평균 : 한 집단을 대표할 수 있는 대푯값.
 – 두 모집단의 평균의 차이유무를 판단하는 통계적 검정방법
단순히 차이의 존재여부를 떠나 그 정도의 통계적 유의미성까지 검정하는 방법
유의미성 : 두 집단의 평균 차이

양측검정, 단측검정 : 보수적인(conservative) 정도(degree)에 차이

양측검정 (two sided test) = 좌측검정
단측검정 (one sided test) = 우측검정

모비율 검정
: 모집단 비율에 대해 가설검정을 할 때 '이항분포의 Normal분포 근사(approximation)'를 활용
- 모집단 비율에 대한 가설검정을 할 때도, 다른 가설검정과 같이 검정통계량을 구한 다음 임계치와 비교해서 검정통계량이 분포의 바깥부분

<ex>
어떤 제품의 길이의 표준편차는 0.1%이다. 원가절감을 위해 제품에 사용되는 원자재의 공급업체를 바꾸고자 m사의 원자재를 사용하여 시험한 결과 다음 데이터를 얻었다
[Data : 2.2, 2.4, 2.1, 2.5, 2.0, 2.4, 2.5, 2.3, 2.9, 2.7, 2.8] 시료 수는 11
기존의 제품에 비해 산포가 달라졌다고 할 수 있는 가를 유의수준 5% 로 검정하시오.
또한 유의 하다면 모분산의 신뢰한계를 신뢰율 95%로 추정하시오.

	H_1	판 정	추 정
σ² 기지	$\sigma^2 \neq \sigma_0^2$	$\chi_0^2 > \chi^2_{1-\alpha/2}(\nu)$또는 $\chi_0^2 < \chi^2_{\alpha/2}(\nu)$이면 H_0 기각	$\frac{S}{\chi^2_{1-\alpha/2}(\nu)} \leq \sigma^2 \leq \frac{S}{\chi^2_{\alpha/2}(\nu)}$
	$\sigma^2 > \sigma_0^2$	$\chi_0^2 > \chi^2_{1-\alpha}(\nu)$이면 H_0 기각	$\chi_L^2 = \frac{S}{\chi^2_{1-\alpha}(\nu)}$
	$\sigma^2 < \sigma_0^2$	$\chi_0^2 < \chi^2_{\alpha}(\nu)$이면 H_0 기각	$\chi_U^2 = \frac{S}{\chi^2_{\alpha}(\nu)}$

▶ 검정결과 귀무가설이 기각되면 모분산의 신뢰한계를 추정하고, 귀무가설이 채택되면 모분산의 신뢰한계의 추정은 의미가 없음.

모분산에 대한 검정과 추정

모평균의 검정.추정: 정확도에 대하여 추론. 모분산의 검정.추정은 정밀도에 관하여 추론.
A, B 집단의 서로 독립된 두 모집단으로부터 두 표본의 산포에 대하여 두 모분산의 차이를 검, 추정할 경우에는 F분포를 따릅니다.

1. 주장하고자 하는 사실(대립가설)을 바탕으로 귀무가설과 대립가설을 설정합니다.

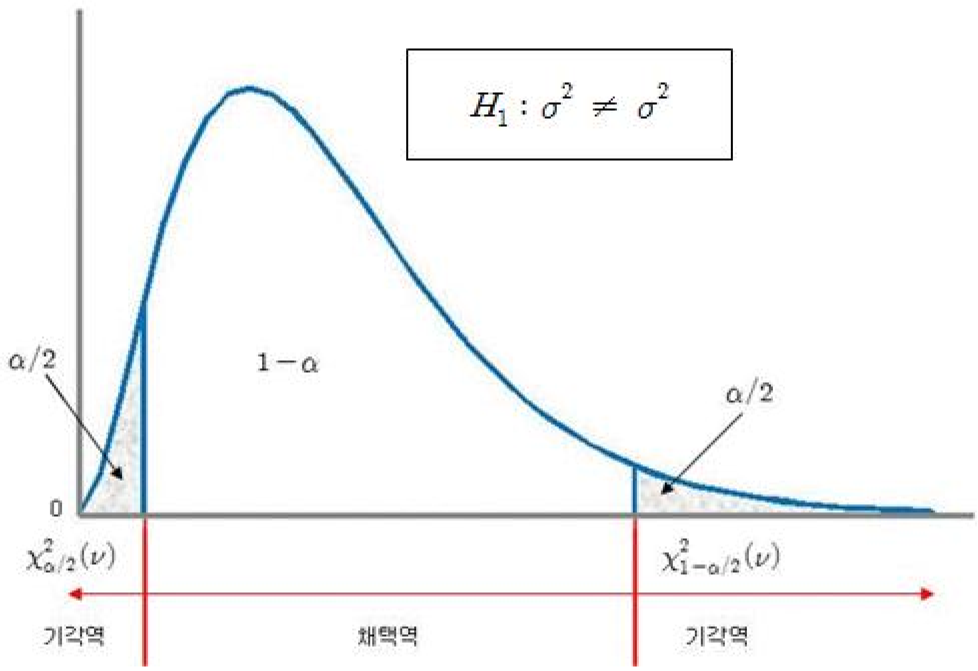
구분		귀무가설	대립가설
한 개의 모분산 (χ^2분포)	양쪽검정	$H_0 : \sigma^2 = \sigma_0^2$	$H_1 : \sigma^2 \neq \sigma_0^2$
	한쪽검정	$H_0 : \sigma^2 \leq \sigma_0^2$	$H_1 : \sigma^2 > \sigma_0^2$
		$H_0 : \sigma^2 \geq \sigma_0^2$	$H_1 : \sigma^2 < \sigma_0^2$

2. 유의수준을 설정
   - p-value : 0.05

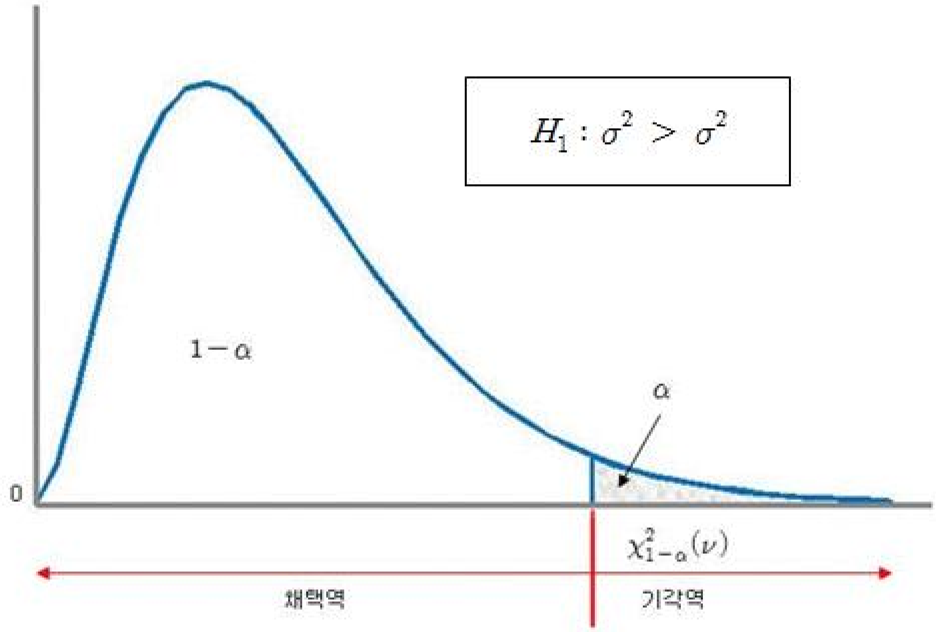
3. 검정통계량을 계산:

$$\chi_0^2 = \frac{S}{\sigma_0^2}$$

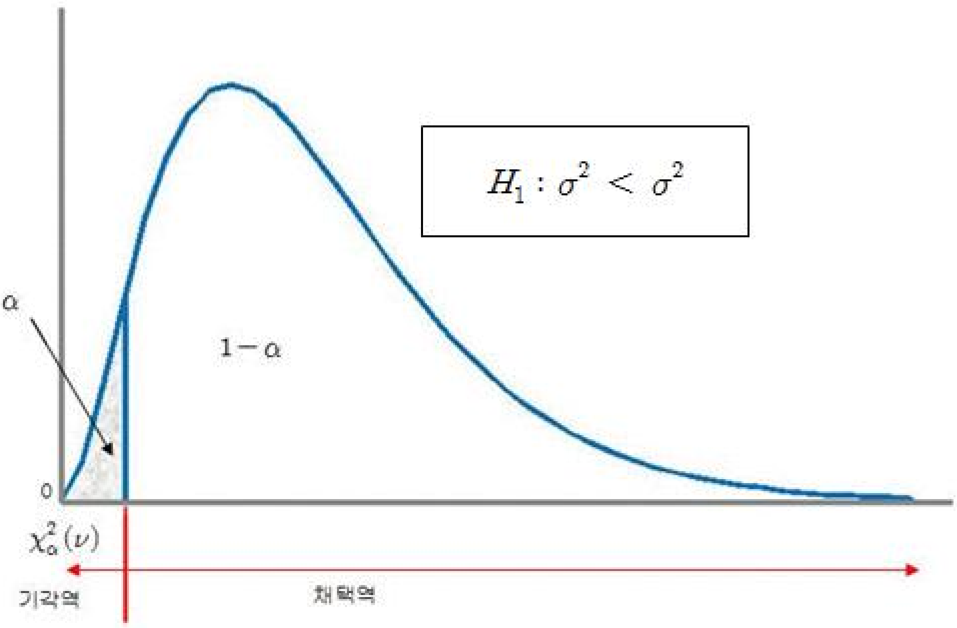
4. 기각역 설정



?



?



?

▶ 자유도(ν)를 확정하고 기각역을 정하기 위하여 χ²분포표를 보고 설정합니다.

▶ χ²분포표 보는 법: http://blog.naver.com/lchry/220477603859

?

5. 판정: 검정통계량의 값과 기각역을 비교하여 판정

	H_1	판 정	추 정
σ² 기지	$\sigma^2 \neq \sigma_0^2$	$\chi_0^2 > \chi^2_{1-\alpha/2}(\nu)$또는 $\chi_0^2 < \chi^2_{\alpha/2}(\nu)$이면 H_0 기각	$\frac{S}{\chi^2_{1-\alpha/2}(\nu)} \le \sigma^2 \le \frac{S}{\chi^2_{\alpha/2}(\nu)}$
	$\sigma^2 > \sigma_0^2$	$\chi_0^2 > \chi^2_{1-\alpha}(\nu)$이면 H_0 기각	$\chi_L^2 = \frac{S}{\chi^2_{1-\alpha}(\nu)}$
	$\sigma^2 < \sigma_0^2$	$\chi_0^2 < \chi^2_{\alpha}(\nu)$ 이면 H_0 기각	$\chi_U^2 = \frac{S}{\chi^2_{\alpha}(\nu)}$

▶ 검정결과 귀무가설이 기각되면 모분산의 신뢰한계를 추정하고, 귀무가설이 채택되면 모분산의 신뢰한계의 추정은 의미가 없음.

그럼 예제를 통하여 모분산의 검정과 추정방법에 대하여 알아보도록 하겠습니다.

< 예제 1 >

어떤 제품의 길이의 표준편차는 0.10%이다. 원가절감을 위해 제품에 사용되는 원자재의 공급업체를 바꾸고자 M사의 원자재를 사용하여 시험한 결과 다음의 데이터를 얻었다.

[Data: 2.2, 2.4, 2.1, 2.5, 2.0, 2.4, 2.5, 2.3, 2.9, 2.7, 2.8] 시료 수는 11

기존의 제품에 비해 산포가 달라졌다고 할 수 있는가를 유의수준 5%로 검정하시오. 또한 유의하다면 모분산의 신뢰한계를 신뢰율 95%로 추정하시오.

◆ 모분산의 검정

1. 가설설정:

$$H_0 : \sigma^2 = 0.10^2\% \;,\; H_1 : \sigma^2 \neq 0.10^2\%$$

2. 유의수준: α = 0.05

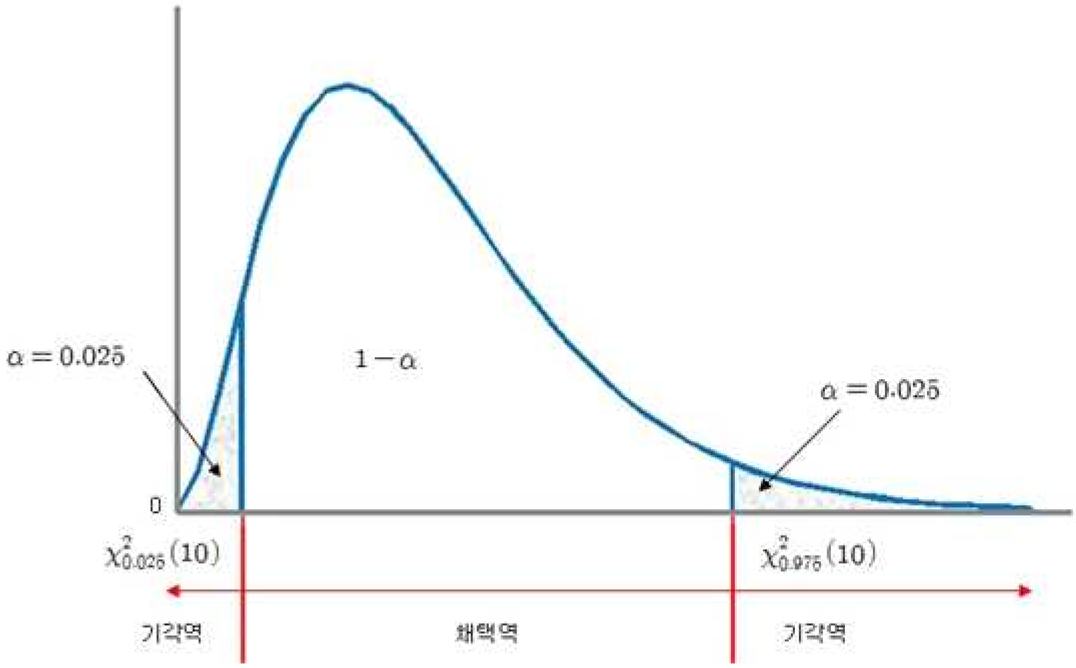
3. 검정통계량:

$$\chi_0^2 = \frac{S}{\sigma_0^2} \quad = \frac{0.805}{0.10^2} = 80.5$$

4. 기각역:

$$\text{하한측기각역} : \chi^2_{\alpha/2}(\nu) = \chi^2_{0.025}(10) = 3.25$$

$$\text{상한측기각역} : \chi^2_{1-\alpha/2}(\nu) = \chi^2_{0.975}(10) = 20.48$$



?

기각역은 산포의 변화 유, 무를 검정함으로 상측과 하측에 설정되고 검정통계량 값이 상한측 분위점 보다 더 크고, 또는 하한측 분위점 보다 더 작으면 H?를 기각한다는 뜻임

$$\text{즉 검정통계량}(\chi_0^2) < \chi^2_{0.025}(10) \text{ 또는}$$

$$\text{검정통계량}(\chi_0^2) > \chi^2_{0.975}(10) \text{ 이면 } H_0 \text{ 기각}$$

5. 판정:

$$\chi_0^2 > 20.48 \rightarrow H_0 \text{ 기각}$$

즉 새로운 원자재를 사용한 제품의 산포가 달라졌다고 할 수 있다.

◆ 모분산의 신뢰한계 추정

귀무가설(H?)이 기각 되었으므로 모분산의 신뢰구간을 추정하여야 함

▶ 모분산의 95% 신뢰구간 추정

$$1-\alpha = P(\chi_L^2 \le \chi^2 \le \chi_U^2)$$

$$= P\left(\chi_{\alpha/2}^2(\nu) \le \frac{S}{\sigma^2} \le \chi_{1-\alpha/2}^2(\nu)\right) \rightarrow \text{역수를 취하면}$$

$$= P\left(\frac{S}{\chi_{1-\alpha/2}^2(n-1)} \le \sigma^2 \le \frac{S}{\chi_{\alpha/2}^2(n-1)}\right)$$

$$= \frac{S}{\chi_{0.975}^2(10)} \le \sigma^2 \le \frac{S}{\chi_{0.025}^2(10)} = \frac{0.805}{20.48} \le \sigma^2 \le \frac{0.805}{3.25}$$

$$\text{따라서 } \chi^2 = 0.03931\% \le \sigma^2 \le 0.24769\%$$

< 예제 2 >

어떤 제품의 과거 길이간의 표준편차는 10.0mm 이하인 것을 알고 있다.최근 15개의 제품의 표준편차는
14.50이였다. 최근 제품의 산포가 커졌다고 할 수 있는지를 유의수준 5%로 검정하시오. 또한 유의하다면 모분산의 신뢰한계를 신뢰율 95%로 추정하시오.

◆ 모분산의 검정

1. 가설설정: 최근 생산한 제품의 산포가 과거 생산한 제품보다 산포가 커졌는가를 검정하여야 함으로 한쪽검정이고 대립가설을(최근제품 >과거제품) 설정합니다.

$$H_0 : \sigma^2 \le \sigma_0^2,\ H_1 : \sigma^2 > \sigma_0^2$$

2. 유의수준: α = 0.05

3. 검정통계량:

$$\chi_0^2 = \frac{S}{\sigma_0^2} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{14 \times 14.5^2}{10.0^2} = 29.435$$

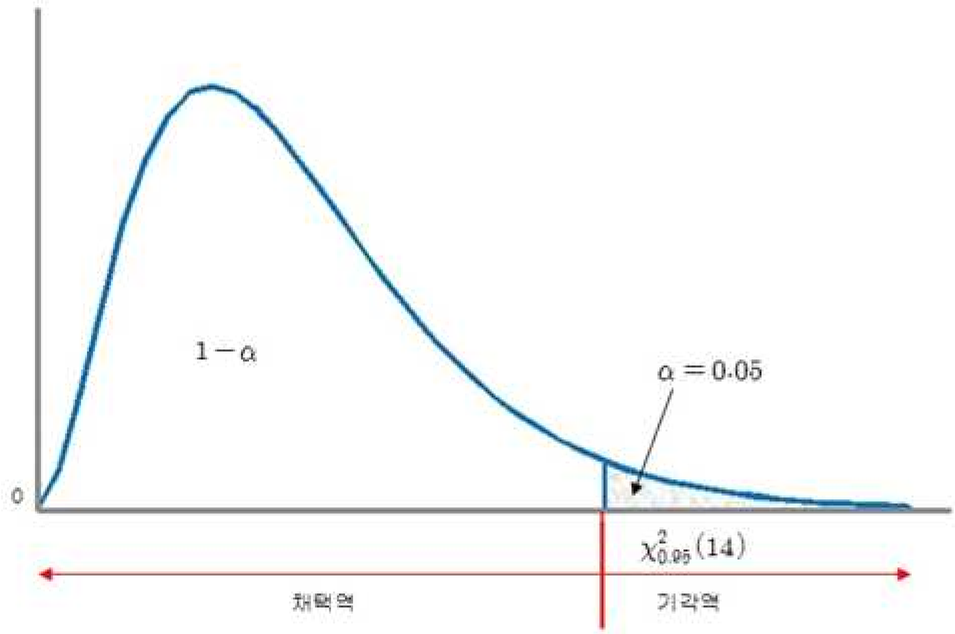
4. 기각역:

?

$$\chi_{1-\alpha}^2(n-1) = \chi_{0.95}^2(14) = 23.68$$

?

?



?

기각역은 최근제품의 산포가 큰가를 검정함으로 상측에 설정되고 검정통계량 값이 상한측 분위점 보다 더 크면 H?를 기각한다는 뜻입니다.

즉 검정통계량(χ_0^2) > $\chi^2_{0.95}(14)$이면 H_0 기각

5. 판정:

$\chi_0^2 > 23.68 \rightarrow H_0$기각

즉 최근 제품의 산포가 과거보다 커졌다고 할 수 있다.

◆ 모분산의 95% 신뢰하한의 추정

검정결과 모분산이 과거보다 산포가 커졌으므로 신뢰하한을 추정해야 함

왜 신뢰하한을 추정을 하는가 하면 산포가 커졌다는 것을 검정을 통해서 증명이 되었고 그럼 산포가 얼마만큼 커졌는가를 알아보기 위해서 산포의 최소값을 구해야 합니다.

$$\sigma_L^2 = \frac{(n-1)s^2}{\chi^2_{1-\alpha}(n-1)} = \frac{(15-1)14.5^2}{\chi^2_{0.95}(14)}$$

$$= \frac{14 \times 14.5^2}{23.68} = 124.30321mm$$

$$\text{따라서 } \sigma^2 \geq 124.30321mm$$

< 예제 3 >

어떤 제품의 종전 길이 간의 표준편차는 10.0mm 이하로 측정되고 있음을 알고 있다.

최근 금형을 개선한 후 15개의 제품을 조사하여 표준편차를 구해보았더니 6.5mm로 나타났다.

개선된 금형에서 생산되고 있는 제품의 표준편차가 작아졌다고 할 수 있는가를 유의수준 5%로

검정하시오. 그리고 검정결과가 유의하다면 모분산의 신뢰한계를 신뢰율 95%로 추정하시오.

◆ 모분산의 검정

1. 가설설정: 최근 생산한 제품의 산포가 과거 생산한 제품보다 산포가 작아졌는가를 검정하여야 함으로

한쪽검정이고 대립가설을(최근제품 <과거제품)설정합니다.

$$H_0 : \sigma^2 \geq \sigma_0^2$$

$$H_1 : \sigma^2 < \sigma_0^2$$

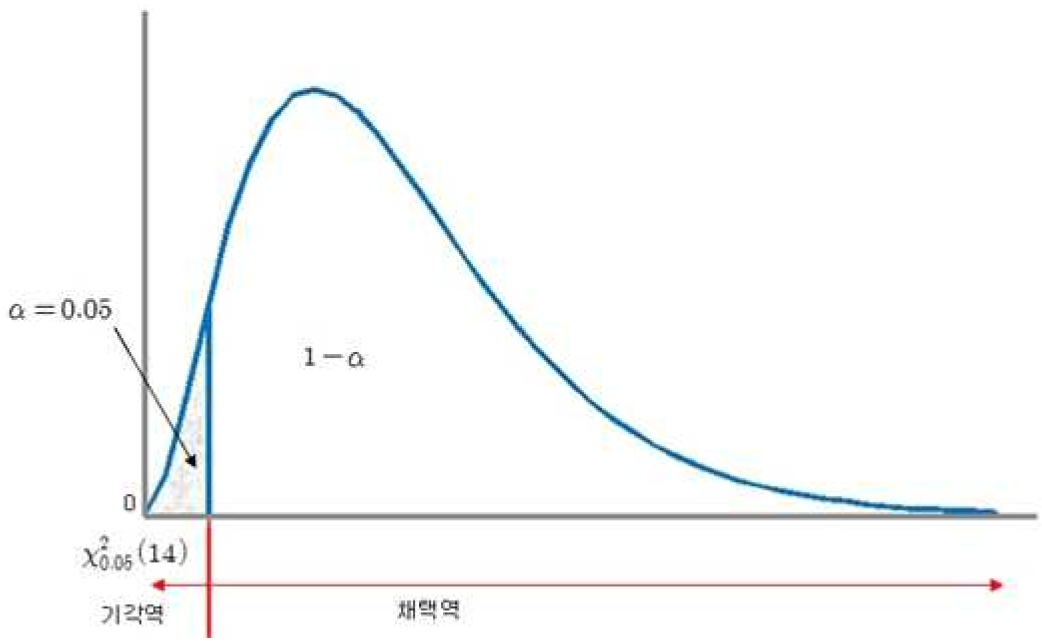
2. 유의수준: α = 0.05

3. 검정통계량:

$$\chi_0^2 = \frac{S}{\sigma_0^2} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{14 \times 6.5^2}{10.0^2} = 5.915$$

4. 기각역:

$$\chi_{1-\alpha}^2(n-1) = \chi_{0.05}^2(14) = 6.57$$



?

기각역은 최근제품의 산포가 작은가를 검정함으로 α는 하측에 설정되고, 검정통계량 값이 하한측 분위점 보다 더 작으면 H?를 기각한다는 뜻입니다.

즉 검정통계량(χ_0^2) < $\chi_{0.05}^2(14)$이면 H_0 기각

5. 판정:

$$\chi_0^2 < 6.57 \rightarrow H_0 \text{기각}$$

즉 개선된 금형에서 생산되는 제품의 산포가 작아졌다고 할 수 있다.

◆ 모분산의 95% 신뢰상한의 추정

검정결과 모분산이 과거보다 산포가 작아졌으므로 신뢰상한을 추정해야 함

왜 신뢰상한을 추정을 하는가 하면 산포가 작아졌다는 것을 검정을 통해서 증명이 되었고 그럼 산포가
얼마만큼 작아졌는가를 알아보기 위해서 산포의 최대값을 구해야 합니다.

$$\sigma_U^2 = \frac{(n-1)s^2}{\chi^2_{0.05}(n-1)} = \frac{(15-1)6.5^2}{\chi^2_{0.05}(14)}$$

$$= \frac{14 \times 6.5^2}{6.57} = 90.03044mm$$

$$\text{따라서}\ \sigma^2 \leq 90.03044mm$$