Seoul National University

제 5장. 통계적 추론

추정의 원리

- **모수 (parameter)** : 모집단의 특성을 나타내는 상수 ($\mu,\ \sigma^2,\ p,\ \rho$)
- **추정량 (estimator)** : 모수를 추정하는 데 사용하는 통계량 ($\hat{\theta}$)
- **추정값 (estimate)** : 추정량의 관측값

- **추정(estimation)**

  : N개의 원소로 된 모집단에서 n 개의 표본을 추출한 후, 이를 토대로 모집단의 모수의 값을 추측하는 과정

  : 표본을 이용하여 이러한 모집단의 특성치에 대한 추측값을 제공하고 그 오차한계를 제시하는 과정

  - 점추정(point estimation) : 모수를 어떤 하나의 값으로 추측하는 것
  - 구간추정(interval estimation): 모수를 어떤 구간으로 추측하는 것

점추정

- 점추정(point estimation)

  : 선정된 추정량을 추출된 표본자료에 대입하여 얻은 단일 추정치에 의하여 모수를 추정하는 방법

  - 점추정량(point estimator) : 점추정에 사용된 통계량, 예) $\hat{\mu} = \bar{X}, \ \hat{\sigma}^2 = S^2$
  - 점추정치, 점추정값(estimate): 실제 표본에서 계산 된 값, 예) $\bar{x}, \ s^2$

- 표집오차(sampling error)

: 모집단 전체를 관측하지 않고, 부분 집합인 표본만 관측함으로써 생기는 오차

: 모집단의 모수(θ)와 모수의 추정량($\hat{\theta}$)의 차이

: 표집오차 = $\hat{\theta} - \theta$

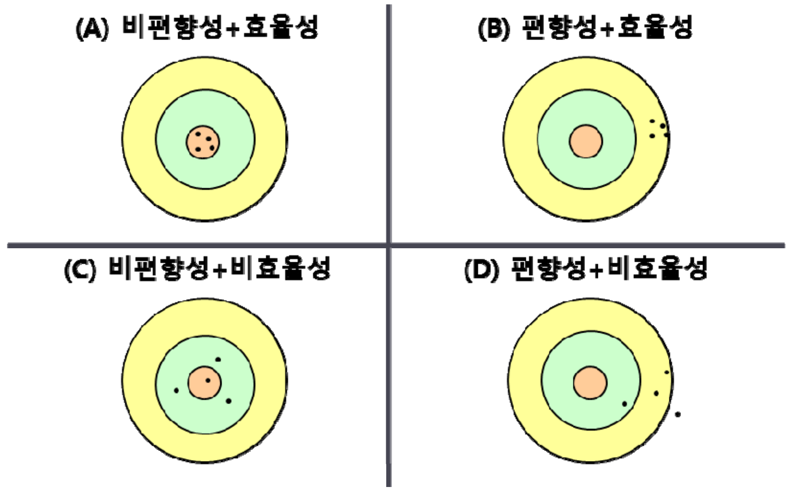
- 표집오차는 변동(variation)과 편의(bias)로 분해될 수 있다.

$$\hat{\theta} - \theta = \hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta$$
$$= \left[\hat{\theta} - E(\hat{\theta}) \right] + \left[E(\hat{\theta}) - \theta \right]$$

점추정의 바람직한 성질

- 불편성 (Unbiasedness)
  : 추정량의 기대값이 모수와 같아짐
  : $E(\hat{\theta})=\theta$ 이면 $\hat{\theta}$ 는 θ 의 불편추정량 (unbiased estimator)
  : (의미) 주어진 표본추출 방법으로 표본을 뽑고 그 추정량을 표본으로부터 계산하는 작업을 무한히 반복하였을 경우, 그렇게 해서 얻어지는 추정량 값들이 평균적으로는 모수와 일치하게 되는 것.

- 효율성(Efficiency)
  : 추정량의 분산이 더 작은 것이 상대적으로 더 효율적(relatively efficient)
  : 추정량의 분산이 작다는 것은 모수에 가까운 값이 나올 확률이 더 높다는 것을 의미하므로 좋은 추정량의 성질이 될 수 있음



- 최소분산 불편추정량
  : 같은 표본에서 도출된 불편추정량 중에서 분산이 최소가 되는 추정량

구간 추정

- **구간추정(interval estimation)**
  : 모수를 하나의 값으로 추정하는 것이 아니라 점추정량 값에 표집오차의 크기를 함께 반영하여 어떤 구간으로 나타내는 것
  : 점추정은 모수를 하나의 값으로 예측하기 때문에, 우리가 구한 추정치가 실제 모수와 얼마나 떨어져있는지에 대한 정보가 없는데 이를 보완한 추정 방법이 구간추정임.
- **신뢰구간(confidence interval)**
  : 구간추정을 통하여 얻어 지는 구간
  : 신뢰수준과 구간의 크기를 통해 결정 됨
- **신뢰수준(confidence level)**
  : 여러 번 구한 신뢰구간 중 추정하고자 하는 모수를 포함하는 신뢰구간의 비율
  : 신뢰수준은 높을수록 바람직함
- **구간의 크기**
  : 신뢰구간의 폭을 의미함
  : 구간의 크기가 작을 수록 더욱 정교한 구간 추정이 됨

구간 추정

- 신뢰수준은 높을수록, 구간의 크기는 작을 수록 바람직함.
- 그러나 신뢰수준과 구간의 크기는 서로 상충됨
  : 신뢰수준을 높이면 구간의 크기가 길어지고, 반대로 구간의 크기를 줄이면
  신뢰수준은 떨어지게 됨.

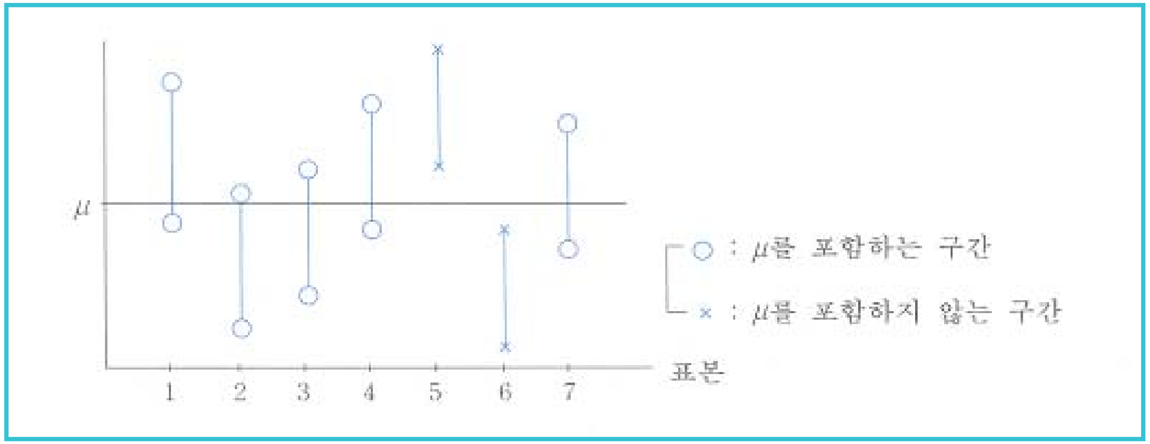
- 따라서 신뢰수준을 고정시켜 놓은 상태에서, 신뢰수준을 만족하면서 구간의 길이가 가장 작은 신뢰구간을 도출하는 방법을 사용함

- 즉, $P(a < \mu < b) = 1 - \alpha$ 를 만족시키는 수많은 (a,b) 구간들 중, 가장 길이가 짧은 구간을 신뢰구간으로 선택하게 됨

- 신뢰구간은 점추정량이 지닌 오차의 크기를 반영한 것
- 따라서 신뢰구간은 점추정량을 기준으로 오차의 크기를 가감한 형태를 지니게 된다

신뢰수준의 의미

- ‘모수가 신뢰구간에 포함될 비율’ vs ‘신뢰구간이 모수를 포함할 비율’ ?
- 모수는 우리가 모르지만 불변의 값인 반면, 신뢰구간은 표본에 따라서 달라질 수 있음
  - 그림 : 표본에 따라서 신뢰구간이 달라지는 예



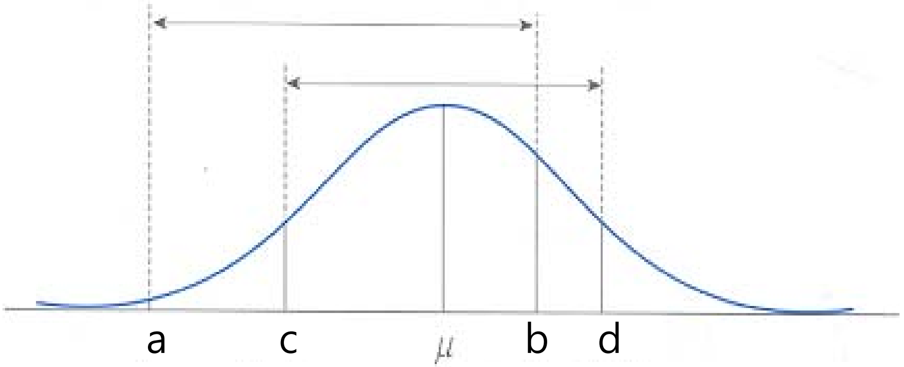
- 따라서 ‘모수가 신뢰구간에 포함될 비율’이 아니라, 여러 번 표본을 뽑아 신뢰구간을 구하였을 때 ‘**신뢰구간이 모수를 포함할 비율**’을 고려해야 하는 것
- 그리고 그 비율이 바로 신뢰수준을 의미하는 것!!!

모평균에 대한 구간 추정

- 모평균에 대한 점추정량 = $\bar{X}$
- 모평균에 대한 $100(1-\alpha)\%$ 신뢰구간 (L , U)

  $\Leftrightarrow$ $P(L<\mu<U)=1-\alpha$ 를 만족하는 구간 중, 가장 길이가 짧은 (L , U)

- 표본평균의 표집분포 : $\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$



: (a,b)와 (c,d) 모두 주어진 신뢰도를 만족하지만 (c,d)의 구간의 크기가 더 작다.
따라서 **좌우 대칭**의 형태로 신뢰구간을 만드는 것이 최선임을 알 수 있다.

모평균에 대한 구간 추정

- 정규분포에서 표본평균의 분포로부터 다음이 성립한다

$$1-\alpha = P\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2} \right)$$

$$= P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \right)$$

- 모평균에 대한 $100(1-\alpha)\%$ 신뢰구간

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = \left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \right)$$

  - **표준오차 (standard error)** : $SE(\bar{X}) = \sqrt{Var(\bar{X})} = \sigma/\sqrt{n}$
  - **오차한계 (margin of error)** = $z_{\alpha/2}\,\sigma/\sqrt{n}$

- 예 : $\alpha = 0.05$ 일 때, 모평균의 95% 신뢰구간

$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}} = \left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1.96\frac{\sigma}{\sqrt{n}} \right)$$

신뢰구간에 대한 이해

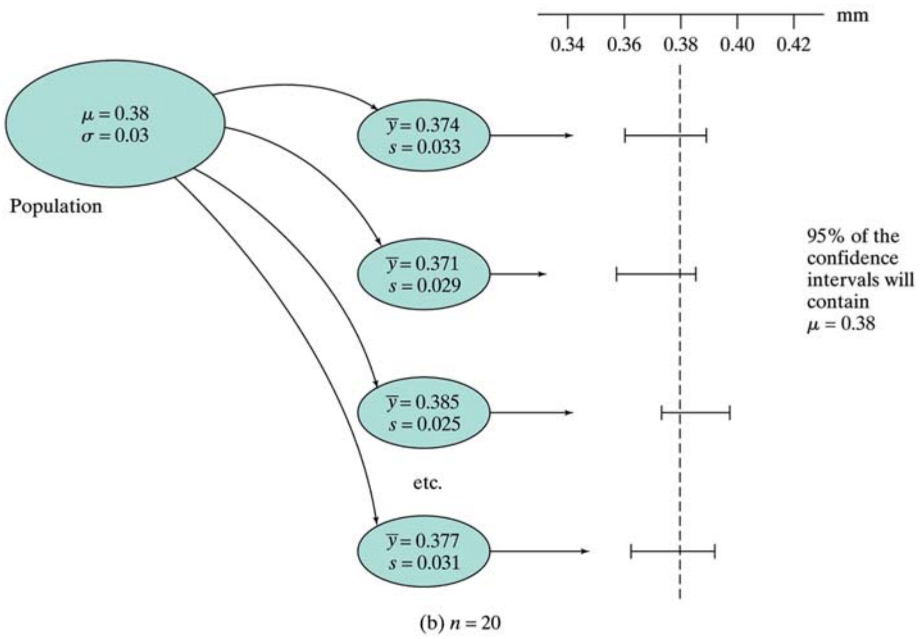
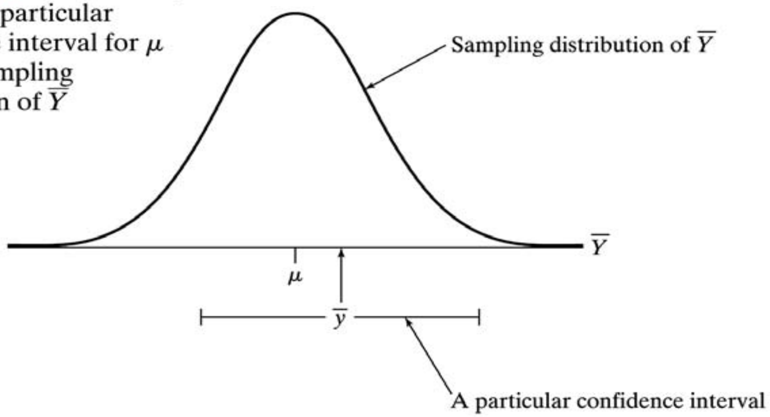


Figure 6.3.7 Relationship between a particular confidence interval for μ and the sampling distribution of $\bar{Y}$



모평균에 대한 구간 추정 : 예제

- 예제 5.1 : 중앙아메리카의 저소득층 원주민을 대상으로 49명의 표본조사를 한 결과 혈청내의 콜레스테롤양의 평균이 157.02(mg/L)였다고 한다. 이들 원주민 전체에서 혈청 내의 콜레스테롤 양이 정규분포를 하고, 모표준편차가 30(mg/L) 라고 할 때, 원주민 전체에서 혈청 내의 콜레스테롤 양의 평균에 대한 95% 신뢰구간을 구하여라.

모평균 추정에서 표본크기의 결정

- 모표준편차를 아는 경우, $100(1-\alpha)\%$ 신뢰수준에서의 오차한계

$$d = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- $100(1-\alpha)\%$ 오차한계를 d 이하로 만족하는 표본의 크기

$$n \ge \left(z_{\alpha/2}\frac{\sigma}{d}\right)^2$$

- 예제 5.2 : 예제 5.1에서 모평균에 대한 추정의 95% 오차한계를 5(mg/L)이하로 하기를 원한다면, 몇 명의 저소득층 원주민을 표본조사 했어야 하는가?

통계적 추론 : 가설검정

- 통계적 추론(statistical inferential)
  : 표본 데이터를 이용하여 모집단의 특성값인 모수에 대해 추측하는 것

- 추정(estimation) : 데이터를 이용하여 모수에 대한 추측값을 구하는 것
  - 예 : 표본가구의 소득을 조사하여 우리나라 가구당 평균 소득을 추측
  - 예 : 일정수의 유권자들을 추출하여 특정 후보자의 지지율을 예측하는 것
  - 점추정(point estimation)
  - 구간추정(interval estimation)

- 가설검정(hypothesis testing) : 모수에 대한 특정 주장을 받아들일지 여부를 표본 데이터를 이용하여 판단하는 것
  - 예 : 2004년도 고등학생들의 평균 키는 10년전 학생들의 평균키보다 커졌는가?
  - 예 : 새로 개발한 진통제는 기존의 진통제보다 효능이 뛰어나다고 할 수 있는가?

유의성 검정

- 예 5.4 : 한 과학자가 유전공학적 기법을 사용하여 변종시킨 완두콩에 대하여는 멘델의 이론과는 다른 이론이 성립하지 않는가라는 착상을 하게 되었다. 따라서 변종된 완두콩들을 사용하여 교배 실험을 한 결과, 2세대에서 860개의 둥글고 노란 완두, 340개의 둥글고 녹색인 완두, 240개의 모나고 노란 완두와 160개의 모나고 녹색인 완두를 얻었다. 멘델의 이론에 의하면 이들 네 종류의 비는 9:3:3:1이어야 하는데, 과연 이 과학자의 실험결과는 새로운 이론의 존재를 강하게 시사하는 것일까?

- 과학자의 실험의 목적은 멘델의 이론을 하나의 가설(hypothesis)로 간주하고, 이를 부정할 수 있는 반증이 이 실험의 결과로서 뚜렷이 나타나는가를 확인하고자 하는 것

- 귀무가설 (null hypothesis) : 반대의 증거를 찾기 위하여 상정된 가설
- 대립가설 (alternative hypothesis) : 귀무가설의 대안으로서 상정되는 가설

- **유의성 검증 (test of significance)**

  : 귀무가설에 대한 반증의 강도를 제공하는 과정

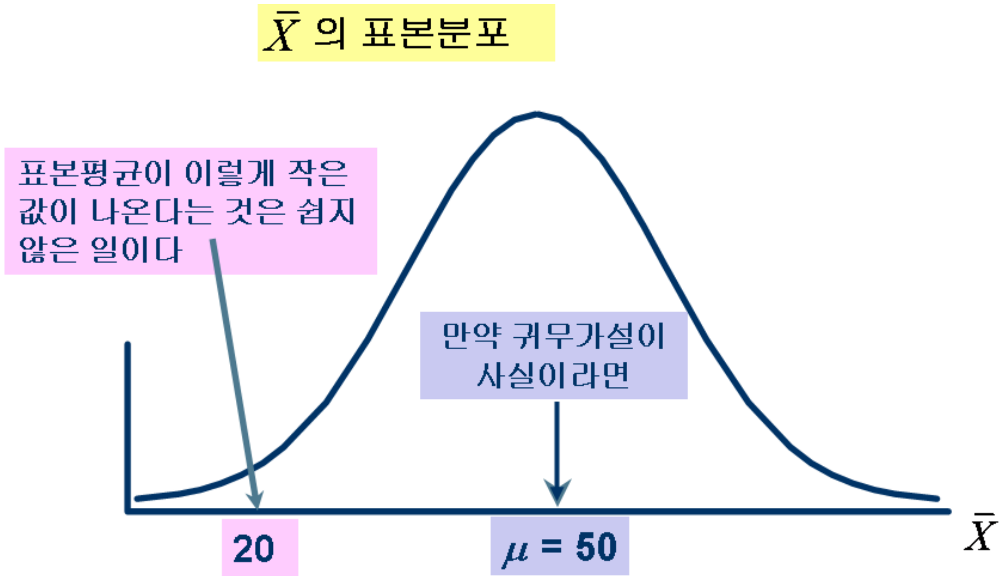
  : 귀무가설이 옳다면 일어나기 매우 어려운 현상을 관측한 경우라면, 귀무가설을 대신 할 새로운 가설에 대한 탐구가 필요할 것이라는 논리

가설 검정

- 현재 모집단의 평균 연령은 50세인 것으로 알려져 있다.
- 하지만 나의 주장은 모집단의 평균 연령은 50세 미만이라는 것이다.
- 나의 주장을 뒷받침하기 위해 데이터를 수집한다
- 표본을 뽑아서 평균 연령을 구해본 결과, 평균 연령은 20세로 나타났다.

- Q : 만약 현재 알려진 사실대로, 모집단의 평균 연령이 50세라고 하자. 그렇다면 그러한 모집단에서 뽑힌 표본들의 평균 연령이 20세가 되는 것이 흔한 일인가?
- A : 그렇지 않다!! 모집단의 평균 연령이 50세라면, 표본 평균 역시 평균이 50세인 분포를 따르게 된다. 하지만 그러한 분포에서 평균값이 20세가 나오는 일은 쉽지 않은 일이기 때문이다.

- 따라서 모집단의 평균 연령이 50세라는 것은 기각 된다.



가설 검정의 단계

모집단의 특성에 대한 주장, 즉 가설을 설정

⇩

검정의 유의수준 α 를 정함

⇩

표본을 추출하여 검정통계량을 계산

⇩

검정통계량에 대한 p 값을 구하여 가설에 대한 결론 도출
or 기각역에 속하는지 검토하여 가설에 대한 결론 도출

통계적 가설

- **통계적가설(statistical hypothesis)**

  : 모집단의 특성 또는 모수에 대한 대립되는 두 가지 주장에 대하여 통계적으로 다루기 편리하도록 정리한 것

- **귀무가설(null hypothesis, H_0)**

  : 알려진 사실, 현재 믿어지고 있는 사실을 나타내는 가설

  : 대립가설에 대한 확실한 근거가 없을 때 받아들이는 가설

  : 모집단의 모수를 하나의 값 또는 구간 등으로 표시

- **대립가설(alternative hypothesis, H_1)**

  : 연구자의 주장을 나타내는 가설

  : 표본으로부터 입증하고자 하는 가설

  : 귀무가설에서 제시하는 모수의 값을 제외한 나머지 영역에서 모수의 값을 정의

통계적 가설

- 예 5.5 : 건물의 소화용으로 사용되는 살수장치를 생산하는 회사에서 이 살수장치가 실내온도 55 도에서 작동되도록 제조하려고 한다. 제조공정의 이상여부를 판단하기 위하여 생산품 중에서 표본을 추출하여 작동 시작 온도를 조사하고자 한다. 이러한 조사에서 공정에 이상이 있다는 증거가 뚜렷하면 후속되는 기술적 조치를 하려고 한다. 이러한 경우, 이 공정에 이상이 있다는 것은 $\mu < 55$ 이거나 $\mu > 55$ 인 경우, 즉 $\mu \neq 55$ 인 경우이다.

  또한, 조사의 목적이 공정이 정상이 아니라는 뚜렷한 증거가 있는가를 알아보기 위한것이므로, 귀무가설과 대립가설은 각각 다음과 같다.

$$H_0 : \mu = 55 \qquad H_1 : \mu \neq 55$$

- 한 페인트 제조회사에서 생산하는 특수 페인트의 건조시간은 평균 75분, 표준편차 9.4분인것으로 알려져있다. 건조시간을 줄이기 위하여 새로운 첨가제를 개발한 연구소에서는 실제로 평균 건조시간이 줄어들 것이라고 주장한다. 이에 따라 새로운 첨가제를 사용한 시제품을 생산하여 건조시간을 조사하고 실제 건조시간이 단축된다는 뚜렷한 증거가 있으면 이를 전체 공정에 적용하려고 한다. 이러한 경우의 귀무가설과 대립가설은 다음과 같다.

$$H_0 : \mu = 75 \qquad H_1 : \mu < 75$$

검정 통계량

- **검정통계량(test statistic)**

  : 가설 검정에 쓰이는 통계량

  : 검정하기에 편리한 형태로 표본자료를 바꾸어 놓은 것

  : 일반적으로 검정하려는 모수의 점추정량이 되기도 하고, 이 점추정량을 표준화 한 것을 사용하기도 함

- 귀무가설이 사실이라는 가정 하에, 표본데이터가 얻어 질 가능성을 살펴야 함

  ⇒ 점추정량의 분포가 필요함

- 예 5.7 : 예 5.5에서 귀무가설과 대립가설이 각각 $H_0 : \mu = 55 \quad vs \quad H_1 : \mu \neq 55$ 인 경우에는 모평균에 대한 추정량인 표본평균을 사용하여 유의성검증에서의 증거를 찾게 된다.

검정 오류

- 검정오류(test error) : 가설을 채택하거나 기각할 때 확률적으로 틀릴 가능성

  - **제1종 오류(type Ⅰ error: α)**

    : 귀무가설이 맞는데도 잘못하여 이를 기각하고 대립가설을 채택할 확률

    : 검정의 **유의수준 significance level**)

$$P\left[H_0 \text{ reject} \mid H_0 \text{ true}\right] = \alpha$$

  - **제2종 오류(type Ⅱ error: β)**

    : 대립가설이 사실임에도 불구하고 귀무가설을 채택하게 되는 확률

$$P\left[H_0 \text{ accept} \mid H_a \text{ true}\right] = \beta$$

실제현상 / 검정결과	H_0 true	H_1 true
H_0 채택	옳은 결정	**제 2종 오류**
H_1 채택	**제 1종 오류**	옳은 결정

검정오류

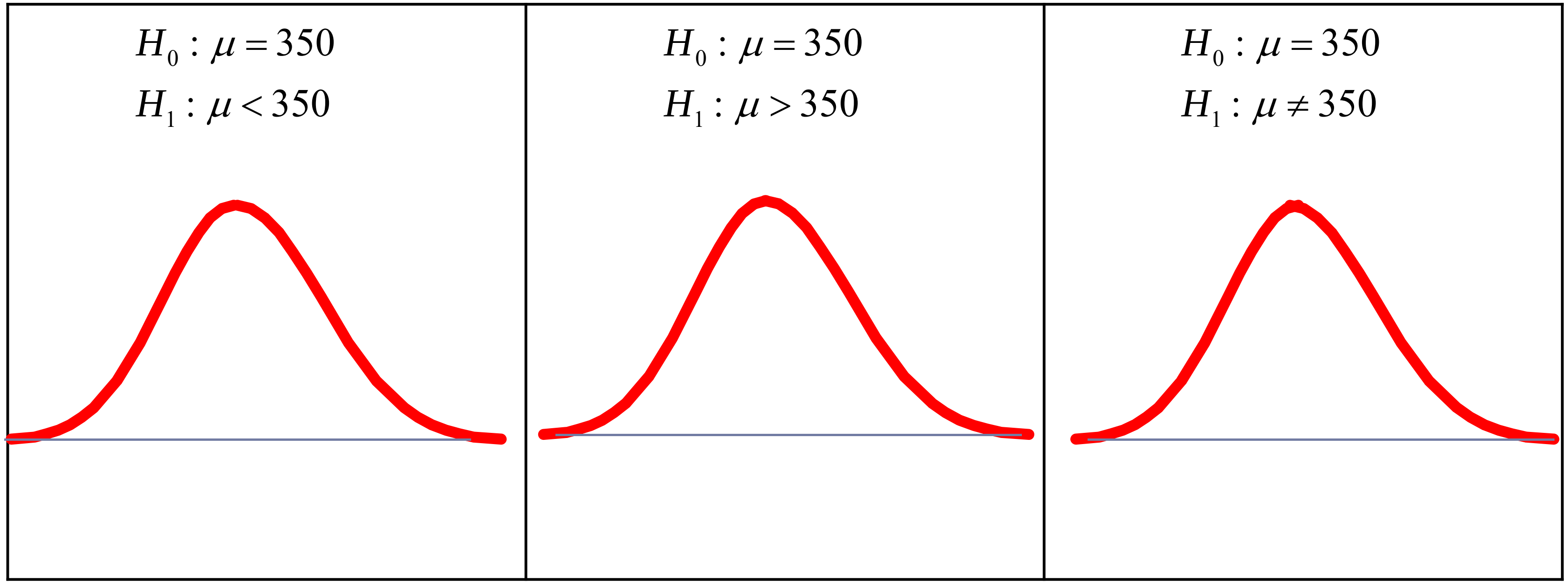
- 두 가지 검정오류를 동시에 작게 하는 것이 최선이지만, 제 1종 오류와 제 2종 오류는 서로 상충되는 관계에 있음
  : 제 1종 오류를 작게 하면 제 2종 오류가 증가, 반대로 제 2종 오류를 작게 하면 제 1종 오류가 증가하게 됨

- 따라서 둘 중 더욱 심각한 오류라고 여겨지는 제 1종 오류를 고정시켜 놓은 상태에서 제 2종 오류를 최소화 하는 검정 방법을 사용하게 됨
  : 주어진 α 에 대하여 $(1-\beta)$ 를 크게 하는 검정을 선호

- 유의수준 α 의 결정
  : 자료와 관련된 학문 분야에서 이미 사용되고 있는 유의수준 또는 선행 연구 논문 등을 참고하여 유의수준을 결정한다.

결론 도출 : 기각역

- **기각역(rejection region) :** 귀무가설을 기각하게 되는 영역
  - 검정통계량이 기각역 안에 속하게 되면 귀무가설을 기각하게 됨
  - 기각역의 넓이 = 유의수준
  - 따라서, 주어진 유의수준에 따라 기각역을 결정하고 이 영역 안에 검정통계량이 포함되면 귀무가설을 기각하게 됨
  - 기각역의 방향은 대립가설에 따라 결정이 됨



결론 도출 : p-value

- 기각역 방법에 의한 결론 도출의 경우, 유의수준이 달라지면 새로운 기각역을 반복해서 설정해야 함
- 또한 단순 결론 이외의 기각의 강도 등을 알 수 없는 단점이 있음

- **P-value (significance probability)**
  - 귀무가설이 사실일 때, 표본이 대립가설 방향으로 검정통계량의 값보다 더 어긋날 확률
  - 귀무가설이 참이라는 가정하에 현재 관측된 데이터 이상으로 대립가설을 입증하는 데이터가 관찰 될 확률
  - 즉, p-value가 작다는 것은 **귀무가설 하에서 현재 관측된 표본 데이터보다 더 강력하게 귀무가설을 기각할 데이터가 나올 확률이 작다**는 것이므로, 지금으로서도 충분히 귀무가설을 기각할 수 있다는 것을 의미

- 귀무가설에 대한 반증의 강도에 대하여 기준값을 미리 정해놓고 유의확률을 그 기준값과 비교하여 결론을 도출하게 됨.

결론 도출 : p-value

- P-value에 의한 결론 도출

$$\text{Reject } H_0 \text{ if } \text{p-value} < \alpha$$

: 유의확률 (P값)이 유의수준 이하이면, 조사결과가 **유의수준 α 에서 통계적으로 유의하다**고 하며, 이는 귀무가설에 대한 반증의 강도가 지정된 수준보다 강함을 뜻한다.

: 조사 결과보다 더욱 귀무가설에 대한 반증이 강하게 나타날 수 있는 기회를 유의수준 이하로 요구하는 것.

- (참고) 가설 검정의 결과

모평균에 대한 추론 : 모분산을 아는 경우

1. 가설을 설정한다

$$H_0 : \mu = \mu_0 \quad H_1 : \mu < \mu_0$$ or $$H_0 : \mu = \mu_0 \quad H_1 : \mu > \mu_0$$ or $$H_0 : \mu = \mu_0 \quad H_1 : \mu \neq \mu_0$$

2. 검정의 유의수준을 결정한다

3. 적절한 검정통계량을 결정한다

- 모집단이 정규분포를 따르는 경우

$$\frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$$

- 모집단이 정규분포가 아니지만 표본의 수가 충분히 큰 경우

$$\frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \dot{\sim} N(0,1)$$

4. 기각역 방법 또는 p-value 방법을 이용하여 결론을 도출한다

모평균에 대한 추론 : 모분산을 아는 경우

- 예 5.10 : 새로운 첨가제가 페인트의 건조시간을 단축시키는가를 확인하기 위하여 가설

$$H_0 : \mu = 75 \qquad H_1 : \mu < 75$$

  을 세우고, 시제품 100개를 생산하여 건조시간을 조사한 결과 평균 건조시간이 73.5분 이었다고 하자. 모표준편차는 9.4분으로 주어졌을 때 유의수준 5%에서 가설검정을 하시오.

모평균에 대한 추론 : 모분산을 아는 경우

- 예제 5.7 : 건물의 소화용으로 사용되는 살수장치를 생산하는 회사에서 이 살수장치가 실내온도 55도에서 작동되도록 제조하려고 한다. 제조공정의 이상여부를 판단하기 위하여 생산품 중에서 9개의 표본을 추출하여 작동 시작 온도를 조사한 결과 그 평균온도가 55.63이었다. 전체 제품의 작동 시작 온도가 표준편차 0.9인 정규분포를 따른다고 할 때, 공정의 이상여부를 확인하기 위한 가설

$$H_0 : \mu = 55 \qquad H_1 : \mu \neq 55$$

  을 유의수준 1%에서 검정하여라

모평균에 대한 추론 : 모분산을 아는 경우

- 예제 5.6 : 한 제약회사에서 생산하고 있는 기존의 진통제는 진통효과가 나타나는 시간이 평균 30분, 표준편차 5분인 정규분포를 따르는 것으로 알려져 있다. 이 회사에서 새롭게 개발한 진통제의 효과를 확인하기 위하여 50명의 환자를 랜덤추출하여 새로운 진통제에 의해 그 효과가 나타나는 시간을 관측한 결과, 평균이 28.5분이었다. 새로운 진통제의 진통효과가 더 빠르다고 말할 수 있는지 유의수준 5%에서 검정하여라.

검정력

- 검정력

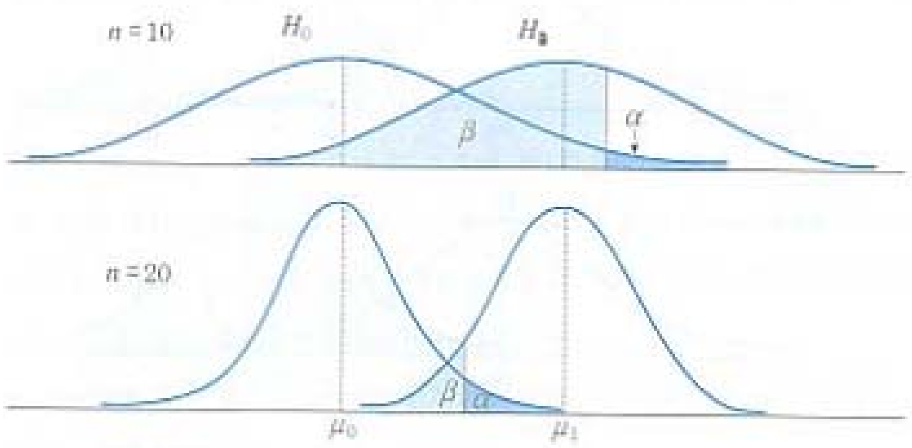
: $H_0:\theta=\theta_0$ vs $H_1:\theta=\theta_1\neq\theta_0$ 일 때,

$$\gamma(\theta_1)=P(X\in R\,|\,\theta=\theta_1)$$

where X = 검정통계량, R = 기각역

- 표본수 역시 검정오류에 영향을 미침

: 표본수가 커지면 일반적으로 모수 추정량의 분산이 작아지기 때문에 고정된 유의수준에 따른 제 2종 오류가 감소하게 된다



검정력 : 예제

- 예 5.13 : 새로운 진통제의 효과가 나타나는 평균시간 μ가 28분이면 이는 괄목할만한 개선이라고 한다. 실제로 $\mu = 28$ 일 때, 검정력은?

제 2종 오류에 근거한 표본의 크기 : 예제

- 예제 5.8 : 예제 5.6에서 실제로 $\mu = 28$ 일 때, 제 2종의 오류를 범할 확률이 $\beta = 0.1$ 이하가 되도록 하며, 유의수준이 $\alpha = 0.05$ 인 검정법을 사용하려고 한다. 이 때, 요구되는 표본의 크기를 구하여라