(2) 표본오차

✦ (표본평균) 표본분포의 평균

- 모집단의 평균과 동일<(식 8-9) 참조>

✦ (표본평균) 표본분포의 표준편차(=표본오차=표준오차=오차한계)

- 모표준편차를 표본의 크기(사례수)의 제곱근으로 나눈 값<(식 8-10) 참조>
- 표본추출이 완벽하지 못하기 때문에 특정 표본평균이 모평균으로부터 떨어진 정도

표본평균 표본분포의 평균 $= \sum \bar{X}_i P(\bar{X}_i) = \mu$ (식 8-9)

표본평균 표본본포의 표준편차 (=표준오차) $= \sigma/\sqrt{n} = s/\sqrt{n}$ (식 8-10)

• 출처 : 신민철 저 "경영경제 통계학의 기초, 창민출판사, 2010년.

✦ 모집단의 표준편차를 알아야 표본오차가 계산됨

➪ 통계학자들은, '하나의 표본에서 계산된 표본표준편차가 표본오차의 근사치'라고 함을 밝혀냄

➪ 결국, 표본평균, 표본표준편차 및 정규분포를 따르는 표본분포의 특징을 알게 되면 표본오차(=표준오차)를 알게 됨

➪ 표본평균도 알고, 표본오차도 알게 되므로, 표본평균이 모수와 얼마나 유사한지도 알게 됨 (☞ 모수 = 표본통계량 – 표본오차)

4) 표본분포의 기타 유형

✦ 표본분포가 정규분포를 따르지 않을 때에도 추리통계 수행 가능

예) 표본의 사례 수(즉, 표본의 크기)가 매우 작은 경우

(1) *t* 분포

✦ *t* 분포(*t* -distribution)

- 표본의 크기가 30 이하인 경우에 사용가능
- 영국의 Gosset이 1908년에 학생(student)이라는 익명으로 소표본(N <30) 에서의 표본분포는 정규분포를 따르지 않는다는 것을 발표하여 알려지게 됨

➪ *t* 분포는 '학생의 *t* 분포(student's *t* -distribution)'라고 불리기도 함

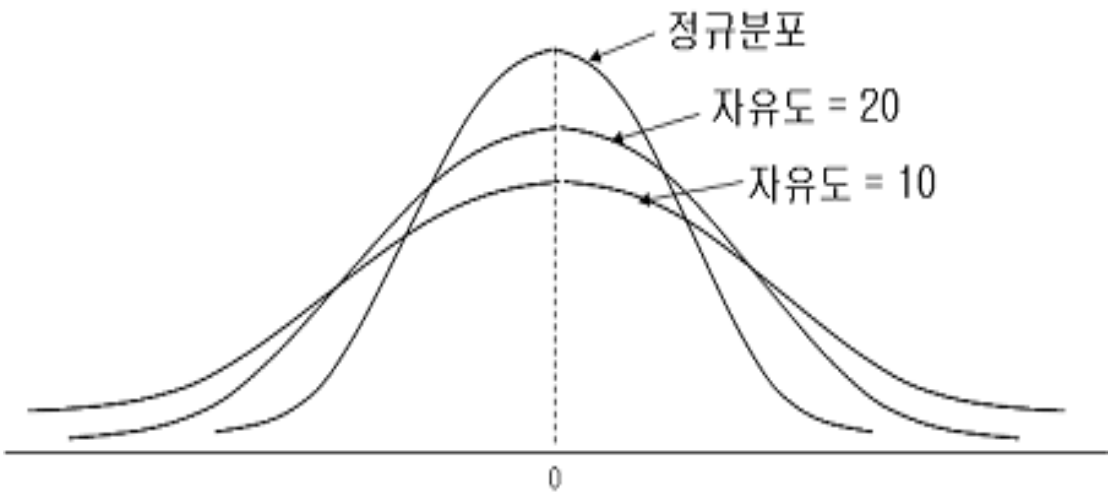
✦ *t* 분포란 표준정규분포처럼 단일분포가 아니고 표본의 크기에 따라 표본분포(sampling distribution)가 변하는 특징을 보유

➪ 즉, 자유도에 따라 표본분포곡선의 모습이 변화([그림 8-5] 참조).

✦ **자유도**(degrees of freedom: **df**)

- *t* 분포뿐만 아니라 다른 유형의 통계검정에도 자주 사용되는 중요한 통계개념
- 개념적으로 자유도란 *표본분포(sampling distribution)를 구성하기 위해 자유롭게 반복해서 추출할 수 있는 표본(repeated random sample)의 수*
- 구체적으로 자유도는 표본크기에서 표본에 부여되는 제약조건의 수를 차감해서 계산 (p.236 예)

[그림 8-5] 표본의 크기와 *t* 분포



* 두 개의 표본간에 존재하는 평균의 차이검정에 사용

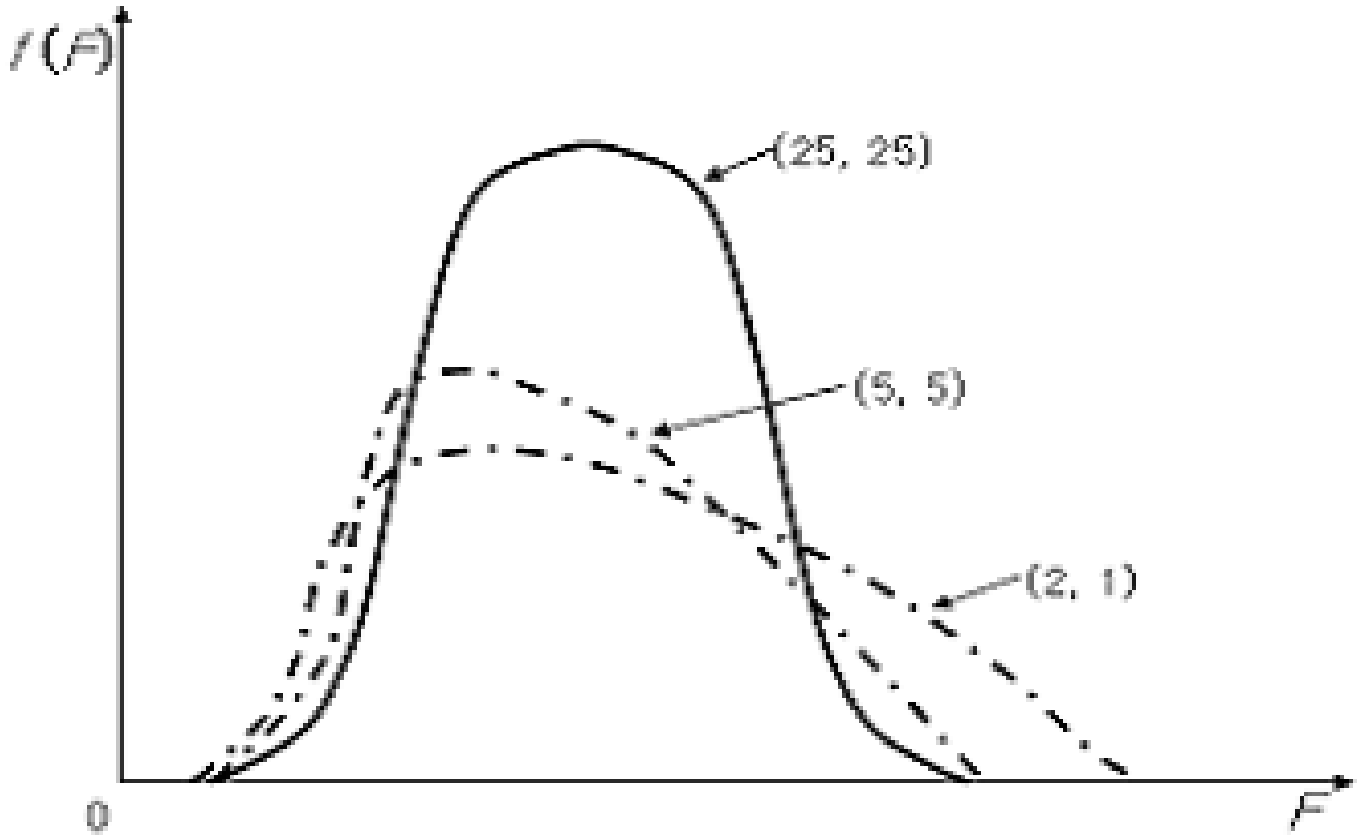
-> 자유도 : $(n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$

(2) *F* 분포

- *F* 분포(*F*-distribution)
  - *F* 값을 나타내는 분자와 분모 각각의 자유도에 의하여 규정되는 분포
  - *F* 분포의 표본분포(sampling distribution)는 자유도의 값에 따라 다양한 수의 분포 가능
  - 비교집단이 2개보다 큰 경우(즉, 3개 이상 집단간의 비교)에도 집단간의 차이를 설명할 수 있는 표본분포
  - 분산분석(analysis of variance)에서 대표적으로 사용되고 있는 분포 (종종 분산분석을 **F검정**이라고 부르기도 함)

- *F* 값(*F* ratio)
  - 집단간 분산의 추정값을 집단내 분산의 추정값으로 나눈 것

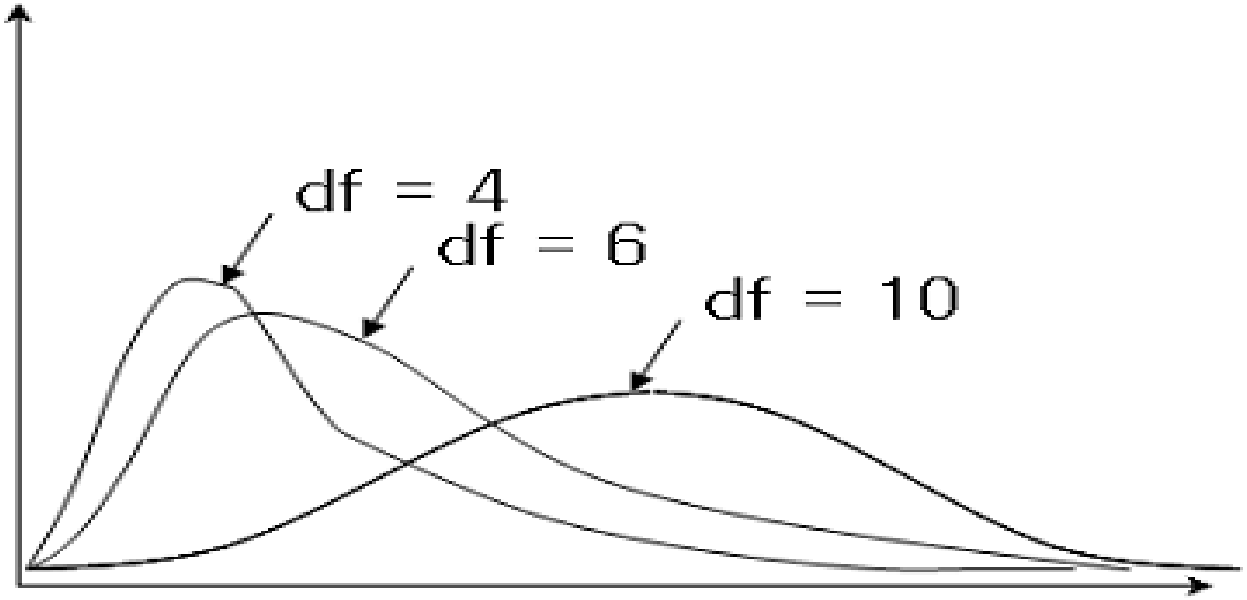
[그림 8-6] 자유도에 따른 *F* 분포곡선



(3) χ^2 분포

- χ^2 분포 (Pearson, 1900)
  - χ^2 값에 따르는 표본분포(sampling distribution)
  - *t* 분포와 마찬가지로 자유도에 의하여 구체적 분포가 결정됨
  - 일반적으로 크기 *n* 인 하나의 표본에서 χ^2 자유도는 *n* -1로 결정됨

[그림 8-7] χ^2 분포곡선



✦ χ^2 값

◆ 실제로 관찰된 빈도가 기대한 빈도와 얼마나 가까운가를 검정하는 도구

⇨ 명목척도로 측정된 두 변수간의 상관관계 검정 시 χ^2 검정 이용

⇨ χ^2 검정은 χ^2 분포를 사용하여 가설의 진위를 판단

χ^2 값 계산 (식 8-11)

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

단, $E_{ij} = ij$ 번째 칸(cell)의 기대빈도

$O_{ij} = ij$ 번째 칸의 실제 빈도

제2절 추리통계분석 결과는 어찌 해석할 것인가?

✦ 추리통계분석의 목적

- 표본통계량(sample statistic)으로 모수(population parameter)를 추정하는 것

✦ 추리통계분석의 구분

- 추정(parameter estimation)과 가설검정(hypothesis testing)

⇨ 추정이나 가설검정의 구체적 예는 매우 다양

⇨ 통계적 추론의 정확성을 판단해 줄 공통의 도구 필요

제2절 추리통계분석 결과는 어찌 해석할 것인가?

1. 추 정

✦ 50명 학생들의 통계학 평균점수가 100점 만점 기준 50점(= $\bar{X}$)일 때, 6,000명 모집단의 평균점수(= μ_X) 추정

⇨ 모집단 평균점수(= μ_X) 추측의 예

(a) 모수는 50점일 것이다. ⇨ 이상적 추론?

(b) 모수는 25점에서 75점 사이일 것이다. ⇨ 상대적으로 신뢰성 있는 추론?

(c) 모수는 0점에서 100점 사이일 것이다. ⇨ 100% 확실한 추론?

⇨ 어떻게 추정의 정확성을 증가(or 오류가능성을 감소)시킬 것인가?

제2절 추리통계분석 결과는 어찌 해석할 것인가?

✦ 추정의 구분

◆ 점추정과 구간추정

- ▫ **점추정**(point estimation)

  = 하나의 값으로 모수를 추측하는 것

  예) (a)

- ▫ 구간추정(interval estimation)

  = 모수가 포함되는 구간을 추측하는 것

  예) (b), (c)

제2절 추리통계분석 결과는 어찌 해석할 것인가?

1) 점추정

✦ 추정량과 추정치

- 추정량(estimator)

  = 모수를 추정하기 위해 이용하는 표본통계량, 확률변수임

- 추정치/추정값(estimate)

  = 모수를 추정한 구체적인 값

✦ **바람직한 점추정량**

◆ 개념적으로는 추정량 $\hat{\theta}$과 모수 θ의 차이, 즉 $(\hat{\theta}-\theta)$를 최소화하는 추정량

➪ 구체적으로는 평균제곱오차[(식 8-12)]가 작은 추정량

$$MSE(\theta) = E[(\hat{\theta} - \theta)^2] = Var(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2 \qquad (식\ 8\text{-}12)$$

제2절 추리통계분석 결과는 어찌 해석할 것인가?

- **불편성**(unbiasedness)
  - $\hat{\theta}$의 평균[$=E(\hat{\theta})$]이 θ와 동일한 경우
    - → 불편추정량(unbiased estimator)

  예) 표본평균 $\bar{X}$는 모집단 평균($=\mu$)의 불편추정량

- **효율성**(efficiency)
  - $\hat{\theta}$ 의 분산이 최소인 특성
    - → 최소분산 불편추정량(minimum variance unbiased estimator)
      (가장 효율적인 불편추정량)

- **일치성**(consistency)
  - 표본의 크기($=n$)가 한없이 증가할 경우 추정량 $\hat{\theta}$이 모수 θ에 한 없이 가까워지는 특성

* 충분성 : 표본자료가 내포하고 있는 모수에 대한 정보와 지식을 포괄적으로 요약해주는 추정량, (표본평균 $\bar{X}$, *중앙값*)

※ 평균제곱오차(mean squared error: MSE)

$$MSE(\theta) = E[(\hat{\theta} - \theta)^2] = Var(\hat{\theta}) + [E(\theta) - \hat{\theta}]^2 \qquad (식\ 8\text{-}12)$$

- 점추정의 상대적 신뢰도 정보는 없음

⇨ $E(\bar{X}) = \mu$ 이지만 $\bar{X} = \mu$ 이 얼마나 신뢰할 수 있는지 판단할 수 없음

기대치: 같은 일이 무한히 반복될 때, 해당 확률변수의 평균의 의미

2) 구간추정

✦ 구간추정(interval estimation)

- 모수가 존재할 구간을 제시
- 이 구간이 얼마나 믿을만한 지에 대한 정보도 제시

※ 점추정량

- '평균적'으로는 모수를 대변
- 점추정치가 상대적으로 얼마나 믿을 만한 지에 대한 정보가 없음

제2절 추리통계분석 결과는 어찌 해석할 것인가?

- **유의수준**(significance level)
  - α
  - 모수가 특정 구간 내에 포함되지 않을 가능성

- **신뢰수준/신뢰도**(confidence level)
  - (1- α)
  - 모수가 특정 구간 내에 포함될 가능성

- 100(1- α)%의 신뢰구간(confidence interval)
  - 모수 θ가 포함될 가능성이 100(1- α)% 인 구간

    예) 관행적으로 95%(즉 α =0.05) 혹은 99%(즉 α =0.01)의 신뢰구간을 사용

- 표본분포를 포함한 모든 정규분포는 표준정규분포([그림 8-8] 참조)로 변환 가능

예1) 모수 중심 좌우 $1\sigma_{\bar{X}}$ 내에는 68.26%,

좌우 $2\sigma_{\bar{X}}$ 내에는 95.44%,

좌우 $3\sigma_{\bar{X}}$ 내에는 99.74%,

좌우 $4\sigma_{\bar{X}}$ 내에는 거의 100%의 표본통계량(즉, 표본평균) 값 위치

• 출처 : 신민철 저 “경영경제 통계학의 기초, 창민출판사, 2010년.