

8. 분산 분석

8.1 분산 분석의 이해

분산 분석 (ANOVA, Analysis of Variance)

관심 집단이 3개 이상이 되면 기존의 검정 통계량을 통해 평균이 동일한지 여부를 판단할 수 없음

정규 분포를 따르는 검정 통계량을 만들어 낼 수 없기 때문

따라서 3개 이상 집단간 평균의 동일성을 판단하고자 할 때, 분산들의 상대적 크기를 비교하면서 이를 검정

8.1 분산 분석의 이해

분산 분석 용어

1. 요인 (Factor)

: 결과에 영향을 주는 **원인에 해당하는 변수** ex) 학습시간에 영향을 미치는 원인은 학년 (요인이 학년)

2. 수준 (Level)

: **요인을 나누는 기준** ex) 학년을 1학년, 2학년, 3학년 등으로 나눔

3. 반응 (Response)

: 실험 또는 조사의 **결과로 얻어지는 양적 자료**를 의미

요인에 대한 수준은 질적 자료로 그룹을 구분하는 단위로 이해

반응은 이러한 그룹별로 시행되는 실험 또는 조사에 의해 얻어지는 양적 자료

8.2 일원배치 분산 분석의 절차

일원배치 분산분석 (One-Way ANOVA)

그룹	1	2	...	k
자료	Y_{11}	Y_{21}	...	Y_{k1}
	Y_{12}	Y_{22}	...	Y_{k2}

	Y_{1n}	Y_{2n}	...	Y_{kn}

1. 등분산성의 가정 (각 집단의 분산이 동일)

: 각 집단의 분산이 동일 ($\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$)

2. 정규성의 가정 (관찰값들이 모두 정규분포를 따름)

: $Y_{k1}, Y_{k2}, \dots, Y_{kn} \sim N(\mu_1, \sigma^2)$

3. 독립성의 가정

: 각 집단별로 관측된 자료는 서로 독립임을 가정

8.2 일원배치 분산 분석의 절차

일원배치 분산 분석의 검정 절차

가설의 설정	$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = 0 \text{ vs } H_1 : \text{Not } H_0$
모형의 분해	$(Y_{ij} - \bar{Y}_{..}) = (\bar{y}_{i.} - \bar{Y}_{..}) + (Y_{ij} - \bar{y}_{i.})$
제곱합의 분해 (분산의 분해)	$SST = SS_{Trt} + SSE$
자유도 계산	$N - 1 = (k - 1) + (N - k)$
제곱 평균의 계산	$MS_{Trt} = \frac{SS_{Trt}}{k - 1}, \quad MSE = \frac{SSE}{N - k}$
검정 통계량의 계산과 의사결정	$F = \frac{MS_{Trt}}{MSE}$