분산분석

서론

- 분산분석(ANOVA: Analysis of Variance)

- 일원배치 분산분석법(One-Way ANOVA)

일원배치 분산분석법(One-Way ANOVA)

	관측자료	평균	제곱합
처리 1	$y_{11}, y_{12}, \cdots, y_{1n_1}$	$\bar{y}_1$	$\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2$
처리 2	$y_{21}, y_{22}, \cdots, y_{2n_2}$	$\bar{y}_2$	$\sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
처리 k	$y_{k1}, y_{k2}, \cdots, y_{kn_k}$	$\bar{y}_k$	$\sum_{j=1}^{n_k} (y_{kj} - \bar{y}_k)^2$
총평균 $\bar{y} = \dfrac{\text{관측값들의 총합계}}{\text{총 자료의 수}} = \dfrac{n_1\bar{y}_1 + \cdots + n_k\bar{y}_k}{n_1 + \cdots + n_k}$			

일원배치 분산분석법(One-Way ANOVA)

- 예제. 안경의 표면 손상을 방지하려고 네 종류 A, B, C, D 의 코팅처리에 대하여 표면보호에 얼마나 효과가 있는지를 비교하여 보았다. 다음의 표는 네 개의 코팅처리 된 안경에서 표면의 마모도를 측정한 자료이다.

코팅	관측자료	평균	제곱합
A	10, 15, 8, 12, 15	$\bar{y}_1 = 12$	$\sum_{j=1}^{5}(y_{1j} - \bar{y}_1)^2 = 38$
B	14, 18, 21, 15	$\bar{y}_2 = 17$	$\sum_{j=1}^{4}(y_{2j} - \bar{y}_2)^2 = 30$
C	17, 16, 14, 15, 17, 15, 18	$\bar{y}_3 = 16$	$\sum_{j=1}^{7}(y_{3j} - \bar{y}_3)^2 = 12$
D	12, 15, 17, 15, 16, 15	$\bar{y}_4 = 15$	$\sum_{j=1}^{6}(y_{4j} - \bar{y}_4)^2 = 14$
총평균 $\bar{y} = 15$			

일원배치 분산분석법(One-Way ANOVA)

- $(y_{ij} - \bar{y})$: 개개의 관측값의 총평균에 대한 편차
- $(\bar{y}_i - \bar{y})$: 각 코팅처리 간의 평균값의 차이에서 기인하는 부분
- $(y_{ij} - \bar{y}_i)$: 동일한 코팅처리 내에서 발생하는 측정값의 오차에 의한 부분
- $(y_{ij} - \bar{y}) = (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i)$
- $y_{ij} = \bar{y} + (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i)$
- 관측값 = (총평균) + (처리에 의한 편차) + (잔차)

일원배치 분산분석법(One-Way ANOVA)

- 총제곱합(SST: total sum of squares): $SST = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$

- 처리제곱합(SStr: treatment sum of squares): $\mathrm{SStr} = \sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y})^2$

- 오차제곱합(SSE: error sum of squares): $SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$

- $SST = SStr + SSE$

일원배치 분산분석법(One-Way ANOVA)

- $(y_{ij} - \bar{y}) = (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i)$
- $(y_{ij} - \bar{y})^2 = (\bar{y}_i - \bar{y})^2 + (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \bar{y})(y_{ij} - \bar{y}_i)$
- $\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i) = 0$
- $\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2 = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$
- $SST = SStr + SSE$

일원배치 분산분석법(One-Way ANOVA)

- (제곱합의 자유도) = (제곱을 하여 더하는 항의 수) – (각 항들에 의하여 만족되는 선형 제약조건의 수)

- 총제곱합 $\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2$ 의 자유도: $\sum_{i=1}^{k} n_i - 1$
  (제약식: $\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{y}_i\right) = 0$)

- 처리제곱합 $\sum_{i=1}^{k} n_i(\bar{y}_i-\bar{y})^2$ 의 자유도: $k-1$
  (제약식: $\sum_{i=1}^{k} n_i\,(\bar{y}_i-\bar{y}) = 0$)

- 오차제곱합 $\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2$ 의 자유도: $\sum_{i=1}^{k} n_i - k$
  (제약식: $\sum_j\left(y_{ij}-\bar{y}_i\right) = 0 \quad for \quad i = 1,\cdots,k$)

- $\bar{y} = \frac{n_1\bar{y}_1+n_2\bar{y}_2+\cdots+n_k\bar{y}_k}{n_1+n_2+\cdots+n_k}$

일원배치 분산분석법(One-Way ANOVA)

➢ 분산분석표(ANOVA Table)

요인	제곱합	자유도	평균제곱
처리	$SStr=\sum_{i=1}^{k} n_i(\bar{y}_i-\bar{y})^2$	$k-1$	$MStr=\frac{SStr}{k-1}$
오차	$SSE=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2$	$\sum_{i=1}^{k} n_i-k$	$MSE=\frac{SSE}{\sum_{i=1}^{k} n_i-k}$
합계	$SST=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2$	$\sum_{i=1}^{k} n_i-1$	

일원배치 분산분석법(One-Way ANOVA)

- 제곱합의 간편 계산식

$T_i = \sum_{j=1}^{n_i} y_{ij}$: 처리 i 에서의 모든 관측값의 합계

$\mathrm{T} = \sum_{i=1}^{k} T_i = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}$: 모든 관측값의 총계

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{n_i} {y_{ij}}^2 - \frac{T^2}{n} \quad where \quad n = \sum_{i=1}^{k} n_i$$

$$SStr = \sum_{i=1}^{k} \frac{{T_i}^2}{n_i} - \frac{T^2}{n}$$

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} {y_{ij}}^2 - \sum_{i=1}^{k} \frac{{T_i}^2}{n_i} = SST - SStr$$

일원배치 분산분석모형에서의 추론

- 처리 k 개를 비교하기 위한 모형

비교하려는 처리가 k 개이고 각 처리에서 반복측정을 n_i $(i = 1, \cdots, k)$ 번 할 때, i 번째 처리에서 j 번째 관측한 반응값을 Y_{ij} 라 하면, Y_{ij} 는 다음과 같이 표현할 수 있다.

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad j = 1, \cdots, n, \quad i = 1, \cdots, k$$

여기에서 μ_i 는 i 번째 처리의 모평균을 나타낸다. 오차항 ε_{ij}는 모두 서로 독립이고 평균이 0, 분산이 σ^2 인 정규분포를 따른다.

일원배치 분산분석모형에서의 추론

- k 개의 모집단의 모평균이 차이가 없다는 귀무가설과 대립가설

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k \quad vs. \quad H_1: not\ H_0$$

- F 분포

$$F = \frac{MStr}{MSE} = \frac{SStr/(k-1)}{SSE/(n-k)} \sim F(k-1, n-k) \quad Under\ H_0, \quad where\ n = \sum_{i=1}^{k} n_i$$

$$R: F = \frac{MStr}{MSE} \geq F_\alpha(k-1, n-k)$$

수고하셨습니다.

- 과제 X