실험계획과 분석

심송용(한림대학교 데이터과학스쿨)

http://jupiter.hallym.ac.kr

대비(Contrast)

대비(contrast)

- 앞의 튀김 기름에서 2번 기름과 나머지 기름과의 차이가 있는지

$$\mu_2 = \frac{\mu_1 + \mu_3 + \mu_4}{3} \Leftrightarrow 3\mu_2 - \mu_1 - \mu_3 - \mu_4 = 0$$

- 만일 1번 2번은 동물성 기름이고 3번 4번은 식물성 기름일 때 식물성 기름과 동물성 기름의 차이가 있는지

$$\frac{\mu_1 + \mu_2}{2} = \frac{\mu_3 + \mu_4}{2} \Leftrightarrow \mu_1 + \mu_2 - \mu_3 - \mu_4 = 0$$

일반적으로

$$C = c_1\mu_1 + c_2\mu_2 + \cdots + c_a\mu_a \,, \quad \sum_{i=1}^{a} c_i = 0$$

를 대비라고 하며 추정량은

$$\hat{C} = c_1\overline{y}_{1.} + c_2\overline{y}_{2.} + \cdots + c_a\overline{y}_{a.} = \sum_{i=1}^{a} c_i\overline{y}_{i.}$$

대비(Contrast)

가 되며 $\bar{y}_{i.} \sim N(\mu_i, \sigma^2/n_i)$이고 독립이므로

$$E(\hat{C}) = c_1\mu_1 + c_2\mu_2 + \cdots + c_a\mu_a = \sum_{i=1}^{a} c_i\mu_i = C \text{ (불편추정량)}$$

$$Var(\hat{C}) = c_1^2\frac{\sigma^2}{n_1} + c_2^2\frac{\sigma^2}{n_2} + \cdots + c_a^2\frac{\sigma^2}{n_a} = \sum_{i=1}^{a} c_i^2\frac{\sigma^2}{n_i}$$

이므로

$\hat{C} \sim N\left(\sum_{i=1}^{a} c_i\mu_i,\ \sum_{i=1}^{a} c_i^2\frac{\sigma^2}{n_i}\right) \Leftrightarrow \dfrac{\sum_{i=1}^{a} c_i\bar{y}_{i.} - \sum_{i=1}^{a} c_i\mu_i}{\sigma\sqrt{\sum_{i=1}^{a}\frac{c_i^2}{n_i}}} \sim N(0,1)$ 이다. σ 대신 $\sqrt{MSE}$를 사용하면

$$\frac{\sum_{i=1}^{a} c_i\bar{y}_{i.} - \sum_{i=1}^{a} c_i\mu_i}{\sqrt{MSE}\sqrt{\sum_{i=1}^{a}\frac{c_i^2}{n_i}}} \Leftrightarrow \frac{\hat{C} - C}{\sqrt{MSE}\sqrt{\sum_{i=1}^{a}\frac{c_i^2}{n_i}}} \sim t_{N-a} \text{ 이다.}$$

대비(Contrast)

따라서 대비 $C=\sum_{i=1}^{a} c_i\mu_i$에 대한 $100(1-\alpha)\%$ 신뢰구간은

$$1-\alpha=\Pr\left[-t_{N-a;\alpha/2}<\frac{\hat{C}-C}{\sqrt{MSE}\sqrt{\sum_{i=1}^{a}\frac{c_i^2}{n_i}}}<t_{N-a;\alpha/2}\right]$$ 에서

$$C=\sum_{i=1}^{a}c_i\bar{y}_{i.}\pm t_{N-a;\alpha/2}\sqrt{MSE}\sqrt{\sum_{i=1}^{a}\frac{c_i^2}{n_i}}$$

가 된다.

대비(Contrast)

귀무가설 $H_0 : C = \sum_{i=1}^{a} c_i \mu_i = 0$ 에 대해 H_0가 참이면

검정통계량 $t_0 = \dfrac{\sum_{i=1}^{a} c_i \bar{y}_{i.}}{\sqrt{MSE}\sqrt{\sum_{i=1}^{a} \frac{c_i^2}{n_i}}}$ 이 자유도 $N-a$인 t분포를 따르므로

- 대립가설이 $H_1 : C > 0$ 이면 $t_0 > t_{N-a;\alpha}$
- 대립가설이 $H_1 : C < 0$ 이면 $t_0 < -t_{N-a;\alpha}$
- 대립가설이 $H_1 : C \neq 0$ 이면 $|t_0| > t_{N-a;\alpha/2}$

일 때 H_0 기각한다.

t-분포와 F-분포의 관계에 의해서

$F_0 = t_0^2 = \dfrac{(\sum_{i=1}^{a} c_i \bar{y}_{i.})^2}{MSE \sum_{i=1}^{a} \frac{c_i^2}{n_i}} \sim F_{1,N-a}$ 임을 알 수 있다.

대비(Contrast)

두 개이상의 대비에 대한 동시 추론도 가능한데 이 경우에는

- 행렬대수를 통해서 이론적 전개가 가능
- 검정통계량의 분포는 F-분포
- 이런 이유로 SAS의 경우 한 개의 대비인 경우에도 F-검정을 함.
- 한 개의 대비인 경우 F-검정을 하면 자동으로 양측 검정만 시행함.

보기: 앞의 튀김 기름 자료에서 1, 2번은 동물성 기름이고 3, 4번은 식물성 기름이라고 할 때 이 동물성 기름과 식물성 기름의 차이에 대한 신뢰구간을 구해보고 이들의 차이여부에 대한 검정을 해보자.

$MSE = 100.9$, $n_1 = n_2 = n_3 = n_4 = 6$, $t_{20,0.025} = 2.086$, $c_1 = c_2 = 1$, $c_3 = c_4 = -1$ 이므로

$(c_1 = c_2 = -1$, $c_3 = c_4 = 1$로 하여도 부호만 다른 동일한 결과를 얻음)

$\hat{C} = \sum_{i=1}^{a} c_i \bar{y}_{i.} = 172 + 185 - 176 - 162 = 19$ 이고, 표준오차는

대비(Contrast)

$SE(\hat{C}) = \sqrt{MSE\sum_{i=1}^{a}\frac{c_i^2}{n_i}} = \sqrt{100.9\sum_{i=1}^{a}\frac{1}{6}} = 8.201626$ 이므로

대비 C 에 대한 95% 신뢰구간은 $19 \pm 2.086 \times 8.202 = (1.890,\ 36.109)$ 이며 귀무가설

$H_0 : C = 0$에 대한 검정통계량은

$t_0 = \dfrac{\sum_{i=1}^{a} c_i \bar{y}_{i.}}{\sqrt{MSE}\sqrt{\sum_{i=1}^{a}\frac{c_i^2}{n_i}}} = \dfrac{19}{8.202} = 2.317 > 2.086 = t_{20;0.025}$ 이므로 귀무가설을 기각한다. 즉, 동물성 기름과 식물성 기름은 차이가 있다.

SAS를 이용한 대비분석

SAS에서 대비는 CONTRAST 문을 사용하여 분석할 수 있다. CONTRAST 문은

CONTRAST 'label' *class_var* contrast

로 사용하며

- label에는 무엇에 대한 검정인지에 대한 적절한 이름 내지 설명을

대비(Contrast)

- *class_var* 에는 처리에 해당하는 변수의 이름을
- contrast에는 대비 계수를 설정한다. 해당되지 않는 처리수준에 0을 넣는 것에 주의!

SAS에서 대비를 사용하여

- 1번과 3번 기름이 차이가 나는지
- 앞에서 계산한 동물성과 식물성 기름의 차이가 있는지

각각 계산해보자

```
/* contrast1.sas */
data 스텝은 생략함. (앞의 m.comparisons3.sas 등 참고)
proc glm data = fatdata;
 class oil;
 model transfat = oil;
 contrast '동물성과 식물성 비교' oil 1 1 -1 -1;
 contrast '1번과 3번 비교' oil 1 0 -1 0;
run;
```

대비(Contrast)

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
동물성과 식물성 비교	1	541.5000000	541.5000000	5.37	0.0313
1번과 3번 비교	1	48.0000000	48.0000000	0.48	0.4983

이 출력에서

- 첫 번째 동물성과 식물성의 차이는 검정통계량 F_0이 5.37로 앞에서 계산 t_0와 $t_0^2 = 2.316508^2 = 5.36621 = F_0$인 관계로 같은 계산 결과를 보여주며,
- 두 번째 대비 검정은 F_0가 0.48로 귀무가설 $H_0 : \mu_1 - \mu_3 = 0$를 기각할 수 없음을 알 수 있다(대립가설은 $H_1 : \mu_1 - \mu_3 \neq 0$임)

(둘 이상의 동시 대비검정은 생략함)