제11장 통계적 추정과 가설검정 :두 모집단

- 표본의 독립성과 종속성
- 표본평균차의 표본분포
- 두 모집단 평균의 차에 대한 추정과 검정
- 두 모집단 비율의 차에 대한 추정과 검정
- 두 모집단 분산에 대한 추정과 검정

I. 표본의 독립성과 종속성

두 모집단의 평균을 비교하기 위하여 표본을 추출할 때는 세심한 주의가 필요하다.

만일 A집단 표본10개와 B집단 표본10개를 각각 확률표본으로 추출하였다면 이는 독립적인(independent) 표본을 추출하는 것이다.

그러나 두 집단에서 짝을 이루어 표본을 추출하였다면 이는 종속적 표본 또는 대응표본(paired sample)이라고 한다.

예를 들어 다이어트 전과 다이어트 6개월 후의 비교는 같은 사람에 속하는 자료이므로 대응표본으로 추출해야 한다.

Ⅱ. $(\overline{X}_1 - \overline{X}_2)$의 표본분포

$(\overline{x}_1 - \overline{x}_2)$의 표본분포의 특성

평균 : $E(\overline{x}_1 - \overline{x}_2) = \mu_1 - \mu_2$

표준편차(표준오차) :

$$\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

n_1 :모집단1로부터 추출하는 표본크기

n_2 :모집단2로부터 추출하는 표본크기

$(\overline{x}_1 - \overline{x}_2)$ 의 표본분포는 중심극한정리에 의하여

$n_1 \geq 30, n_2 \geq 30$ 일때 정규분포에 접근한다.

Ⅲ. 두 모평균의 차이에 대한 추정과 검정

1. 두 모평균의 차이에 대한 추정과 검정[독립표본]

[독립표본] 두 모집단의 표준편차를 아는 경우

$(\mu_1 - \mu_2)$ 에 대한 $100(1-\alpha)\%$ 신뢰구간 : σ 를 아는 경우

$$(\bar{x}_1 - \bar{x}_2) - Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$(\mu_1 - \mu_2)$ 에 대한 가설검정 : σ 를 아는 경우

Δ 양측검정

$H_0 : \mu_1 - \mu_2 = \mu_0$

$H_1 : \mu_1 - \mu_2 \neq \mu_0$

만일 $\dfrac{(\overline{X}_1 - \overline{X}_2) - \mu_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} < -Z_{\frac{\alpha}{2}}$ 또는 $\dfrac{(\overline{X}_1 - \overline{X}_2) - \mu_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} > Z_{\frac{\alpha}{2}}$ 이면 H_0 를기각

Δ 좌측검정

$H_0 : \mu_1 - \mu_2 \geq \mu_0$

$H_1 : \mu_1 - \mu_2 < \mu_0$

만일 $\dfrac{(\overline{X}_1 - \overline{X}_2) - \mu_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} < -Z_{\alpha}$ 이면 H_0 를기각

Δ 우측검정

$H_0 : \mu_1 - \mu_2 \leq \mu_0$

$H_1 : \mu_1 - \mu_2 > \mu_0$

만일 $\dfrac{(\overline{X}_1 - \overline{X}_2) - \mu_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} > Z_{\alpha}$ 이면 H_0 를기각

[독립표본] 두 모집단의 표준편차를 모르는 경우(대표본)

$(\mu_1 - \mu_2)$ 의 $100(1-\alpha)\%$ 신뢰구간 : 대표본

$$(\overline{X}_1 - \overline{X}_2) - Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \le \mu_1 - \mu_2 \le (\overline{X}_1 - \overline{X}_2) + Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

[독립표본] 두 모집단의 표준편차를 모르는 경우(소표본)

두 모집단의 통합분산

$$S_p^{\,2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$(\bar{x}_1 - \bar{x}_2)$의 표본분포: 소표본

평균 : $E(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2$

표준오차 : $s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}} = s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$

$(\mu_1 - \mu_2)$의 100 $(1-\alpha)$% 신뢰구간 : 소표본, $\sigma_1^2 = \sigma_2^2$

$$(\bar{X}_1 - \bar{X}_2) - t_{(n_1+n_2-2),\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le (\bar{X}_1 - \bar{X}_2) + t_{(n_1+n_2-2),\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

[독립표본] 두 모평균 차에 대한 가설검정 : 소표본 $\sigma_1^2 = \sigma_2^2$

- 양측검정

$$H_0 : \mu_1 - \mu_2 = \mu_0$$

$$H_1 : \mu_1 - \mu_2 \neq \mu_0$$

$$\frac{(\overline{X}_1 - \overline{X}_2) - \mu_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < -t_{(n_1+n_2-2),\frac{\alpha}{2}} \text{ 또는 } \frac{(\overline{X}_1 - \overline{X}_2) - \mu_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{(n_1+n_2-2),\frac{\alpha}{2}} \text{ 이면 } H_0 \text{를기각}$$

– 좌측검정

$$H_0 : \mu_1 - \mu_2 \geq \mu_0$$

$$H_1 : \mu_1 - \mu_2 < \mu_0$$

$$\frac{(\overline{X}_1 - \overline{X}_2) - \mu_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < -t_{(n_1+n_2-2),\alpha} \text{ 이면 } H_0 \text{를기각}$$

– 우측검정

$$H_0 : \mu_1 - \mu_2 \leq \mu_0$$

$$H_1 : \mu_1 - \mu_2 > \mu_0$$

$$\frac{(\overline{X}_1 - \overline{X}_2) - \mu_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{(n_1+n_2-2),\alpha} \text{ 이면 } H_0 \text{를기각}$$

Ⅲ. 두 모평균의 차이에 대한 추정과 검정

2. 두 모평균의 차이에 대한 추정과 검정[대응표본]

$(\mu_1 - \mu_2)$ 의 $100(1-\alpha)\%$ 신뢰구간 : 대응표본

$$\bar{d} = \frac{\sum (x_{i1} - X_{i2})}{n} = \frac{\sum_{i=1}^{n} d_i}{n}$$

$$s_d = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}}$$

$$\bar{d} - t_{n-1,\frac{\alpha}{2}} \frac{s_d}{\sqrt{n}} \le \mu_1 - \mu_2 \le \bar{d} + t_{n-1,\frac{\alpha}{2}} \frac{s_d}{\sqrt{n}}$$

$(\mu_1 - \mu_2)$에 대한 가설검정 : 대응표본

– 양측검정

$H_0 : \mu_1 - \mu_2 = \mu_0$

$H_1 : \mu_1 - \mu_2 \neq \mu_0$

$$\frac{\overline{d}_1 - \mu_0}{S_d / \sqrt{n}} < -t_{n-1,\frac{\alpha}{2}} \text{ 또는 } \frac{\overline{d}_1 - \mu_0}{S_d / \sqrt{n}} > t_{n-1,\frac{\alpha}{2}} \text{이면 } H_0\text{를 기각}$$

– 좌측검정

$H_0 : \mu_1 - \mu_2 \geq \mu_0$

$H_1 : \mu_1 - \mu_2 < \mu_0$

$$\frac{\overline{d}_1 - \mu_0}{S_d / \sqrt{n}} < -t_{n-1,\alpha} \text{이면 } H_0\text{를 기각}$$

– 우측검정

$H_0 : \mu_1 - \mu_2 \leq \mu_0$

$H_1 : \mu_1 - \mu_2 > \mu_0$

$$\frac{\overline{d}_1 - \mu_0}{S_d / \sqrt{n}} > t_{n-1,\alpha} \text{이면 } H_0\text{를 기각}$$

IV. 두 모비율의 차이에 대한 추정과 검정

1. 두 모비율의 차이에 대한 추정과 검정

두 모비율 차의 $100(1-\alpha)\%$ 신뢰구간 : 대표본

$$(\hat{p}_1-\hat{p}_2)-Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \le p_1-p_2 \le (\hat{p}_1-\hat{p}_2)$$

$$+Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

두 모비율 차이에 대한 가설검정 : 대표본

Δ양측검정

$$H_0 : p_1 - p_2 = 0$$

$$H_1 : p_1 - p_2 \neq 0$$

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$$

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}}} < -Z_{\frac{\alpha}{2}} \text{또는 } \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}}} > Z_{\frac{\alpha}{2}} \text{이면 } H_0 \text{를기각}$$

Δ좌측검정

$$H_0 : p_1 - p_2 \geq p_0$$

$$H_1 : p_1 - p_2 < p_0$$

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}}} < -Z_{\alpha} \text{이면 } H_0 \text{를기각}$$

Δ우측검정

$$H_0 : p_1 - p_2 \leq p_0$$

$$H_1 : p_1 - p_2 > p_0$$

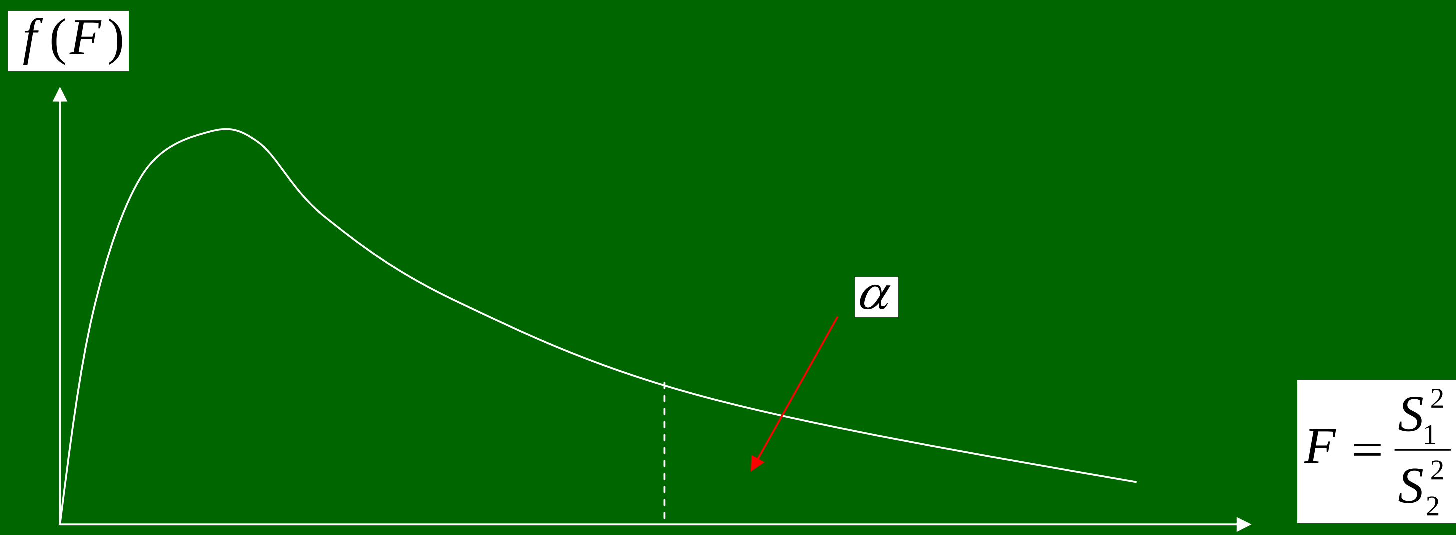
$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}}} > Z_{\alpha} \text{이면 } H_0 \text{를기각}$$

V. 두 모분산에 대한 추정과 검정

1. 두 모집단 분산비율에 대한 추정과 검정

F–분포

$$F_{n_1-1,n_2-1}=\frac{\dfrac{(n_1-1)S_1^2/\sigma_1^2}{n_1-1}}{\dfrac{(n_2-1)S_2^2/\sigma_2^2}{n_2-1}}=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}=\frac{S_1^2}{S_2^2}$$

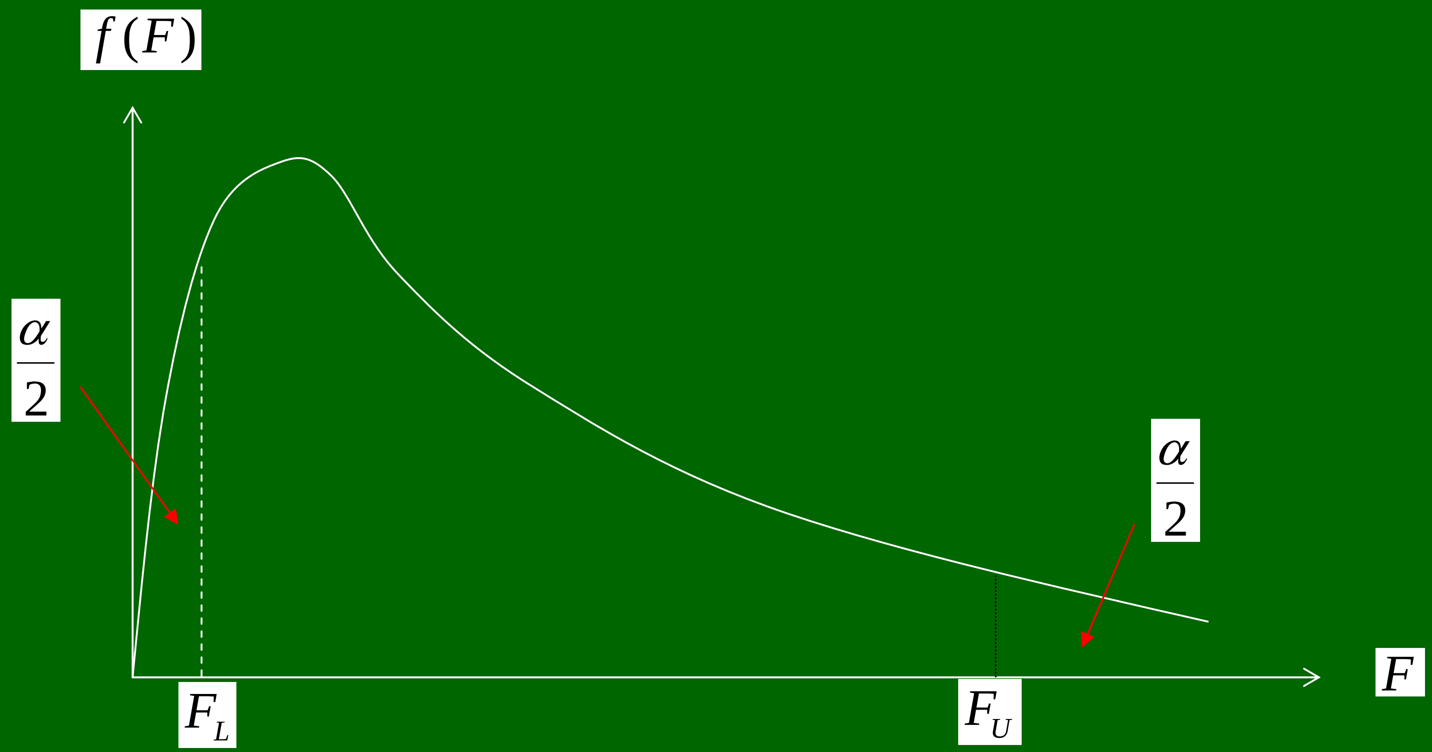


Ⅰ. 두 모분산에 대한 추정과 검정

두 모분산 비율에 대한 $100(1-\alpha)\%$ 신뢰구간

$$\frac{S_1^2/S_2^2}{F_U} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{S_1^2/S_2^2}{F_L}$$

$$F_{L\text{ 또는 }U} = F_{df_1, df_2, \frac{\alpha}{2}} = \frac{1}{F_{df_2, df_1, \frac{\alpha}{2}}}$$



I. 두 모분산에 대한 추정과 검정

두 모분산 비율에 대한 가설검정

– 양측검정

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

만일 $\frac{S_1^2}{S_2^2} < F_{n_1-1, n_2-1, 1-\frac{\alpha}{2}}$ 또는 $\frac{S_1^2}{S_2^2} > F_{n_1-1, n_2-1, \frac{\alpha}{2}}$ 이면 H_0를 기각

– 좌측검정

$$H_0 : \sigma_1^2 \geq \sigma_2^2$$
$$H_1 : \sigma_1^2 < \sigma_2^2$$

만일 $\frac{S_1^2}{S_2^2} < F_{n_1-1, n_2-1, 1-\alpha}$ 이면 H_0를 기각

– 우측검정

$$H_0 : \sigma_1^2 \leq \sigma_2^2$$
$$H_1 : \sigma_1^2 > \sigma_2^2$$

만일 $\frac{S_1^2}{S_2^2} > F_{n_1-1, n_2-1, \alpha}$ 이면 H_0를 기각