제12장 두 모집단에 대한추정

1) 모평균차($\mu_1 - \mu_2$)의 신뢰구간 추정

(a) 상호 독립인 두 확률표본(two independent random samples)을 이용하는 경우

: 두 모집단을 대상으로 첫 번째 모집단으로부터 m개의 표본을 추출하고 이와는 독립적으로 두 번째 모집단으로부터 n개의 표본을 추출함.

$X_{1,} X_{2,} \cdots, X_m \sim$ 정규$(\mu_{1,} \sigma^2)$ (=모집단 1)

$Y_{1,} Y_{2,} \cdots, Y_n \sim$ 정규$(\mu_{2,} \sigma^2)$ (=모집단 2)

[가정]

① 두 모집단은 정규분포를 따른다.

② $\sigma_1^2 = \sigma_2^2 = \sigma^2$,

③ σ^2은 미지수.

$(\mu_1 - \mu_2)$의 점 추정량은 $\overline{X} - \overline{Y}$ $\Rightarrow$ $\overline{X} - \overline{Y}$ 의 분포를 알아야 함.

[$\overline{X} - \overline{Y}$ 의 분포] : $\overline{X} \sim$ 정규$(\mu_{1,} \dfrac{\sigma^2}{m})$, $\overline{Y} \sim$ 정규$(\mu_{2,} \dfrac{\sigma^2}{n})$

$$\Rightarrow \quad \overline{X} - \overline{Y} \sim \text{정규}\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)\right).$$

이를 표준화하면, $\dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\dfrac{1}{m} + \dfrac{1}{n}}} \sim$ 정규$(0, 1)$

그러나 분모에 미지수 σ가 포함되어 있으므로 윗 식을 $\mu_1 - \mu_2$의 추론에 이용불가 $\Rightarrow$ σ (또는 σ^2)의 추정량이 필요

$\Rightarrow$ $s_p^2 = \dfrac{\sum_{i=1}^{m}(X_i - \overline{X})^2 + \sum_{i=1}^{n}(Y_i - \overline{Y})^2}{m + n - 2}$; σ^2의 "pooled" estimator

[$(\mu_1 - \mu_2)$의 신뢰구간] :

$$\Rightarrow \quad t^* = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t_{m+n-2} \text{ 를 이용}$$

$$\Rightarrow \quad P\left(-t_{\alpha/2;\, m+n-2} < \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{m} + \frac{1}{n}}} < t_{\alpha/2;\, m+n-2} \right) = 1 - \alpha$$

$$\Rightarrow \quad P \quad (\overline{X} - \overline{Y}) - t_{\alpha/2;\, m+n-2} \cdot s_p\sqrt{\frac{1}{m} + \frac{1}{n}} < \mu_1 - \mu_2 <$$

$$(\overline{X} - \overline{Y}) + t_{\alpha/2;\, m+n-2} \cdot s_p\sqrt{\frac{1}{m} + \frac{1}{n}} \quad = 1 - \alpha$$

따라서 $(\mu_1 - \mu_2)$의 $100(1-\alpha)\%$ 신뢰구간은,

$$(\overline{X} - \overline{Y}) \pm t_{\alpha/2;\, m+n-2} \cdot s_p\sqrt{\frac{1}{m} + \frac{1}{n}} \, .$$

(b) 짝을 이룬 표본(paired sample)을 이용하는 경우

[예] 우리나라 남성들에게 두 가지의 새로 개발된 두통약 A, B를 적용했을 때 얻어지는 효과의 차이(= $\mu_A - \mu_B$)에 대한 신뢰구간을 구하고자 한다.

방법 1: 20명의 남성(표본)을 무작위로 추출한 후 10명에게는 두통약 A를, 그리고 나머지 10명에게는 두통약 B를 적용하여 자료 얻은 후 신뢰구간 도출,

방법 2: 10명의 남성(표본)을 무작위로 추출한 후 10명 모두에게 두통약 A, B를 시차를 두고 복용케 하여 자료 얻은 후 신뢰구간 도출.

[예] 짝을 이룬 표본의 예 --- twin 자료, (동일 환자의) 복용전, 복용후 자료, (동일 환자의) 오른팔, 왼팔 자료 등등.

[자료]: 두통약 A= X, 두통약 B= Y

표본	X	Y	$D=X-Y$
1	X_1	Y_1	$D_1=X_1-Y_1$
2	X_2	Y_2	$D_2=X_2-Y_2$
·	·	·	· · ·
·	·	·	· · ·
n	X_n	Y_n	$D_n=X_n-Y_n$

[방법]: $D_i=X_i-Y_{i,}\quad i=1,2,\cdots,n,$ 라 하고

$\overline{D}=\frac{1}{n}\sum_{i=1}^{n}D_i=\overline{X}-\overline{Y},\quad s_D^2=\frac{1}{n-1}\sum_{i=1}^{n}(D_i-\overline{D})^2$ 라 하면,

$$\overline{D} \sim 정규(\mu_1 - \mu_2,\ \sigma_D^2/n) \quad \Rightarrow \quad \frac{\overline{D} - (\mu_1 - \mu_2)}{\sigma_D/\sqrt{n}} \sim 정규(0,\ 1)$$

$$\Rightarrow \quad \frac{\overline{D} - (\mu_1 - \mu_2)}{s_D/\sqrt{n}} \sim t_{n-1}$$

따라서 ($\mu_1 - \mu_2$)의 100($1-\alpha$)% 신뢰구간은,

$$(\overline{X} - \overline{Y}) \pm t_{\alpha/2;\, n-1} \cdot \frac{s_D}{\sqrt{n}} .$$

(7) 모비율차($p_1 - p_2$)의 신뢰구간 추정

① 두 이항 모집단을 대상으로 첫 번째 이항 모집단으로부터 m개의 표본을 추출하고, 두 번째 이항 모집단으로부터 n개의 표본을 추출함.

$X_1, X_2, \cdots\cdots, X_m \sim$ 베르누이(p_1), (=모집단 1)

$Y_1, Y_2, \cdots\cdots, Y_n \sim$ 베르누이(p_2) (=모집단 2)

$\Rightarrow \quad X = \sum_{i=1}^{m} X_i$라 하면, $X \sim$ 이항$(m,\ p_1)$,

$Y = \sum_{j=1}^{n} Y_j$라 하면, $Y \sim$ 이항$(n,\ p_2)$.

$\Rightarrow$ 모비율차 $(p_1 - p_2)$의 추정량 $= \dfrac{X}{m} - \dfrac{Y}{n} = \widehat{p}_1 - \widehat{p}_2$.

② $(p_1 - p_2)$의 신뢰구간 추정

$\widehat{p}_1$과 $\widehat{p}_2$이 각각 근사적인 정규분포 따름.

$$\Rightarrow \widehat{p}_1 - \widehat{p}_2 \sim \text{정규}\left(p_1 - p_2,\ \frac{p_1(1-p_1)}{m} + \frac{p_2(1-p_2)}{n}\right), \text{ approximately.}$$

따라서, $(p_1 - p_2)$의 $100(1-\alpha)\%$ 신뢰구간은

$$\Rightarrow (\widehat{p}_1 - \widehat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{m} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n}}\ .$$

(8) 두 모분산比의 신뢰구간추정

$X_{1,} X_{2,} \cdots\cdots, X_m \sim$ 정규$(\mu_{1,}\ \sigma_1^2)$ (=모집단 1)

$Y_{1,} Y_{2,} \cdots\cdots, Y_n \sim$ 정규$(\mu_{2,}\ \sigma_2^2)$ (=모집단 2)

$\Rightarrow$ $\dfrac{\sigma_1^2}{\sigma_2^2}$ 의 100(1- α)% 신뢰구간은?

$$\frac{(m-1)s_1^2}{\sigma_1^2} = \frac{\sum_{i=1}^{m}(X_i - \overline{X})^2}{\sigma_1^2} \sim \chi^2_{m-1}. \quad \frac{(n-1)s_2^2}{\sigma_2^2} = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{\sigma_2^2} \sim \chi^2_{n-1},$$

$$\Rightarrow \frac{\dfrac{(m-1)s_1^2}{\sigma_1^2}/(m-1)}{\dfrac{(n-1)s_2^2}{\sigma_2^2}/(n-1)} = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{m-1,\, n-1}.$$

$$\Rightarrow P\left(F_{1-\alpha/2;\, m-1,\, n-1} < \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} < F_{\alpha/2;\, m-1,\, n-1}\right) = 1-\alpha$$

따라서 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 의 100($1-\alpha$)% 신뢰구간은,

$$\left(\frac{1}{F_{\alpha/2;\, m-1,\, n-1}} \cdot \frac{s_1^2}{s_2^2}, \quad \frac{1}{F_{1-\alpha/2;\, m-1,\, n-1}} \cdot \frac{s_1^2}{s_2^2} \right).$$