두 모집단의 비교

두 개의 독립 표본

- 독립적인 두 모집단의 평균차이 $(\mu_1 - \mu_2)$를 추론하고자 한다.

$$X_1, X_2, \cdots, X_{n_1} \sim Dist(\mu_1, {\sigma_1}^2)$$

$$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i \,, \qquad {s_1}^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1} (X_i - \bar{X})^2$$

$$Y_1, Y_2, \cdots, Y_{n_2} \sim Dist(\mu_2, {\sigma_2}^2)$$

$$\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i \,, \qquad {s_2}^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$$

정규분포의 성질

- 두 확률변수가 서로 독립이고 다음과 같이 정규분포를 따를 때,

$$X \sim N(\mu_1, {\sigma_1}^2), \qquad Y \sim N(\mu_2, {\sigma_2}^2)$$

두 확률변수의 합과 차는 다음과 같은 정규분포를 따른다.

$$X \pm Y \sim N(\mu_1 \pm \mu_2, {\sigma_1}^2 + {\sigma_2}^2)$$

두 개의 독립 표본(표본의 크기가 클 때)

- 표본의 수가 충분히 많을 때, 두 표본평균은 근사적으로 정규분포를 따르고,

$$\bar{X} \sim N\left(\mu_1, \frac{{\sigma_1}^2}{n_1}\right), \qquad \bar{Y} \sim N\left(\mu_2, \frac{{\sigma_2}^2}{n_2}\right) \qquad by\ C.L.T.$$

- 두 표본평균의 차는 다음과 같은 정규분포를 따른다.

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}\right)$$

모평균의 차 $\mu_1 - \mu_2$에 대한 신뢰구간(표본의 크기가 클 때)

- $\bar{X} - \bar{Y}$를 표준화하고, 모분산을 표본분산으로 대체하여도 다음과 같이 근사적으로 표준정규분포를 따른다.

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}} \sim N(0,1)$$

- $\mu_1 - \mu_2$의 신뢰구간을 구하면,

$$(\bar{X} - \bar{Y}) \pm z_{\alpha/2}\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}$$

$H_0: \mu_1 - \mu_2 = \delta_0$에 대한 검정(표본의 크기가 클 때)

- 표본의 크기가 모두 30 이상일 때 가설 $H_0: \mu_1 - \mu_2 = \delta_0$에 대한 검정통계량

$$Z = \frac{(\bar{X} - \bar{Y}) - (\delta_0)}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

H_0 가 참일 때, Z 는 표준정규분포를 따른다. 각 대립 가설에 대하여 유의수준 α를 갖는 기각역은 다음과 같다.

$$H_1: \mu_1 - \mu_2 < \delta_0 \qquad \mathrm{R}: \mathrm{Z} \le -z_\alpha$$
$$H_1: \mu_1 - \mu_2 > \delta_0 \qquad \mathrm{R}: \mathrm{Z} \ge z_\alpha$$
$$H_1: \mu_1 - \mu_2 \ne \delta_0 \qquad \mathrm{R}: |Z| \ge z_{\alpha/2}$$

모평균의 차 $\mu_1 - \mu_2$에 대한 추론(표본의 크기가 작을 때)

- 표본의 크기가 작을 때 필요한 가정

  (1) 두 모집단이 모두 정규분포를 따른다.

  (2) 두 모집단의 표준편차가 일치한다.($\sigma_1 = \sigma_2 = \sigma$) or

  두 모집단의 표준편차가 다르다.($\sigma_1 \neq \sigma_2$)

- 두 모집단의 표준편차가 같은지 판단하는 규칙

$$\frac{1}{2} \leq \frac{s_1}{s_2} \leq 2 \quad \rightarrow \quad \sigma_1 = \sigma_2$$

- 위의 두 가정(일치 가정)을 만족할 때, $(\bar{X} - \bar{Y})$ 는 다음과 같은 정규분포를 따른다.

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

- 공통분산 σ^2의 합동 추정량(Pooled Estimator)

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

모평균의 차 $\mu_1 - \mu_2$에 대한 신뢰구간(표본의 크기가 작을 때, 모표준편차가 같을 때)

- 두 정규모집단에서 독립적으로 추출된 두 표본으로부터 얻게 되는 표준화된 확률변수는 자유도가 $(n_1 + n_2 - 2)$인 t 분포를 따른다.

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

- 두 모집단이 모두 정규분포를 따르고 두 모표준편차가 같을 때 $(\mu_1 - \mu_2)$에 대한 $100(1-\alpha)\%$ 신뢰구간은 다음과 같다.

$$(\bar{X} - \bar{Y}) \pm t_{\alpha/2}(n_1 + n_2 - 2) \times s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$H_0: \mu_1 - \mu_2 = \delta_0$에 대한 검정(표본의 크기가 작을 때, 모표준편차가 같을 때)

- 두 모집단이 정규분포를 따르고, 두 모표준편차가 같을 때 가설 $H_0: \mu_1 - \mu_2 = \delta_0$에 대한 검정통계량은 다음과 같다.

$$t = \frac{(\bar{X} - \bar{Y}) - (\delta_0)}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

검정통계량의 분포는 H_0가 맞을 때 자유도가 $(n_1 + n_2 - 2)$인 t 분포를 따른다. 각 대립가설에 대하여 유의수준 α를 갖는 기각역은 다음과 같다.

$$\begin{array}{ll} H_1: \mu_1 - \mu_2 < \delta_0 & \mathrm{R}: t \le -t_\alpha(n_1 + n_2 - 2) \\ H_1: \mu_1 - \mu_2 > \delta_0 & \mathrm{R}: t \ge t_\alpha(n_1 + n_2 - 2) \\ H_1: \mu_1 - \mu_2 \ne \delta_0 & \mathrm{R}: |t| \ge t_{\alpha/2}(n_1 + n_2 - 2) \end{array}$$

모평균의 차 $\mu_1 - \mu_2$에 대한 신뢰구간(표본의 크기가 작을 때, 모표준편차가 다를 때)

- 두 정규모집단에서 독립적으로 추출된 두 표본으로부터 얻게 되는 표준화된 확률변수는 근사적으로 자유도가 $\min(n_1 - 1, n_2 - 1)$인 t 분포를 따른다.

$$t^* = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}} \sim t(\min(n_1 - 1, n_2 - 1))$$

- 두 집단이 모두 정규분포를 따르고 모표준편차가 다를 때 $100(1-\alpha)\%$ 신뢰구간은 다음과 같다.

$$(\bar{X} - \bar{Y}) \pm t_{\alpha/2}(\min(n_1 - 1, n_2 - 1)) \times \sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}$$

$H_0: \mu_1 - \mu_2 = \delta_0$에 대한 검정(표본의 크기가 작을 때, 모표준편차가 다를 때)

- 두 모집단이 정규분포를 따르고, 두 모표준편차가 다를 때 가설 $H_0: \mu_1 - \mu_2 = \delta_0$에 대한 검정통계량은 다음과 같다.

$$t^* = \frac{(\bar{X} - \bar{Y}) - (\delta_0)}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

검정통계량의 분포는 H_0가 맞을 때 자유도가 $\min(n_1 - 1, n_2 - 1)$인 t 분포를 따른다. 각 대립가설에 대하여 유의수준 α를 갖는 기각역은 다음과 같다.

$$H_1: \mu_1 - \mu_2 < \delta_0 \qquad \mathrm{R}: t \le -t_\alpha(\min(n_1 - 1, n_2 - 1))$$
$$H_1: \mu_1 - \mu_2 > \delta_0 \qquad \mathrm{R}: t \ge t_\alpha(\min(n_1 - 1, n_2 - 1))$$
$$H_1: \mu_1 - \mu_2 \ne \delta_0 \qquad \mathrm{R}: |t| \ge t_{\alpha/2}(\min(n_1 - 1, n_2 - 1))$$

수고하셨습니다.

- 과제 X