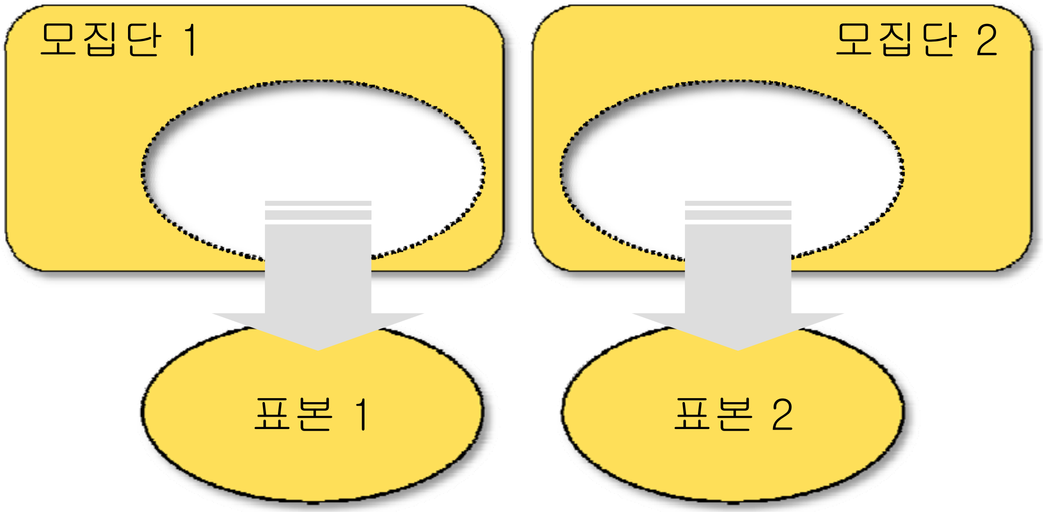
통계학
– 제 12 장 두 모집단의 비교

0011 0010 1010 1101 0001 0100 1011

두 개의 독립 표본

- 독립표본에 의한 두 모평균의 비교
  - 표본추출 과정
    - 정규분포를 따르며 상호 독립인 두 모집단으로부터 표본을 추출



- 표본 1 : 평균이 μ_1 이고 표준편차가 σ_1 인 모집단으로부터 추출된 n_1 개의 표본 – $X_1, X_2, \cdots, X_{n_1}$
- 표본 2 : 평균이 μ_2 이고 표준편차가 σ_2 인 모집단으로부터 추출된 n_2 개의 표본 – $Y_1, Y_2, \cdots, Y_{n_2}$

- 독립표본 t-검정
  - 귀무가설과 대립가설의 설정
    - 귀무가설 (H_0) :
      두 집단의 평균은 서로 같다 ($\mu_1 = \mu_2$)
    - 대립가설 (H_1) :
      두 집단의 평균은 서로 같지 않다 ($\mu_1 \neq \mu_2$)
      모집단 1의 평균이 모집단 2의 평균보다 크다 ($\mu_1 > \mu_2$)
      모집단 1의 평균이 모집단 2의 평균보다 작다 ($\mu_1 < \mu_2$)
  - 유의수준의 설정

- 검정통계량의 결정 및 분포 확인 (표본의 크기가 큰 경우)

두 확률변수 X, Y가 서로 독립이고

$$X \sim N\left(\mu_X,\ \sigma_X^2\right),\quad Y \sim N\left(\mu_Y,\ \sigma_Y^2\right)$$

일 때, 두 확률변수의 합과 차도 다음과 같은 정규분포를 따름

$$\left(X \pm Y\right) \sim N\left(\mu_X \pm \mu_Y,\ \sigma_X^2 + \sigma_Y^2\right)$$

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

- 검정통계량 : $Z = \dfrac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \sim N(0,\ 1)$

- 검정통계량의 결정 및 분포 확인 (표본의 크기가 작은 경우)
  - 두 확률변수 X , Y 가 서로 독립이고

    $$X \sim N\left(\mu_1,\ \sigma_1^2\right),\quad Y \sim N\left(\mu_2,\ \sigma_2^2\right)$$

    이며, 두 모분산이 같다($\sigma_1^2 = \sigma_2^2$)고 가정
  - 검정통계량 :

    $$T = \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_1 - \mu_2\right)}{S_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t\left(n_1 + n_2 - 2\right)$$

    여기서,

    $$S_p^2 = \frac{\left(n_1 - 1\right)S_1^2 + \left(n_2 - 1\right)S_2^2}{n_1 + n_2 - 2}$$: 공통분산의 합동 추정량

- 예 – [어린이 성장과 사회환경에 관한 연구]
  - 부모가 고등교육을 받지 못한 경우와 받은 경우 초등학교 1학년생들의 언어능력 비교 연구
  - 부모가 고등교육을 받지 못한 1학년생 66명의 언어능력 평균 : 305, 표준편차 : 29
  - 부모가 고등교육을 받은 1학년생(집단2) 38명의 언어능력 평균 : 311, 표준편차 : 40
  - 두 집단의 언어능력은 같은지 유의수준 5%에서 검정
  - 가설설정
    - 귀무가설 (H_0) : 두 집단의 언어능력 평균은 서로 같다 ($\mu_1 = \mu_2$)
    - 대립가설 (H_1) : 두 집단의 언어능력 평균은 서로 같지 않다 ($\mu_1 \neq \mu_2$)

- 예 – [어린이 성장과 사회환경에 관한 연구]
  - 검정통계량의 결정 및 분포 확인
    - 표본의 크기가 충분히 크므로
    - 검정통계량 :

$$Z = \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0,\ 1)$$

$$\frac{(305 - 311) - (0)}{\sqrt{\frac{29^2}{66} + \frac{40^2}{38}}} \doteq -0.81$$

    - 유의확률 : $P\left(Z > \left|-0.81\right|\right) = 0.4180$

- 예 – [우유생산량 비교]
  - 목초의 종류에 따른 우유생산량의 차이 확인
  - 3주 동안 우유생산량 측정
  - 13마리의 젖소에게는 인공적으로 건조시킨 목초를 줌
    평균 : 45.15, 표준편차 : 7.998
  - 12마리의 젖소에게는 들판에서 말린 목초를 줌
    평균 : 42.25, 표준편차 : 8.740

  - 가설설정
    - 귀무가설 (H_0) :
      목초의 종류에 따른 우유생산량의 평균은 같다 ($\mu_1 = \mu_2$)
    - 대립가설 (H_1) :
      목초의 종류에 따른 우유생산량의 평균은 다르다 ($\mu_1 \neq \mu_2$)

- 예 – [우유생산량 비교]
  - 검정통계량의 결정 및 분포 확인
    - 표본의 크기가 작으므로
    - 검정통계량 :

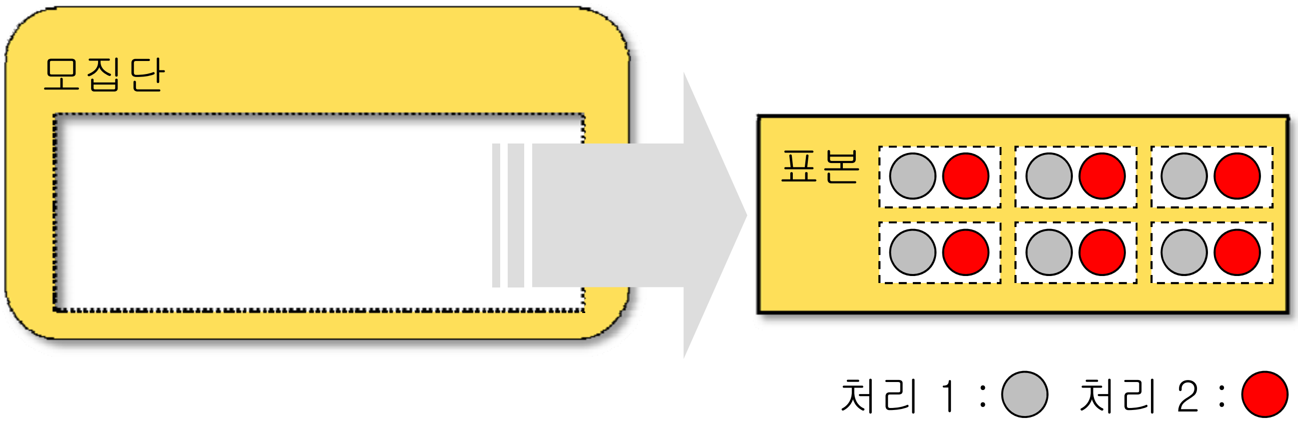
$$T = \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_1 - \mu_2\right)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t\left(n_1 + n_2 - 2\right)$$

$$S_p = \sqrt{\frac{(13-1)\times 7.998^2 + (12-1)\times 8.740^2}{13+12-2}} \doteq 8.361$$

$$\frac{(45.15 - 42.25) - (0)}{8.361 \times \sqrt{\frac{1}{13} + \frac{1}{12}}} \doteq 0.866$$

짝 비교

- 쌍 관측에 의한 두 모평균의 비교
  - 표본추출
    - 하나의 모집단으로부터 표본을 추출하며
      실험단위를 동질적인 쌍으로 묶어 각 쌍에 대해 두 처리를 적용



    - 블럭화 :
      - 동일한 특성에 따라 실험단위를 묶음으로서
        특성치에 의한 변동 요인을 제거하는 작업

- 대응표본 t-검정
  - 귀무가설과 대립가설의 설정
    - 귀무가설 (H_0) :
      두 처리의 평균 차이는 0이다 ($\mu_1 - \mu_2 = \delta_0 = 0$)
    - 대립가설 (H_1) :
      두 처리의 평균 차이는 0이 아니다 ($\mu_1 - \mu_2 = \delta_0 \neq 0$)
      두 처리의 평균 차이는 0보다 크다 ($\mu_1 - \mu_2 = \delta_0 > 0$)
      두 처리의 평균 차이는 0보다 작다 ($\mu_1 - \mu_2 = \delta_0 < 0$)
  - 유의수준의 설정

- 검정통계량의 결정 및 분포 확인
  - 두 처리 차이의 표본평균, 표본분산

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i \qquad S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2$$

  - 표본의 크기가 큰 경우의 검정통계량 :

$$Z = \frac{\overline{D} - \delta_0}{S_D / \sqrt{n}} \sim N(0,\ 1)$$

  - 표본의 크기가 작은 경우의 검정통계량 :

$$T = \frac{\overline{D} - \delta_0}{S_D / \sqrt{n}} \sim t(n-1)$$

- 예 – [약의 부작용]
  - 어떤 약의 부작용으로 혈압강하 현상 발생
  - 15명 환자 대상으로 복용 전과 후의 혈압 측정
  - 복용 전 혈압 – 복용 후 혈압

    평균 : 8.80, 표준편차 : 10.98
  - 부작용이 나타난다고 할 수 있는지 유의수준 1%에서 검정

  - 가설설정
    - 귀무가설 (H_0) :

      복용 전과 복용 후의 평균 차이는 0이다 ($\mu_1 - \mu_2 = \delta_0 = 0$)
    - 대립가설 (H_1) :

      복용 전과 복용 후의 평균 차이는 0보다 크다 ($\mu_1 - \mu_2 = \delta_0 \neq 0$)

- 예 – [약의 부작용]
  - 검정통계량의 결정 및 분포 확인
    - 표본의 크기가 작으므로
    - 검정통계량 :

$$T = \frac{\overline{D} - \delta_0}{S_D / \sqrt{n}} \sim t(n-1)$$

$$\frac{8.80 - 0}{10.98 / \sqrt{15}} \doteq 3.104$$