정규분포? t분포?
σ^2? s^2?

- σ_1^2, σ_2^2 알고있나?
  - ① σ, 정규분포
  - 표본 크기가 충분히 큰가?
    - ② S, 정규분포
    - $\sigma_1^2 = \sigma_2^2$ 가정할 수 있나?
      - ③ S_p, T분포 $df = n_1+n_2-2$
      - ④ S, T분포 $df = v^*$

$$E(\bar{X}_1-\bar{X}_2) = \mu_1-\mu_2$$

$$Var(\bar{X}_1-\bar{X}_2) = Var(\bar{X}_1)+Var(\bar{X}_2) = \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$$

① σ_1^2, σ_2^2 알고있는 경우

$$\left[(\bar{X}_1-\bar{X}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\ (\bar{X}_1+\bar{X}_2)+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right]$$

* X_1, X_2가 정규분포가 아니더라도
  중심극한정리에 의해 $\bar{X}_1, \bar{X}_2$는 정규분포라 할 수 있음

② σ_1^2, σ_2^2 모르지만 표본크기 충분히 큰 경우

$$\left[(\bar{X}_1-\bar{X}_2) + z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}},\ (\bar{X}_1+\bar{X}_2)+z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right]$$

③ σ_1^2, σ_2^2 모르지만 $\sigma_1^2=\sigma_2^2$인 경우

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

$$T = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

④ σ_1^2, σ_2^2 모르고 $\sigma_1^2 \neq \sigma_2^2$인 경우

$$T = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \sim t_{df} \text{ 분포}$$

$$df = \frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}}$$

→ 간편하게 $\min(n_1, n_2)-1$ 을 사용하기도 함

쌍체비교

1개의 독립표본에 대한 대응 비교 (전후 비교)

$d_i = X_i - Y_i$ → 단일 모집단 추정 문제

$$\left[\bar{d} - t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} , \bar{d} + t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} \right]$$

* 전후집단은 각 쌍이 되어야

$\mu_1 - \mu_2 = 0$ 인지를 구간추정으로 보자함

→ d의 추정구간이 0을 포함하는지, 포함안하면 "전후 차이가 있다"

장점

표본크기 감소

동일대상이라 비교의 효과 UP

두개 독립표본일 때보다 분산 작아짐

→ $Var(\bar{X} - \bar{Y}) = Var(\bar{X}) + Var(\bar{Y}) - 2Cov(\bar{X}, \bar{Y})$

단점

자유도 감소 ($2n-2 \rightarrow n-1$)

→ 자유도 down → t 값 커지므로 같은 신뢰수준에서

신뢰구간이 넓어진다는 단점

모분산 추정

$X \sim N(\mu, \sigma^2)$

$s^2 = \frac{\Sigma(X_i - \bar{X})^2}{n-1}$, $s^2(n-1) = \Sigma(X_i - \bar{X})^2$

$$\frac{(n-1)s^2}{\sigma^2} = \frac{\Sigma(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\therefore \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$P\left(\chi^2_{n-1, 1-\alpha/2} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{n-1, \alpha/2}\right) = 1 - \alpha$$

$$\therefore P\left(\frac{(n-1)s^2}{\chi^2_{n-1, \alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{n-1, \alpha/2}}\right)$$

두모집단 간 분산 차이 검정

: 값 차이 X , 비율에 의해서 (F검정)

가정

두 모집단은 정규분포

표본은 두 모집단에서 독립적으로 추출

$$F = \frac{\chi_2^2}{\chi_1^2}$$

$$\frac{s_1^2(n_1-1)}{\sigma_1^2} \sim \chi^2_{n_1-1} \qquad \frac{s_2^2(n_2-1)}{\sigma_2^2} \sim \chi^2_{n_2-1}$$

$$P\left(F_{1-\frac{\alpha}{2}, n_1-1, n_2-1} \le \frac{\frac{s_2^2(n_2-1)}{\sigma_2^2}}{\frac{s_1^2(n_1-1)}{\sigma_1^2}} \le F_{\frac{\alpha}{2}, n_1-1, n_2-1}\right) = 1-\alpha$$

$$\therefore P\left(\frac{\frac{s_1^2}{s_2^2}}{F_{\alpha/2, n_1-1, n_2-1}} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{\frac{s_1^2}{s_2^2}}{F_{1-\alpha/2, n_1-1, n_2-1}}\right) = 1-\alpha$$

$$\frac{s_1^2}{s_2^2} \sim \frac{\sigma_1^2}{\sigma_2^2} F_{n_1-1, n_2-1}$$

$$* F_{1-\alpha/2, n_1-1, n_2-1} = \frac{1}{F_{\alpha/2, n_2-1, n_1-1}}$$

구간추정 관련요인

$1-\alpha \rightarrow \alpha$, σ^2 , n

① $\alpha \uparrow$ 신뢰구간 좁아짐 → 모수가 포함될 확률 떨어짐

신뢰구간 = $2 \times Z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$

② σ^2 커질수록 신뢰구간 大

but 의도적으로 조정할 수 없는 요소.

→ 가능한 한 동질적인 개체들을 뽑아서 분산 줄이는 것이 바람직

③ n이 커질수록 신뢰구간 좁아짐 → 오차확률 올라가는 문제

n이 작을수록 신뢰구간 넓어짐 → 부정확하고 부적합한 정보 제공.

→ 최선 대안은 표본 크기 늘리는 것

→ "최소 표본 크기" ← 최소한 요만큼은 되어야된다~

$$n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{MOE}\right)^2 \text{ or } \left(\frac{t_{\alpha/2, n-1} \cdot S}{MOE}\right)^2$$

$$MOE = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \text{ or } t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$$