통계적 추정

- 통계적 추정 : 표본의 측정 자료를 근거로 모수를 추측하는 것.
- 모수(θ), 추정량($\hat{\theta}$)

<점추정량>

▶ 좋은 추정량의 조건

1. 불편추정량 만족

- $E(\hat{\theta}) = \theta$
- $E(\overline{X}) = \mu$

$$E(S^2) = E(\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}) = \sigma^2$$

2. 최소분산 불편추정량

- 어떤 모수의 불편추정량은 여러 개 있을 수 있음. -> 분산이 작을수록 좋은 추정량
- ex) 표본평균, 표본중앙값 -> 모평균의 불편추정량. but, 표본중앙값 대신 표본평균을 모평균의 추정량으로 사용하는 이유 : 표본중앙값의 분산보다 표본평균의 분산이 더 작기 때문

<구간추정>

- 모수를 신뢰구간으로 나타내는 것
- $P(\widehat{\theta_L} < \theta < \widehat{\theta_U}) = 1-\alpha$: 모수 θ에 대한 $1-\alpha$% 신뢰구간(모수가 해당 구간 안에 있을 확률이 $1-\alpha$)

▶ 모평균의 구간추정

-> $\overline{X}$이 정규분포를 따른다면

- 신뢰구간 : $P(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$
- $1-\alpha$%에서의 최대오차($\overline{X}$을 μ의 점추정량으로 사용했을 때의) : $e = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
- 최대 오차(e)를 넘지 않는 n = $(\frac{z_{\alpha/2}\sigma}{e})^2$

-> $\overline{X}$가 t분포를 따른다면

- 신뢰구간 : $P(\overline{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}) = 1-\alpha$

▶ 서로 독립인 두 모평균 차이의 구간추정

-> 두 표본평균이 정규분포를 따르는 경우

- 신뢰구간 : $P((\overline{X_1} - \overline{X_2}) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{X_1} - \overline{X_2}) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}) = 1-\alpha$

-> 두 표본평균이 t분포를 따르고, 모평균이 같은 경우

- 신뢰구간 : $P((\overline{X_1}-\overline{X_2})-T_{\alpha/2}(n_1+n_2-2)S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}<\mu_1-\mu_2<$

$$(\overline{X_1}-\overline{X_2})+T_{\alpha/2}(n_1+n_2-2)S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}})=1-\alpha$$

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

(풀이)

- $Z=\frac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}}$, $V(n_1+n_2-2)=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\sigma^2}$
- $\frac{Z}{\sqrt{V/(n_1+n_2-2)}}\sim T(n_1+n_2-2)$
- $\frac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim T(n_1+n_2-2)$, $S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ (합동추정량)

-> 두 표본평균이 t분포를 따르고, 모평균이 다른 경우

$$P((\overline{X_1}-\overline{X_2})-T_{\alpha/2}(v)\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}<\mu_1-\mu_2<(\overline{X_1}-\overline{X_2})+T_{\alpha/2}(v)\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}})=1-\alpha$$

$, v=\frac{(S_1^2/n_1+S_2^2/n_2)^2}{[(S_1^2/n_1)^2/(n_1-1)]+[(S_2^2/n_2)^2/(n_2-1)]}$=satterthwaite 자유도 (등분산 자유도)

$$\frac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}\sim T(v)\ ,\ v=\frac{(S_1^2/n_1+S_2^2/n_2)^2}{[(S_1^2/n_1)^2/(n_1-1)]+[(S_2^2/n_2)^2/(n_2-1)]}$$

▶ 서로 종속인(독립이 아닌) 두 모평균 차이의 구간추정 - ex) 쌍체표본

- 두 집단이 서로 독립이 아니기 때문에 ($\overline{X_A}-\overline{X_B}$와 같은 방식으로 구할 수 없음)
- 그 차이 자체에 대한 확률분포를 구해서 차이의 모평균를 추정해야 함.
- 차이의 확률분포에 따라 적절한 방식으로 차이의 모평균 신뢰구간을 구하면 됨.

▶ 모비율의 구간검정

- 표본의 확률변수 X는 베르누이 분포를 갖고, 표본에서 추출한 $E(X)=np$은 sample의 크기가 크면 정규분포를 띔

- $\hat{p} \sim N(p, pq/n)$
- 신뢰구간 : $P(\hat{p} - z_{\alpha/2}\sqrt{pq/n} < p < \hat{p} + z_{\alpha/2}\sqrt{pq/n}) = 1-\alpha$
- 최대오차 e를 만족하는 n : $\hat{p}\hat{q}(\frac{z_{\alpha/2}}{e})^2$

$\frac{\hat{p}-p}{\sqrt{pq/n}} \sim N(0,1)$ -> 모집단의 p를 알지 못한다면 n이 충분히 크기 때문에 $\hat{p}$를 p대신 사용가능

$\frac{\hat{p}-p}{\sqrt{\hat{p}\hat{q}/n}} \sim N(0,1)$

▶ **서로 독립인 모비율 차이의 구간검정**

- 신뢰구간 : $P((\hat{p_1} - \hat{p_2}) - z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} < p_1 - p_2 < (\hat{p_1} - \hat{p_2}) + z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} = 1-\alpha$

**이항분포의 정규근사*
- 이항분포를 띄는 모집단 -> 표본 추출
- 이항분포에서 각 샘플은 베르누이분포(p)를 가짐
- $\overline{X} = Y_1 + Y_2 + Y_3 + ... + Y_n$
- $\overline{X} \sim N(np, npq), n \geqq 30$r (중심극한정리에 의해)
- $\hat{p} \sim N(p, pq/n)$

▶ **모분산의 구간추정**

- 신뢰구간 : $P(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}) = 1-\alpha$

▶ **각 독립인 확률분포의 모분산 비 구간추정**

- 신뢰구간 : $P(\frac{S_1^2}{S_2^2}\frac{1}{F_{\frac{\alpha}{2}}(n_1-1, n_2-1)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\frac{1}{F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)}) = 1-\alpha$

*카이제곱, F분포는 대칭이 아니기 때문에 주의해야 함!