1-1. 큰 수의 약법칙

= n이 충분히 클 때, 표본 평균이 모평균에 가까워질 확률이 1에 가까워진다

증명.

표본 평균 $\overline{X_n}$ 에서 $E(\overline{X_n})=\mu$, $Var(\overline{X_n})=\frac{\sigma^2}{n}$

체비셰프 부등식 => 모든 $k>0$ 에 대해 $P(X-\mu \ge k\sigma) \le \frac{1}{k^2}$

체비셰프 부등식을 적용하면 $P(|\overline{X_n}-\mu| \ge \varepsilon) \le \frac{Var(\overline{X_n})}{\varepsilon^2}$

$$P(|\overline{X_n}-\mu| \ge \varepsilon) \le \frac{\sigma^2}{n\varepsilon^2} \quad (\because Var(\overline{X_n}) = \frac{\sigma^2}{n}) \cdots\cdots ①$$

①에서 $n\to\infty$ 이면 $\frac{\sigma^2}{n\varepsilon^2} \to 0$ 이므로

$$\lim_{n\to\infty} P(|\overline{X_n}-\mu| \ge \varepsilon) = 0$$

$$\therefore \lim_{n\to\infty} P(|\overline{X_n}-\mu| < \varepsilon) = 1$$

2-1. CLT 의 정의

크기가 n인 랜덤 표본 $X_1, X_2 \cdots X_n$ 이 모평균 μ, 모분산 σ^2인 분포에서 왔을 때, 확률변수 W의 분포는

$W = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma}$ 로 표준 정규분포 $N(0,1)$로 수렴한다.

2-2. 구간추정에서 CLT 의 정의

* CLT = 모집단의 분포에 관계없이 n이 충분히 클 때, 표본평균 $\bar{X}$는 정규분포에 수렴

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

CLT는 구간추정에서 모집단의 분포가 알려지지 않은 경우에도 $\bar{X}$가 정규분포에 근사함을 보장하여 표본평균을 $N(0,1)$로 변환한다. (표준화)

∴ 구간추정의 전제 조건으로 활용된다.

3. 모분산에 관한 추론

a) 귀무가설(H_0) : 공정에 이상이 없다 (모표준편차가 1.5mm 이하이다) = $\sigma^2 \le 1.5^2$

대립가설(H_1) : 공정에 이상이 있다 (모표준편차가 1.5mm 초과이다) = $\sigma^2 > 1.5^2$

b)

$$\bar{x} = \frac{226+228+226+225+232+228+227+229+225+230}{10} = \frac{2276}{10} = 227.6$$

$$S^2 = \frac{1}{10-1}\left((226-227.6)^2 + \cdots + (230-227.6)^2\right) = 5.3778$$

χ^2 검정 $\Rightarrow$ $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2} = \dfrac{(10-1)\cdot 5.3778}{(1.5)^2} = 21.5112$

유의수준 $\alpha = 5\%$ (0.05) $\chi^2_{0.05,9} = 16.919$

카이제곱 분포표에서 자유도 9, 유의수준 0.05일 때, 임계값은 $\chi^2_{0.05,9} = 16.92$

검정통계량 χ^2이 16.92보다 크면 H_0 기각, $\chi^2 = 21.5112$로 귀무가설 H_0 기각.

$\therefore$ 공정의 표준편차 1.5를 초과한다

90% 신뢰구간 $\Rightarrow$ $\alpha = 0.1$

$$\left[\frac{9\times 5.3778}{\chi^2_{0.05,9}},\ \frac{9\times 5.3778}{\chi^2_{0.95,9}}\right] \qquad \chi^2_{0.95,9} = 3.33, \quad \chi^2_{0.05,9} = 16.92$$

$$\left[\frac{9\times 5.3778}{16.92},\ \frac{9\times 5.3778}{3.33}\right], \quad [2.86,\ 14.54]$$

모분산 σ^2의 90% 신뢰구간 = $[2.86,\ 14.54]$

4. 통계적 방법론

a) 귀무가설(H_0) = DSL 학회원의 평균 키가 DSL 학회원이 아닌 사람의 평균 키보다 작거나 같다.

대립가설(H_1): DSL 학회원의 평균 키가 DSL 학회원이 아닌 사람의 평균 키보다 크다.

b) t-검정 통계량

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}, \quad S_p = \text{공통표준편차} = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

$n_1, n_2 = 125$

$S_1^2, S_2^2 = 7.05$, $S_p = \sqrt{\frac{(125-1)\cdot(7.05)^2+(125-1)\cdot(7.05)^2}{125+125-2}} = \sqrt{49.70} = 7.05$

$$t = \frac{173.5-171.4}{7.05\sqrt{\frac{1}{125}+\frac{1}{125}}} = \frac{2.1}{7.05\times 0.13} = 2.29$$

유의수준 $\alpha = 0.05$, 자유도 $df = 248$ 일때, $t_{0.05,248} \approx 1.645$

$t = 2.29 > 1.645$ 이므로 귀무가설 기각

∴ DSL 학회원의 평균 키가 DSL 학회원이 아닌 사람의 평균 키보다 크다