<기초과제 1> - 응용통계학과 2019122053 유현동

#문제 1

(1-1). 중심극한정리란, 모집단이 정규분포가 아니더라도 평균 μ, 분산 σ^2 $(\sigma^2 < \infty)$인 때, 임의로 추출한 n개의 표본의 크기가 충분히 크다면 표본평균 $\bar{X}$의 분포는 $N(\mu, \frac{\sigma^2}{n})$으로 분포 수렴한다는 것이다.
중심극한정리가 갖는 의미는, 모집단의 분포를 모름에도 불구하고 모집단의 분포에 상관없이 표본평균의 분포가 근사적으로 정규분포를 따르며, 이 사실을 이용하여 모평균에 대한 통계적 추정과 검정이 가능해진다는 것이다.

(1-2). 중심극한정리를 이용하여, 표본분산 S^2에 대해서 표본평균을 표준화시킨 값인 $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ 의 신뢰구간을 구하는 과정에 있어서 T가 근사적으로

$N(0,1)$을 따른다는 사실을 이용할 수 있다.

(1-3). n개의 표본 $X_1, \dots, X_n$을 추출했다고 할 때,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n}{n-1}\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right\}$$

$$= \frac{n}{n-1}\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2\bar{X}X_i + \bar{X}^2)\right\} = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2 - \bar{X}^2\right)$$

$$\xrightarrow{P} 1 \cdot \{E(X^2) - \mu^2\} = \sigma^2$$

표본분산 S^2은 모분산 σ^2으로 확률수렴한다.
따라서 n이 충분히 클 때 모분산을 표본분산으로 대체할 수 있다.

문제 2

(2-1). ③ $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

proof) $V = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}+\bar{X}-\mu}{\sigma}\right)^2$

$= \sum_{i=1}^{n}\left\{\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 + \left(\frac{\bar{X}-\mu}{\sigma}\right)^2\right\} = \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 + n\left(\frac{\bar{X}-\mu}{\sigma}\right)^2$

$= \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 + \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 = \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$

$V = \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$의 양변에 mgf를 취해주면

좌변은 $V \sim \chi^2(n)$의 mgf이고 우변은 $\frac{(n-1)S^2}{\sigma^2}$의 mgf와

$\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2(1)$의 mgf의 곱이다.

따라서 uniqueness of mgf에 따라 $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$이다

(2-2). ④ $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

proof) $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{1}{n-1}\cdot(n-1)\cdot\frac{S^2}{\sigma^2}}}$

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

$V = \frac{(n-1)S^2}{\sigma^2}$, $r = n-1$ 이라 하면

$T = \frac{W}{\sqrt{V/r}}$로 t분포의 정의에 해당한다.

따라서 $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$이다.

문제 3.

(3-1). (a). μ_{DSL} : DSL 학회원(동문 포함)의 평균 키

μ_{etc} : DSL 학회원(동문포함)이 아닌 사람의 평균 키

귀무가설 : $\mu_{DSL} = \mu_{etc}$

대립가설 : $\mu_{DSL} > \mu_{etc}$

(b). $n_{DSL} = 101$, $\bar{X}_{DSL} = 178.5$, $S_{DSL} = 7.05$

$n_{etc} = 101$, $\bar{Y}_{etc} = 179.9$, $S_{etc} = 7.05$

$$\text{t-value } (t^*) = \frac{\bar{X}_{DSL} - \bar{Y}_{etc}}{\sqrt{\frac{S^2}{n_{DSL}} + \frac{S^2}{n_{etc}}}} \sim t(200)$$

$$\left(S^2 = \frac{\sum_{i=1}^{101}(X_i - \bar{X}_{DSL})^2 + \sum_{j=1}^{101}(Y_j - \bar{Y}_{etc})}{n_{DSL} + n_{etc} - 2} = \frac{100 \times (7.05)^2 + 100 \times (7.05)^2}{101 + 101 - 2}\right.$$

$$\left. = 49.7025\right)$$

$$t^* = \frac{178.5 - 179.9}{\sqrt{2 \times \frac{49.7025}{101}}} \fallingdotseq -\frac{1.4}{\sqrt{0.9842}} \fallingdotseq -1.4112$$

유의수준 5%일 때 $t(200)$-분포의 임계값은 1.645이다

따라서 $t^* < 1.645$이므로 귀무가설을 기각한다