1-1 큰수의 약한 법칙의 정의를 서술하고 증명하라.

정의: 확률론에서 동일한 확률 분포를 따르는 독립적인 확률 변수들의 평균이 모집단의 기대값에 점근적으로 수렴한다는 것을 의미한다.

증명: 동전을 여러 번 던져서 앞면이 나오는 확률을 계산한다고 생각해보자. 동전 던지기를 많이 반복하면, 앞면이 나오는 비율은 점점 동전이 공정할 때의 실제 확률 (50%)에 가까워진다.

우리가 관심 있는 값은 평균 $\bar{X}_n$ 이다. 표본평균은 다음과 같이 계산된다. $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

여기서 $X_1, X_2, X_3, \dots, X_n$은 각 동전을 던져 나온 결과이다.

각 동전의 결과 X_i의 평균, 즉 기대값은 $\mu = E[X_i]$이다. 동전이 공정하면, $E[X_i] = 0.5$ 이다.

n번 동전을 던졌을 때, 표본 평균 $\bar{X}_n$의 분산은 다음과 같다. $Var(\bar{X}_n) = \frac{\sigma^2}{n}$

여기서 σ^2는 각 동전 던지기 결과 X_i의 분산이다. 동전에서는 $\sigma^2 = 0.5 \times 0.5 = 0.25$ 일 것이다.

체비셰프 부등식에 따르면, 표본 평균이 실제 평균에서 멀어질 확률은 다음과 같이 제한된다.

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{Var(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n \cdot \epsilon^2}$$

↳ 분모 n이 커질수록 $\frac{\sigma^2}{n \cdot \epsilon^2}$은 점점 작아진다.

$P(|\bar{X}_n - \mu| \geq \epsilon)$은 $n \to \infty$ 에서 0으로 수렴된다.

∴ 동전을 던지는 횟수(n)이 많아질수록, 표본 평균이 참값에서 크게 벗어날 확률은 0에 가까워진다.

따라서, 표본평균은 참값에 가까워진다.

2-1 : 중심극한 정리의 정의를 서술해요.

A. 독립적이고 동일한 분포를 따르는 확률 변수들의 합(또는 평균)이 적당히 큰 표본 크기에서 정규분포에 가까워진다는 것을 의미!

수학적 정의

확률 변수 $X_1, X_2, \cdots, X_n$ 이 다음 조건을 만족한다고 하자.

1. $X_1, X_2, \cdots, X_n$ 은 독립적이고 동일한 분포를 따름.
2. 각 X_i 의 기댓값이 $\mu = E[X_i]$, 분산이 $\sigma^2 = Var(X_i)$

표본 평균 $\bar{X}_n$ 을 다음과 같이 정의한다.

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

그러면, 다음이 성립한다.

$$Z_n = \frac{\sqrt{n}\cdot(\bar{X}_n - \mu)}{\sigma} \longrightarrow N(0,1)$$

즉, Z_n은 표준 정규 분포 $N(0,1)$에 점근적으로 수렴한다.

2-2, 중심극한 정리가 통계적 추론 중 '구간추정'에서 어떻게 활용되는지 서술하시오.

중심극한 정리(CLT)는 구간추정에서 모집단의 평균이나 다른 모수를 추정할 때 핵심적인 역할을 한다.
이를 통해 모집단의 분포를 정확히 모르더라도, 표본 통계량의 분포를 정규분포로 근사할 수 있기 때문이다.
∴ 모집단이 정규분포를 따르지 않더라도, 표본 크기 n이 충분히 크다면 중심극한 정리에 의해 표본 평균은 정규분포로 근사된다.
따라서 구간추정을 수행할 때 모집단의 분포를 정확히 몰라도 통계적 추론이 가능하다.

3-1.

이 문제는 판 두께의 표준편차가 기준값(1.5mm)을 초과하는지를 검정하는 분산 검정 문제이다.

a)

① 귀무가설

판 두께의 표준편차 σ가 기준값을 초과하지 않는다. ($H_0: \sigma \le 1.5$)

② 대립가설

판 두께의 표준편차 σ가 기준값을 초과한다. ($H_a: \sigma > 1.5$)

b)

검정통계량 $\chi^2 = \frac{(n-1)\cdot S^2}{\sigma_0^2}$ (n: 표본 크기, S^2: 표본 분산, σ_0^2: 기준 분산, χ^2: 카이제곱 분포를 따름).

유의수준 $\alpha = 0.05$에서, $H_0: \sigma \le 1.5$, $H_a: \sigma > 1.5$

카이제곱 검정의 기각역은 다음과 같다. $\chi^2 > \chi^2_{0.05,9}$

$$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = \underbrace{\frac{226+228+226+225+232+228+227+229+225+230}{10}}\ = 227.6$$

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

$$\chi^2 = \frac{(n-1)\cdot S^2}{\sigma_0^2} \longrightarrow 20.62 \ / \ \underbrace{\chi^2_{0.05,9}}_{}{}^{\text{임계값}} = 16.92$$

검정통계량이 임계값을 초과하므로, 귀무가설을 기각한다.

판 두께 분산의 90% 신뢰구간: (2.74, 13.95)

4-1

a)

귀무가설: DSL 학회원의 평균 키가 비회원의 평균 키보다 크지 않다. (같거나 작다.) (H_0)

$H_0 : \mu_{DSL} \leq \mu_{비회원}$

대립가설: DSL 학회원의 평균 키가 비회원의 평균 키보다 크다. (H_A)

$H_A : \mu_{DSL} > \mu_{비회원}$

b)

DSL 학회원 $(n_1 = 125)$

평균 = 173.5 표준편차 = 7.05

비회원 $(n_2 = 125)$

평균 = 171.4 표준편차 = 7.05

유의수준 $(\alpha) = 0.05$

공동분산 : $S_p^2 = \dfrac{(n_1-1)\cdot S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$

t- 검정통계량 : $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

자유도: $df = n_1+n_2-2 = 248$

$t = 2.35$

$t_{critical} = 1.65$

p값 : 0.0097

t는 임계값보다 크므로, 귀무가설을 기각한다.

& P값은 유의수준보다 작으므로 귀무가설을 기각한다.

∴ DSL 학회원의 평균 키가 비회원의 평균 키보다 크다는 주장은 지지할 충분한 근거가 있다.

4-2.

a)

귀무가설 (H_0):

$$\mu_{DSL} = \mu_{ESC} = \mu_{기타}$$

대립가설 (H_A):

적어도 한 그룹의 평균 키가 다르다.