1-1

중심극한정리: 동일한 확률분포를 가진 확률변수 n개의 평균의 분포는 n이 충분히 크다면 정규분포에 가까워진다.

→ 평균 μ, 분산 σ^2인 임의의 모집단에서 크기가 n인 표본 $(X_1, X_2, \dots, X_n)$의 평균 $\overline{X}$의 분포는

$n \to \infty$인 때 (충분히 큰 때), $N(\mu, \frac{\sigma^2}{n})$에 근사하고 $Z = \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$의 분포는 $N(0, 1^2)$에 근사한다.

1-2

원본 데이터의 분포가 알려지지 않거나 정규분포를 따르지 않더라도, 중심극한정리를 통해 표본 평균의 분포가 정규분포를 따른다고 가정할 수 있으므로, 표준오차를 사용해서 모평균에 대한 신뢰구간을 계산 가능하다.

2-1

(a) $\begin{bmatrix} 5 & 0 \\ 0 & 4 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ (c) $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

2-2

scipy.linalg.svd 함수 사용 → $\Sigma = \begin{bmatrix} 1.4142 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

3-1

For $(n-1)S^2/\sigma^2$, consider $V = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2$

$(X_i - \mu)^2/\sigma^2 \sim \chi^2(1)$ (each $i = 1, 2, \cdots, n$)

X_i 가 independent 나 한 때, $\frac{(X_i-\mu)^2}{\sigma^2}$ 도 independent, $V = \chi^2(n)$ 의 r.v

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

$$E(e^{t(X+Y)}) = E(e^{tX}e^{tY}) = E(e^{tX})E(e^{tY})$$

$$(1-2t)^{-n/2} = E[\exp(t(n-1)S^2/\sigma^2)](1-2t)^{-1/2}$$

Thus, moment generating function of $(n-1)S^2/\sigma^2$ → $E[\exp(t(n-1)S^2/\sigma^2)] = (1-2t)^{-(n-1)/2}$

3-2

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

so $T = \dfrac{W}{\sqrt{V/r}}$ where $W = (\overline{X}-\mu)/(\sigma/\sqrt{n}) \sim N(0,1)$

$V = (n-1)S^2/\sigma^2 \to \chi^2(n-1)$ dist. by C and $r = n-1$ and W & V are independent

Thus, T has a t-distribution with r degrees of freedom

4-1

(a) $H_0 : \mu_1 = \mu_2$ (μ_1 = DSL 학회원의 avg height, μ_2 = Non-DSL 의 avg height)

$H_1 : \mu_1 > \mu_2$ (〃)

(b) 독립 표본 t-검정

$n_1 = n_2 = 210$

$\bar{X}_1 = 173.5$ cm, $\bar{X}_2 = 171.4$ cm, $\sigma = 7.05$ cm

t-value = 3.05, df = n1 + n2 − 2 = 418

유의수준 5% critical value = 1.65

t-value > critical value → reject H_0

결론 → 유의수준 5% DSL 학회원들의 평균 키가 DSL 학회원이 아닌 사람들의 평균 키보다 유의미하게 크다.

4-2

(a) H_0 : 모든 학회의 평균 키는 같다.

$\mu_1 = \mu_2 = \mu_3$ (μ_1 = DSL 평균 키, μ_2 = ESC 평균 키, μ_3 = 그 외 평균 키)

H_1 : 적어도 하나의 학회의 평균 키가 다른 학회와 다르다.

μ_1, μ_2, μ_3 중 최소 1개는 값이 다름.