기수 : 12기　이름 : 복지민

24-1 DSL 정규 세션

기초과제 1 통계적 사고

☑ 본 과제는「통계학입문」,「통계방법론」,「선형대수」및「수리통계학(1)」일부에 상응하는 내용의 복습을 돕기 위해 기획되었습니다. 평가를 위한 것이 아니므로, 주어진 힌트(🖐)를 적극 활용하시고 학회원 간 토론, Slack 의 질의응답을 활용하시어 해결해주십시오. 단, 답안 표절은 금지합니다.
☑ 서술형 문제는 ✎ , 코딩 문제는 © 으로 표기가 되어 있습니다. 각 문제에서 요구하는 방법에 맞게 해결하며, 서술형 문제들은 따로 작성하시어 pdf 로 제출해주시고 코드 문제들은 ipynb 파일에 답안을 작성하시어 제출해주십시오.
☑ **7/25 (목) 23 시 59 분까지** Github 에 PDF 파일과 ipynb 파일을 모두 제출해주십시오. Github 에 제출하는 방법을 모른다면 학술부장 혹은 과제 질의응답을 위한 오픈채팅방을 활용해주시오.
☑ 참고 도서 : 통계학입문(3 판, 강상욱 외), Introduction to Mathematical Statistics(8 판, Hogg et.al.)

문제 1　Central Limit Theorem

중심극한정리는 확률변수의 합 형태 (Sum of Random Variables) 의 극한분포를 손쉽게 구할 수 있도록 해 주기에 통계학에서 가장 자주 사용하는 정리입니다. 이 문제에서는 중심극한정리의 정의와 그 활용에 대해 짚어보겠습니다.

1-1 ✎ : 중심극한정리(Central Limit Theorem)의 정의를 서술하시오.

🖐 통계학입문 (3 판) 7 장 참고
🖐 Hogg(8 판) 4 장 2 절, 5 장 3 절 참고

중심극한정리란 평균 μ, 분산 σ^2인 임의의 모집단으로부터 크기 n의 랜덤표본을 반복적으로 추출할 때, 표본평균 $\bar{X}$는 n이 충분히 크면 ($n \geq 30$; 대표본) 근사적으로 정규분포 $N(\mu, \frac{\sigma^2}{n})$을 따른다는 개념이다.

1-2 ✎ : 중심극한정리가 통계적 추론 중 "구간추정"에서 어떻게 활용되는지 서술하시오.

🖐 Hogg(8 판) 4 장 2 절

중심극한정리를 사용하여 $Z = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim N(0,1)$ 를 나타낼 수 있다. 즉, $\bar{X}$로 모평균 μ의 구간 추정이 가능하다.

$$1-\alpha \approx P_\mu\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < z_{\alpha/2}\right)$$

$$1-\alpha \approx P_\mu\left(\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} < \mu \leq \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right)$$

$$\therefore \mu\text{의 구간} : (\bar{x} - z_{\alpha/2}\cdot S/\sqrt{n},\ \bar{x} + z_{\alpha/2}\cdot S/\sqrt{n})$$

문제 2 Linear Algebra

선형대수학은 머신러닝을 위한 수학 중에서 가장 중요한 요소 중 하나이며, 이 중에서 가장 중요한 것 중에서 하나는 바로 SVD (Singular Value Decomposition, 특이값 분해) 입니다. 이것을 알기 위해서 고유값과 고유벡터를 활용한 Diagonalization 에 대해서 먼저 알아본 후, SVD 를 사용하여 실제로 이미지 압축을 적용해보겠습니다.

2-1 ✎ **:** Diagonalization 의 정의가 다음과 같이 주어졌습니다.

Diagonalization 이란 정방행렬(A) 를 Eigenvalue, Eigenvector 를 통해서 대각행렬 (D) 를 만드는 것이며, 즉 $D = P^{-1} A P$ 를 통해서 대각행렬 (D) 를 찾는 것입니다.

조건들 : 1.) A 는 정방행렬 (Square Matrix) 이다. 2.) $A(n \, x \, n)$ 는 n 개의 독립인 고유벡터를 가지고 있다.

1.) A 에 대한 고유벡터들을 찾으며, 이것을 각각 $P_1, P_2, \cdots, P_n$ 으로 놓는다
2.) $P = [P_1, P_2, \cdots, P_n]$ 메트릭스를 만든다
3.) $P^{-1} A P$ 를 구하면 다음과 같은 형태가 나오게 된다

$$P^{-1} A P = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = D$$

만약에 A 가 대칭 (Symmetric) 행렬이면 다음과 같은 꼴이 나오게 됩니다.

$$A = P D P^T$$

다음과 같은 정방행렬에 Diagonalization 을 적용시켜서 나오게 되는 대각행렬을 쓰시오.

D

a) $\begin{bmatrix} 6 & -1 \\ 2 & 3 \end{bmatrix}$ $D = \begin{bmatrix} 5 & 0 \\ 0 & 4 \end{bmatrix}$

b) $\begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 1 \\ -1 & 0 & 1 \end{bmatrix}$ $D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

c) $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ $D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(a)

$$\begin{bmatrix} 6 & -1 \\ 2 & 3 \end{bmatrix}$$

$$\det(\lambda I - A) = \begin{vmatrix} \lambda-6 & 1 \\ -2 & \lambda-3 \end{vmatrix} = (\lambda-6)(\lambda-3)+2$$

$$= \lambda^2 - 9\lambda + 20$$

$$(\lambda-4)(\lambda-5)$$

$$\lambda_1 = 5,\ \lambda_2 = 4$$

$$(\lambda I - A)x = 0$$

i) $\lambda = 5$,

$$\begin{bmatrix} 5-6 & 1 \\ -2 & 5-3 \end{bmatrix} \Rightarrow \begin{bmatrix} -1 & 1 \\ -2 & 2 \end{bmatrix} x = 0$$

$$\left[\begin{array}{cc|c} -1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right] \Rightarrow \text{eigenvector} : \overset{(P_1)}{\begin{pmatrix} 1 \\ 1 \end{pmatrix}}$$

ii) $\lambda = 4$,

$$\begin{bmatrix} 4-6 & 1 \\ -2 & 4-3 \end{bmatrix} x = \begin{bmatrix} -2 & 1 \\ -2 & 1 \end{bmatrix} x = 0$$

$$\left[\begin{array}{cc|c} 2 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right] \Rightarrow \text{eigenvector} : \overset{(P_2)}{\begin{pmatrix} 1 \\ 2 \end{pmatrix}}$$

$$P = [P_1 \ \ P_2] = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$P^{-1} = \frac{1}{2-1}\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$

$$D = P^{-1}AP = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 6 & -1 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 10 & -5 \\ -4 & 4 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 4 \end{bmatrix}$$

(b) $\begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 1 \\ -1 & 0 & 1 \end{bmatrix}$

$\det(\lambda I - A) = 0$

$$\begin{vmatrix} \lambda-2 & 0 & 0 \\ -1 & \lambda-2 & -1 \\ 1 & 0 & \lambda-1 \end{vmatrix} = (\lambda-2)(\lambda-2)(\lambda-1)+0(-1\cdot 0)+0(-1\cdot 1)-0\cdot(\lambda-2)(-1)-0\cdot(\lambda-2)(1)-0(\lambda-1)(-1)$$
$$= (\lambda-2)^2(\lambda-1)=0$$
$$\lambda=2,\ \lambda=1$$

i) $\lambda=2$,

$$\begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & -1 \\ 1 & 0 & 1 \end{bmatrix} \Rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \Rightarrow \text{eigenvector: } \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

ii) $\lambda=1$,

$$\begin{bmatrix} -1 & 0 & 0 \\ -1 & -1 & -1 \\ 1 & 0 & 0 \end{bmatrix} x = 0$$

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \rightarrow \text{eigenvector: } \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$$

$$P = [P_1\ P_2\ P_3] = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}$$

$$P^{-1} = \left[\begin{array}{ccc|ccc} -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right] = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{array}\right]$$
$$= \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{array}\right] = \begin{bmatrix} -1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

$$D = P^{-1}AP = \begin{bmatrix} -1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 1 \\ -1 & 0 & 1 \end{bmatrix}\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} -2 & 0 & 0 \\ 2 & 2 & 2 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(c) $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

$\det(\lambda I - A) = 0 \qquad \begin{vmatrix} \lambda-3 & 0 & 0 \\ 0 & \lambda-1 & 0 \\ 0 & 0 & \lambda-2 \end{vmatrix} = (\lambda-3)(\lambda-1)(\lambda-2)$

$\lambda = 1, 2, 3$

i) $\lambda = 3$,

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} x = 0$$

$$\left[\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \Rightarrow \text{eigenvector} : \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

ii) $\lambda = 2$,

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} x = 0$$

$$\left[\begin{array}{ccc|c} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \Rightarrow \text{eigenvector} : \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

iii) $\lambda = 1$,

$$\begin{bmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} x = 0$$

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \rightarrow \text{eigenvector} : \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$P = [P_1 \; P_2 \; P_3] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$P^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$D = P^{-1}AP = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

2-2 ✐ **:** SVD 의 정의가 다음과 같이 주어졌습니다.

SVD 란 Diagonalization 과는 달리 모든 행렬 (A) 에 대해서 사용이 가능합니다.

1.) $A^T A,\ A\ A^T$ 행렬들을 만듭니다. 이것은 항상 대칭 (Symmetric) 행렬이 됩니다.
2.) $A^T A\ =\ V\,D\,V^T,\ A\ A^T\ =\ U\,D'\,U^T$ 으로 대각화를 진행을 하고 나서 정규직교화까지 하게 된다면 U 와 V 를 얻게 됩니다.
3.) 여기에서 0 이 아닌 고유값들이 내림차순으로 나열된 것이 바로 D 가 되며, 이것은 바로 Σ 행렬의 대칭 원소들이 됩니다.
4.) 결국 $A = U\,\Sigma\,V^T$ 의 관계를 가지기 때문에 위에서 구한 U 와 V 를 대입시키면 되며 Σ 도 3.) 에서 구했던 걸로 대입을 하면 됩니다.

참고 자료 :

- https://www.youtube.com/watch?v=rziHzFk5JyU
- https://www.youtube.com/watch?v=HeGdlgB8450 (해당 자료를 참고하여 문제를 풀어주세요.)
- https://angeloyeo.github.io/2019/08/01/SVD.html#google_vignette

다음과 같은 행렬에 SVD 를 적용하여 나오는 Σ 행렬을 구하시오.

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$\Sigma : \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

2-3 © **:** .ipynb 파일에서는 SVD (Singular Value Decomposition, 특이값-분해) 를 이미지 압축에 활용하는 예시를 보여주고 있습니다. 해당 코드를 확인한 후, 새로운 사진에 대해 원본에 비해서 적은 용량을 차지하면서도 원본에 대한 정보를 유지해주는 차원 수가 무엇인지 알아냅시오.

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\det(\lambda I - (A^TA)) \Rightarrow \begin{vmatrix} \lambda-1 & -1 & 0 \\ -1 & \lambda-1 & 0 \\ 0 & 0 & \lambda-1 \end{vmatrix} = (\lambda-1)^3 - (\lambda-1)$$

$$= (\lambda-1)(\lambda^2 - 2\lambda + 1 - 1)$$

$$= (\lambda-1)\cdot\lambda(\lambda-2)$$

$$\lambda = 2, 1, 0$$

$\lambda = 2$,

$$\left[\begin{array}{ccc|c} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] = \left[\begin{array}{ccc|c} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \rightarrow \text{eigenvector}: \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad v_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

$\lambda = 1$,

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \rightarrow \text{eigenvector}: \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad v_2 = \frac{1}{\sqrt{1}}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$\lambda = 0$,

$$\begin{bmatrix} -1 & -1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} = \left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \rightarrow \text{eigenvector}: \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad v_3 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$$

$$V = [v_1 \ v_2 \ v_3] = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$u_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$u_2 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$U = [u_1 \ u_2] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix} = U\Sigma V^T$$

문제 3 모분산에 관한 추론

카이제곱 분포는 모집단의 모분산 추정에 유용하게 쓰이며, 정규분포에서의 랜덤표본에서 표본분산과 관계되는 분포입니다. 표준정규분포를 따르는 서로 독립인 확률변수 Z_1, Z_2, $Z_3, \ldots, Z_k$가 있을 때, $V = Z_1, + Z_2 + Z_3 + \cdots + Z_k \Rightarrow V \sim$ 자유도가 k 인 X^2 분포를 따른다고 할 수 있습니다. 대개 모분산에 관한 추론에 사용되며, 검정통계량으로 $X^2 = \frac{(n-1)S^2}{\sigma^2} \sim X^2(n-1)$가 쓰입니다.

3-1 : 플라스틱 판을 제조하는 공장이 있다. 판 두께의 표준편차가 1.5mm 를 넘으면 공정 상에 이상이 있는 것으로 간주합니다. 오늘 아침 10 개의 판을 무작위 추출하여 두께를 측정한 결과가 다음과 같았습니다.

{226, 228, 226, 225, 232, 228, 227, 229, 225, 230}

해당 판 두께의 분포가 정규분포를 따른다고 할 때, 공정에 이상이 있는지를 검정하세요.

a) 귀무가설과 대립가설을 설정하시오.

$H_0 : \sigma^2 \le 1.5^2 \qquad H_a : \sigma^2 > 1.5^2$

b) 유의수준 5%에서의 가설검정을 수행하고 판 두께의 분산에 대한 90% 신뢰구간을 구하시오.

어떤 검정통계량이 어떤 분포를 따르는지, 언제 귀무가설을 기각하는지 정해야 합니다.

$\bar{X} = 227.6 \qquad s^2 = 5.156$

$$\chi^2 = \frac{(10-1)(5.156)}{1.5^2} = 20.624 > \chi^2_{0.05}(9) = 16.919$$

∴ 귀무가설 기각 ; 공정 상에 이상이 있다고 할 수 있다.

신뢰구간 :

$$\frac{(n-1)s^2}{\chi^2_{0.05}(9)}, \frac{(n-1)s^2}{\chi^2_{0.95}(9)} = [2.743, 13.935]$$

문제 4	통계적 방법론

t 검정은 모집단이 정규분포를 따르지만 모표준편차를 모를 때, 모평균에 대한 가설검정 방법입니다. 대개 두 집단의 모평균이 서로 차이가 있는지 파악하고자 할 때 사용하며, 표본평균의 차이와 표준편차의 비율 을 확인하여 통계적 결론을 도출합니다. ANOVA Test 의 경우 집단이 2 개보다 많은 경우 모평균에 차이가 있는지 파악하고자 할 때 사용되며, 이것은 코드로만 살펴보겠습니다.

4-1 ✎ **:** 어떤 학우가 DSL 학회원(동문 포함)의 평균 키가 DSL 학회원이 아닌 사람의 평균 키보다 크다고 주장하여, 실제로 그러한지 통계적 검정을 수행하려고 합니다. 며칠간 표본을 수집한 결과 다음과 같은 값을 얻었습니다.

표본 수: 총 250 명, 각 125 명
측정에 응한 DSL 학회원들의 평균 키 : 173.5cm / 표준편차 : 7.05cm
측정에 응한, DSL 학회원이 아닌 사람들의 평균 키 : 171.4cm / 표준편차 : 7.05cm

a) ✎ 귀무가설과 대립가설을 설정하시오.

$H_0: \mu_{DSL} = \mu_{not\,DSL}$; $\mu_{DSL} - \mu_{not\,DSL} = 0$ $\qquad H_a: \mu_{DSL} > \mu_{not\,DSL}$; $\mu_{DSL} - \mu_{not\,DSL} > 0$

b) ✎ 유의수준 5%에서의 가설검정을 수행하고 결론을 도출하시오. (단, 키는 정규분포를 따르며 각 집단의 분산은 같다고 가정한다.)

☝ 통계학입문(3 판) 7 장 참고
☝ 어떤 검정통계량이 어떤 분포를 따르는지, 언제 귀무가설을 기각하는지 정해야 합니다.

$s_p = \sqrt{\frac{(124)(7.05)\cdot 2}{248}} = 2.655$ X : DSL, Y : notDSL

$t = \frac{(173.5-171.4)-(\mu_X-\mu_Y)}{7.05\sqrt{\frac{2}{125}}} = 2.35489 > z_{0.05/2} = 1.96$

∴ 귀무가설 기각 ; 유의수준 5%에서 DSL 학회원의 평균 키가 DSL 아닌 학회원의 평균 키보다 크다고 할 수 있다.

4-2 © **:** 한 학우가 이번에는 각 학회의 평균 키가 똑같다는 주장을 하였습니다. 해당 학우가 제공한 ESC 학회의 학회원별 키 데이터를 활용해 가설검정을 진행하고자 합니다.

a) ✎ 귀무가설과 대립가설을 설정하시오.

$H_0: \mu_{DSL} = \mu_{ESC} = \mu_{Else}$ $\qquad H_a$: At least one of $\mu_{DSL}, \mu_{ESC}, \mu_{Else}$ is different.

b) © 파이썬의 scipy.stats 을 활용해서 유의수준 5%에서의 가설검정을 수행하고 결론을 도출하시오. 결론은 .ipynb 파일에 쓰셔도 괜찮습니다.

☝ One-way Anova Test 를 활용해서 사용하는 문제입니다.
☝ 활용해야 될 함수는 scipy.stats.f_oneway 입니다.

[기초과제1.ipynb 참조]

문제 5	© Numpy + Pandas 활용

기초과제.ipynb 파일에 제공된 문제를 참고하여 수행하기 바랍니다.

Reference

- Introduction to Mathematical Statistics(8 판, Hogg et.al)
- 23-2 기초과제 1 (9 기 이성균)
- 24-1 기초과제 1 (10 기 신재우)

Data Science Lab

담당자 : 11 기 김현진, 11 기 김정우

Rlaguswls186790@yonsei.ac.kr

kjungwoo@yonsei.ac.kr