

성적 생성 및 분석하기

송실대학교
베어드교양대학
강의선 교수
iami86@ssu.ac.kr

random 함수 : 난수 생성

- 0.0~1.0사이의 실수형 난수
 - `np.random.rand(a, b)`
 - ✓ (a, b) : 난수로 이루어진 2차원 array 생성
 - # b 생략시 난수로 이루어진 1차원 array 생성
- a부터 b까지의 정수형 난수 생성
 - `np.random.randint(a, b, size=(x,y))`
 - # size(x,y) : 2차원 array 로 생성
- 정규분포를 갖는 난수 생성
 - `np.random.normal(a, b, c)`
 - #a : 평균, b : 표준편차, c : 생성할 개수

2

기본 통계함수

- 변수명의 합계 계산
 - `np.sum(변수명)`
- 변수명의 평균 계산
 - `np.mean(변수명)`
- 변수명의 분산 계산
 - `np.var(변수명)`
- 변수명의 표준편차 계산
 - `np.std(변수명)`

3

정형 데이터 생성 및 조건에 맞는 데이터 추출

- 행과 열로 구성된 정형 데이터 생성
 - `변수명 = pd.DataFrame({'열이름' : '변수명' })`
 - ✓ 열이름 : 칼럼 이름 설정
 - ✓ 열 : 열이름 칼럼에 대한 열 데이터
 - # 여러 개의 열이름을 지정하고자 할 때 : {'열이름1': '변수명1', '열이름2': '변수명2',...}
- 조건식에 맞는 index 추출
 - `변수명.query('조건식')`
 - #조건식 : 논리연산자, 관계연산자 사용
- 인덱스 리셋
 - `변수명.reset_index(drop=True, inplace=True)`
 - #drop=True 옵션은 기존 인덱스는 버리고 새로 인덱스 설정

4

시각화 하기

import matplotlib.pyplot as plt

- **plt.scatter**(변수명['열이름1'], 변수명['열이름2']) #스캐터 차트
- **plt.hist**(변수명['열이름1'], bins=정수) #히스토그램 차트

import seaborn as sns

- **sns.relplot**(data=변수명, x= '열이름' , y= '열이름' , hue= '열이름');
#hue= '열이름' : 열의 데이터를 기준으로 다른 색상으로 표현

수고하셨습니다.