**문제. 1**

Z-스코어는 데이터 포인트가 평균에서 얼마나 떨어져 있는지를 표준 편차로 단위로 나타낸 값입니다. 이를 통해 개별 데이터 포인트가 전체 데이터 분포에서 어느 위치에 있는지를 파악할 수 있으며, 아래와 같이 해석이 가능합니다.

* Z-스코어가 ?1?이면 데이터 값이 평균과 일치함
* Z-스코어가 ?2?이면 데이터 값이 평균보다 큼
* Z-스코어가 ?3?이면 데이터 값이 평균보다 작음
* Z-스코어가 ?4? 이상이거나 ?5?이하인 경우 해당 데이터 포인트는 이상치로 간주될 수 있음

**위 빈칸을 채워주세요!**

?1? :

?2? :

?3? :

?4? :

?5? :

문제. 1 답

?1?: 0 / ?2?: 양수 / ?3?: 음수 / ?4?: 2 / ?5?: -2

빨강 글씨를 빨강색 하이라이팅으로 가려 둔거라, 빨강색 하이라이팅만 제거하면 답 볼 수 있습니다!

아니면 글씨 색깔을 바꿔도 답을 볼 수 있을 거 같아요!

**문제. 2**

import numpy as np

from scipy import stats

data = [12, 15, 12, 18, 20, 22, 18, 18, 15, 14]

mean = ???

median = ???

mode = ???

print("평균:", mean)

print("중앙값:", median)

print("최빈값:", mode)

**Numpy와 scipy 라이브러리를 이용해 평균과 중앙값, 최빈값을 구하려고 할 때, 위 ???에 어떤 식으로 코드를 작성해야 할까요?**

문제. 2 답

np.mean(data)

np.median(data)

stats.mode(data)[0][0] 또는 stats.mode(data)[0] -> 버전에 따라 두 번째로 작성해도 가능

**문제. 3-1, 2**

import pandas as pd

import numpy as np

data = {'A': [1, 2, np.nan, 4, 5],

'B': [5, np.nan, np.nan, 8, 10],

'C': [np.nan, 2, 3, 4, 5]}

df = pd.DataFrame(data)

3-1: 위처럼 데이터 프레임이 있을 때, Pandas를 통해 결측값이 포함된 행을 제거하려면, 어떤 메서드를 사용할까요?

문제. 3-1 답

df.dropna()

3-2: 위처럼 데이터 프레임이 있을 때, Pandas를 통해 결측값이 포함된 행을 제거하지 않고 각 열의 평균값으로 대체하려면, 어떤 메서드를 사용할까요?

문제. 3-2 답

df.fillna(df.mean(), inplace=True)